Social Mechanisms and Political Economy: When Lobbyists Succeed, Pollsters Fail and Populists Win

David K. Levine

Andrea Mattozzi

Salvatore Modica

First Draft: May 27, 2016; This Version: December 19, 2022.

About the Authors

David K. Levine is Professor of Economics and joint chair of the RSCAS at the European University Institute and John H. Biggs Distinguished Professor Emeritus in Economics at Washington University of St. Louis. Email: david@dklevine.com; website: www.dklevine.com.

Andrea Mattozzi is Professor of Economics at the European University Institute. Email: andrea.mattozzi@eui.eu

Salvatore Modica is Professor of Economics at the University of Palermo. Email: salvatore.modica@gmail.com

Preface

This is a book about a theory of social mechanisms and the behavior of groups. Our focus is on applications to democracy and lobbying. The evidence suggests to us that economic fundamentals matter and we present a systematic theory of how and why they do. The book has two goals:

1. To systematically introduce new tools for incorporating sociological elements of peer pressure and social networks into modeling the behavior of interest groups. In doing so we incorporate standard elements of economic theory: especially incentive constraints and auction theory.

2. To show that these tools shed light on democracy and lobbying. In particular we ask why large interest groups succeed in political elections yet are undermined by smaller groups in lobbying. We examine issues such as the efficacy of interest groups, voter turnout, and the rise of populism.

The book is designed for advanced undergraduates and graduate students in economics and related disciplines such as political science or sociology. The basic prerequisite is familiarity with calculus, probability theory, basic non-cooperative game theory and especially Nash equilibrium.

With respect to the exposition we use boxes and appendices for material that is "optional." This includes proofs for which the details involve more advanced mathematics or straightforward but uninteresting calculations, and digressions that we find interesting but are outside the main line of development. We also have marked as "Technical" chapters that have important results useful to those interested in pursuing research in the area but are not essential to understanding the theory. Along with some of the appendices these chapters require higher level mathematical knowledge.

We are grateful to financial support from the European University Institute Research Council and the Ministero dell'Istruzione e Ministero dell'Università e della Ricerca di Italia (MIUR): PRIN "Voting, Lobbying and Disrupting: Political Economy via Economic Sociology" grant number prj-0317_d1.

We owe a special thanks to our co-authors who contributed to the research reported here: Rohan Dutta and Andrea Galeotti in particular. We are grateful also to the people whose work and discussion contributed much to our understanding of the subject: especially Juan Block, Michele Boldrin, Giacomo Calzolari, Drew Fudenberg, Helios Herrera, Philipp Kircher, Cesar Martinelli, Massimo Morelli, Salvatore Nunnari, Thomas Palfrey, Kirill Pogorelskiy, and Guido Tabellini.

This book is dedicated to John and to our partners who put up with much in the writing of the book and the many papers and revisions of papers that form its basis: to Catharina Tilmans, Daniela Iorio, and Marta Terranova.

Contents

[1]This chapter is based on Levine et al. (2020).

[2]This chapter is based on Dutta et al. (2021).

[3]Based on Dutta et al. (2022).

[4]This chapter and the next are based on Levine and Mattozzi (2020) and presentations in 2016 at the Young Economists Workshop and the Maastricht Games Conference, in 2017 at Royal Holloway, at a public lecture at Washington University in St. Louis, at the Robert Schuman Centre for Advanced Studies of the European University Institute, and at the University of the Pacific (Lima), and in 2020 at Erasmus.

[5]This chapter is based on Levine and Modica (2017).
[6]This section is based on unpublished work with Rohan Dutta.
[7]This section is based on unpublished work with Andrea Galeotti.
[8]This chapter is based on Dutta et al. (2018).
[9]This chapter is based on Levine and Mattozzi (2019).

[10]We are grateful to Christian Ewerhart for pointing out to us the equivalence with zero-sum games. This result can be found in Ewerhart and Valkanova (2020) and the mix and match property of minimax equilibria can be found in his earlier work in Ewerhart (2017) and Ewerhart and Sun (2018).

[11]This chapter is based on Levine (2021a).

[12]This chapter is based on Levine (2021b).

CHAPTER 1

Introduction

Following the end of the cold war, democratic institutions spread across the globe bringing with them competition and capitalism. The benign guidance of a "neo-liberal" elite - including the bulk of professional economists - brought an era of peace, growth, and prosperity without parallel in human history. Books were written about the "end of history." Today rich and advanced countries see rioting in the streets; populist charlatans claiming to represent the "people" attack trade and immigration - the geese that laid the golden eggs - making ever changing promises that will never be fulfilled. Expert opinion is condemned and ignored. What went wrong?

Conventional wisdom in economics is that markets work well while political systems are subverted by special interest lobbying. The most extreme expression of this point of view is found in the Chicago school, but the range of disagreement is small: while the Chicago school argues that the benefits of government action are undone by regulatory capture, the MIT school argues that moderate regulation and anti-trust rules can be useful in the occasional circumstance of market failure. In evaluating this near consensus there are two crucial questions that ought to be addressed:

1. Is it true that markets work well? Special interest groups that are effective at lobbying are small in relative terms but large in absolute terms. If these groups are able to overcome the public goods problem in order to lobby why can they not also overcome the public goods problem of cartel formation? If they can, markets will be just as subverted as political systems. More specifically why are some large groups such as farmers successful at lobbing but not at cartel formation, while others such as trade unions are successful at both?

2. When are political systems subverted by special interest lobbying? Both lobbying and elections are forms of political contests. In the conventional view small groups are successful at lobbying but fail in elections. Why are special interests successful in one type of contest but not the other? Yet some small groups such as teachers unions are effective at winning school board elections, while other small groups such as those arguing for minority rights often lack political influence.

Persuasive and systematic answers to these questions are important for a variety of reasons: not least that one of the commonly suggested driving forces behind populism is the subversion of democratic systems by special interests. Without understanding the disease it is difficult to propose a cure.

The purpose of this book is to explore a theory that provides coherent answers to these basic questions. Our conclusions are not always the obvious ones - for example, we find that making lobbying more difficult will increase the influence of special interests, while making voting easier can reduce voter turnout. Our theory enables us to examine the issue of why social and cultural norms at times persist after they become dysfunctional while at other times change with blinding speed. We explain why the taxes most favored by economists - taxes on wealth and Pigouvian taxes - face nearly insurmountable political obstacles. By laying down a firm foundation for thinking about political economy we uncover sources of our current malaise.

Our theory is a theory of group behavior which takes into account individual incentive compatibility. It is a theory that takes as its point of departure that whenever group members are asked to contribute to a public good such as lobbying or voting, or limiting output to exploit market power, that group faces a free rider problem: each member would prefer that other members bear the cost of the public good. The theory is grounded in empirical research showing that groups can be effective at disciplining members through social means including exclusion and ostracism - in a word, through peer pressure - and that the social norms that emerge endogenously in self-organizing groups are functional and effective in overcoming free riding problems. It is a theory that recognizes that groups need to monitor compliance with social norms. Above all, it is a theory that recognizes that groups respond to incentives in different ways than do individuals. It recognizes, for example, that distribution is not neutral so which group gets the proceeds of a tax is crucial to understanding the political consequences of tax policy.

Formally, our theory views a group of individuals with relatively common interests as facing a mechanism design problem. The design problem is the choice of a social norm in the form of a target contribution to public good production together with a punishment for failing to comply with that norm. With imperfect information about the compliance of individual members, how optimally should the group discipline members and what does this mean for the optimal production of public goods? To solve this problem we draw on existing mechanism design theory.

In particular, our theory is *individualistic* in the sense that group members do not act out of altruism or concern for group welfare but pursue their own interest. It is *collectivist* - as is most mechanism design theory - in the sense that the group is assumed to be able to agree on an optimal mechanism that satisfies individual incentive constraints. It is *behavioral* in that it recognizes the importance of social interaction and that exclusion is a powerful punishment that can be used to overcome free riding.

Our flavor of mechanism design theory has a number of features that have not played a prominent role in earlier work on mechanism design:

1. It emphasizes the use of punishments rather than rewards. This recognizes the fact that political and social groups, unlike business firms, do not generate much by way of revenue that can be used for rewards. Hence, it leads to an emphasis not on implementation of the first best, but rather on the costs of implementing a second best;

2. Our monitoring technology is endogenous. Standard cartel and mechanism design theory assume that there is an exogenous noisy signal of individual behavior. We assume that the signal depends upon the targets established by the group: once the target is established the monitoring technology is specialized to the detection of violations. Not only is this a more sensible assumption but it greatly simplifies analysis of the mechanism design problem;

3. We focus on mechanisms that coexist and compete with each other - as is the case, for example, when two political parties each design a mechanism to employ in an electoral contest;

4. We explicitly account for adjustment costs. While it is costly for individuals to change plans in response to circumstances it is especially costly for groups. Moreover adjustment costs for groups differ from adjustment costs for individuals: groups have options not available to individuals including both disbanding and engaging in political activity.

Our perspective on economic research is that theory directs our thinking about data - what questions to ask, what data to collect, and how to analyze it. Strong theoretical and conceptual underpinnings we believe are essential to that thinking, so our theory is directed at foundational issues. We understand as well that analysis and policy cannot always wait for the most satisfactory theoretical foundations and a great deal of research in political economy has been conducted using different underpinnings. This does not mean that we reject this research - especially not as much of it has substantial empirical validation. A new theory if it is to be useful has to have substantial backwards compatibility - it must not unexplain that which has already been explained - it must agree with valid insights that already exist. While the primary focus of the book is on what is new and not on what has been done by others, we wrap up with explicit consideration of existing models and explanations and how and when our theory is compatible with them.

The theory of social mechanisms is a new one founded in research we began about a decade ago. It raises new questions and proposes new answers to old questions. As it is a new theory it has not been subject to the scrutiny applied to existing theories. The goal of this book is to lay out the theory in a coherent way along with relevant facts and to argue that it has enough potential to deserve careful empirical and theoretical scrutiny.

CHAPTER 2

Social Mechanisms: Politics and Markets[1]

In practice large groups have little difficulty in overcoming free-rider problems. Often coercion is involved: for example through mandatory voting laws, a military draft or penalties for tax evasion. In the context of political groups this kind of direct coercion is seldom relevant - farm lobbies cannot imprison non-contributors. There is, however, another form of coercion: peer pressure. This is not a new idea - indeed we might almost argue that it is common sense. In the context of voting, for example, Della Vigna et al. (2014) demonstrate that an important incentive for citizens to vote is to show others that they have voted, while Gerber et al. (2018) show that social pressure significantly increase turnout. Amat et al. (2018), using historical elections data of Spain's Second Republic, show that turnout was driven by political parties and trade unions' social pressure. Peer pressure in turn operates by punishing individuals for violating social norms: generally by some form of exclusion from social activities as described by Coleman (1988), but also through lesser and greater punishments ranging from sneering to beating.[2] People adhere to social norms because they want to keep the good opinion of their friends and neighbors.

Social norms and the punishments that enforce them are not arbitrary. In the setting of public goods, scholars such as Coase (1960) and Ostrom (1990) have shown that social norms are highly functional in solving free-rider problems. In the context of voting we know that turnout in U.S. national elections, for example, is strategic - it is considerably higher in presidential election years than off-years, and in general participation rates and the social norms that support them adjust strategically to reflect the stakes in the elections. We would prefer that interest groups were less strategic - it would be well if lobbying groups were not effective in looking out for their best interests. Unfortunately it rarely seems so. For better or worse social norms are well adapted to group circumstances.

The issue of how a group uses incentives to overcome a public goods problem is the heart of mechanism design theory. In the context of social organization this approach was pioneered by Townsend (1994) in his studies of risk sharing. This is the approach we adapt. In this chapter we introduce the idea of a social mechanism that solves free-rider problems and show how we can derive insights by varying the economic context. We consider first a canonical public goods problem showing how monitoring costs play a key role in determining the amount of public good that will be provided. We then consider a market setting and show that the inability of firms to increase output plays a key role in the formation of large cartels. Hence while the public goods problem of lobbying will be solved in industries where monitoring

[1] This chapter is based on Levine et al. (2020).

[2] Our focus on enforcement is in contrast to models of social conformity such as Akerlof and Kranton (2005), which do not explicitly consider punishments or rewards.

costs are low, the same industries may none-the-less fail to form cartels if firms can easily increase their output. Here we find an answer to our first fundamental question of why markets are more difficult to subvert than political systems.

2.1. The Canonical Public Goods Problem

To introduce the idea of a social mechanism we start with the canonical public goods problem. We consider a large organized group with many members. Each member i chooses the amount of public good $x^i \geq 0$ to produce at unit marginal cost. Each member faces a capacity constraint $X > 1$. The average output of the group $\overline{x}$ represents a public good benefiting each member by $W(\overline{x}) \equiv (V+1)\overline{x} - (V/2)\overline{x}^2$ up to a satiation point $\overline{X}$ and with $V > 0$. Hence a representative group member receives utility

$$W(\overline{x}) - x^i.$$

Since W is strictly concave, the *first best* is where the marginal benefit of the public good $W'(\overline{x}) = (V+1) - V\overline{x}$ is equal to the marginal cost of producing it (which is equal to 1). The condition $W'(\overline{x}) = 1$ gives $\overline{x} = 1$, which is then the first best. In other words, we have normalized the units of output so that $\overline{x}$ is measured as a fraction of the first best. Notice also that $W'(\overline{x}) = 0$ at $\overline{x} = (V+1)/V$ which we therefore take to be the satiation point $\overline{X}$. Finally observe that per capita utility at the first best is $W(1) - 1 = V/2$ so that V is a measure of the value of the public good.

The effect of any individual member in a large group on average output is negligible, so there is a severe free-rider problem. Each group member would prefer not to produce at all, leaving production up to the others. The key element of our model of peer pressure is that individual production can be monitored and those who fail to produce can be punished. Specifically the group may establish a target level of output - a *social norm* - φ and receive a noisy signal $z^i \in \{0,1\}$ about whether member i respected the social norm where 0 means "good, respected the social norm" and 1 means "bad, failed to respect the social norm." If the social norm was respected ($x^i = \varphi$) the bad signal occurs with probability $\pi \geq 0$; if the social norm was violated ($x^i \neq \varphi$) the probability of the bad signal is at least as high $\pi_1 \geq \pi$. When the signal is bad the group imposes an endogenous utility penalty of $P \geq 0$ on the member with the bad signal.

The combination of a social norm φ and penalty P is a *social mechanism*. However, a social mechanism is only meaningful if group members are willing to adhere to it: that is, if it is *incentive compatible*. As all group members share the same interest we assume that the group collectively chooses the incentive compatible social mechanism (φ, P) that when adhered to provides the greatest expected utility $W(\varphi) - \varphi - \pi P$ to its members. The incentive compatibility constraint is that an individual deviation should not be profitable, that is $W(\varphi) - \varphi - \pi P \geq \max_{x^i} W(\varphi) - x^i - \pi_1 P$.

We emphasize three key features of this model, all to be examined later:

- Punishment is costly to the individual hence to the group. The probability π a bad signal is received when the social norm is adhered to plays a crucial role because it makes punishment costly even when the social norm is

adhered to. When $\pi > 0$ punishments are not hypothetical threats, they must sometimes be carried out.
- Signals and punishments are individual. Punishments are not based on how well the group does nor are they collective.
- The monitoring technology is specialized to the social norm. Choosing a different social norm changes which individual behavior will generate good and bad signals.

Following the insight of mechanism design theory we abstract from many details. We do not specify how the mechanism is to be implemented - whether the punishments are ostracism, social disapproval or even monetary fines; who receives the signals; and who carries out the punishments. Similarly the process by which the group agrees on the mechanism is not described. Many of these details of implementation we will discuss subsequently. Our current and central focus here is on the economic fundamental of what information is available to the group about individual behavior.

We may then summarize the social mechanism problem as follows: The goal of the group is to maximize $W(\varphi) - \varphi - \pi P$ subject to the incentive constraint $W(\varphi) - \varphi - \pi P \geq \max_{x^i} W(\varphi) - x^i - \pi_1 P$.

2.1.1. Direct, Monitoring, and Total Cost. It is useful in dealing with social mechanisms to break the problem of finding an optimal mechanism into two parts. First we consider the problem of minimizing the cost of achieving a particular social norm. This requires us to find the incentive compatible punishment with the least social cost. Specifically, for given φ we must minimize πP subject to incentive compatibility. Letting $\hat{P}$ be the solution to that problem, there are two sources of cost of producing output φ: there is the direct cost $D(\varphi) = \varphi$ and there is the monitoring cost $M(\varphi) = \pi\hat{P}$. Notice that the monitoring cost is the cost of punishing the compliant member - the cost of punishing the innocent. The total cost $C(\varphi)$ is simply the sum of the two.

THEOREM 2.1.1. *A social norm $\varphi > 0$ is feasible if and only if $\pi_1 > \pi$ in which case the optimal incentive compatible punishment is $\hat{P} = \varphi/(\pi_1 - \pi)$. Defining the* monitoring difficulty *as $\theta \equiv \pi/(\pi_1 - \pi)$, costs are given by $D(\varphi) = \varphi$, $M(\varphi) = \theta\varphi$ and $C(\varphi) = (1 + \theta)\varphi$.*

PROOF. If a group member chooses to violate the social norm, the individual objective $-x^i - \pi_1 P$ is maximized by producing $x^i = 0$. Hence the incentive constraint may be written as $-\varphi - \pi P \geq -\pi_1 P$ or $(\pi_1 - \pi)P \geq \varphi$. Thus $\pi_1 = \pi$ forces $\varphi = 0$. When $\pi_1 > \pi$, since the goal is to minimize πP, the least cost is the least value of P satisfying the inequality $(\pi_1 - \pi)P \geq \varphi$, that is, the value such that this holds with equality, $(\pi_1 - \pi)\hat{P} = \varphi$. That $M(\varphi) = \theta\varphi$ follows directly. □

This result is simple but important.

- The optimal *gain to deviating* is the cost saving φ of ignoring the social norm and not contributing to the public good - thereby getting $W(\varphi)$ instead of $W(\varphi) - \varphi$. The binding incentive compatibility constraint $\varphi = (\pi_1 - \pi)\hat{P}$ says that the gain to deviating should equal the corresponding increase in expected punishment.

- The monitoring cost is the monitoring difficulty times the gain to deviating. The monitoring difficulty θ is made up of two parts. The denominator $\pi_1 - \pi$ is a measure of the quality of the signal - how much more likely is the bad signal if the social norm is not adhered to? The lower the quality of the signal the greater the difficulty of monitoring. The numerator π is the chance of an erroneous bad signal and is the frequency with which it is necessary to erroneously carry out punishments. The greater this is, the greater is the monitoring difficulty.

The consequence of monitoring difficulty θ is simply to raise the marginal cost of producing output from 1 to $1+\theta$. Insofar as we deal with situations in which θ remains fixed and costs and benefits must be estimated from data this theory is observationally equivalent to a theory in which there are no monitoring costs. As that theory is widely used in political economy, we see that this model will not contradict existing results. Rather it adds a new dimension of understanding - what happens if θ changes - and as we shall see enables us to analyze new problems not so sensibly solved by a theory of the first best.

2.1.2. The Optimal Social Norm. Having minimized cost the mechanism design problem can be reduced to the problem of choosing an optimal social norm φ to maximize $W(\varphi) - C(\varphi)$. From this we immediately compute that

THEOREM 2.1.2. *The optimal social norm is given by* $\hat{\varphi} = \max\{0, (V-\theta)/V\} \leq 1$.

PROOF. From Theorem 2.1.1 the objective is $W(\varphi) - C(\varphi) = (V+1)\varphi - (V/2)\varphi^2 - (1+\theta)\varphi$. This is strictly concave and the first order condition is $V - V\varphi - \theta = 0$; if $V < \theta$ one has $W'(\varphi) - C'(\varphi) < 0$ for all $\varphi \geq 0$. □

The takeaways from this are

- For a given value of the public good V, output is decreasing in the monitoring difficulty θ, ranging from the first best when $\theta = 0$ and reaching zero when $\theta = V$.
- For a given monitoring cost θ, output is increasing in the value of the public good, ranging from zero if $V \leq \theta$ and approaching the first best as $V \to \infty$.

Neither of these conclusions are surprising, nor are they intended to be. They show that a social mechanism model that capture the key elements of peer pressure and endogenous choice of social norm solves the problem that is was set up to solve: it shows how the production of a public good depends as we would expect on the difficulty of monitoring and the value of the public good relative to that monitoring difficulty. The key point is that since group organization does not much depend on the particular free-rider problem that is to be solved, this same model delivers interesting and less obvious, yet valid, predictions about a variety of other economic questions that it was not set up to solve. We proceed to the first of these.

2.2. The Canonical Cartel Problem

The public good problem above could well be that of an industry raising resources from firms to engage in a lobbying effort. Another public goods problem faced by an industry is that of forming a cartel - and we now apply social mechanism theory to that problem.

We continue to consider a large organized group with many members, which we now explicitly identify as firms. Note that this differs from standard cartel theory in which there are only a few firms. A representative firm i chooses the amount of *private* good $x^i \geq 0$ to produce at a unit cost. Each firm continues to face a capacity constraint $X > 1$. The firms produce perfect substitutes and sell them into a market that values average output $\overline{x}$ at $U(\overline{x}) \equiv (V+1)\overline{x} - (V/2)\overline{x}^2$ with $V > 0$ up to the satiation point $\overline{X} = (V+1)/V$. The price in this market is $U'(\overline{x})$ so the profit of a representative firm i is revenue $U'(\overline{x})x^i$ minus cost, that is

$$U'(\overline{x})x^i - x^i.$$

Here price $U'(\overline{x}) = (V+1) - V\overline{x}$ is equal to the marginal cost at $\overline{x} = 1$: this is the *competitive equilibrium.*

Again the group faces a free-rider problem as each firm has a negligible effect on average industry output, and consequently on price. Behaving as price takers, if the price-cost margin $\mu(\overline{x}) \equiv U'(\overline{x}) - 1$ is positive each firm will try to produce to capacity pushing the industry to competitive equilibrium and zero profits. The group would prefer the first best which is here the monopoly output $\overline{x} = 1/2$.

Monopoly output is derived from the first order condition $U'(\overline{x})+\overline{x}U''(\overline{x})-1 = 0$ for maximizing industry profit. The monopoly profits are $V/4$ while the total surplus from competition is $V/2$ (triangle of base 1 and height V). Consequently we may think of V as a measure of the value of output in this industry. Notice that in contrast to the previous public good case, here the higher the output the worse the result for the group.

As in the public good case, we assume the group, now a cartel, may establish a social norm in the form of a target level of output φ and receive a noisy signal $z^i \in \{0, 1\}$ about whether firm i respected the social norm where as before 0 means "good, respected the social norm" and 1 means "bad, failed to respect the social norm." If the social norm was respected ($x^i = \varphi$) the bad signal occurs with probability $\pi \geq 0$; if the social norm was violated ($x^i \neq \varphi$) the probability of the bad signal is at least as high $\pi_1 \geq \pi$. When the signal is bad the cartel imposes an endogenous utility penalty of $P \geq 0$ on the firm with the bad signal. Again the cartel collectively chooses the incentive compatible social mechanism (φ, P) that when adhered to provides the greatest utility $U'(\varphi)\varphi - \varphi - \pi P$ to its members. As it would yield negative profits it is a bad idea to choose a social norm $\varphi > 1$ so we restrict attention to $0 \leq \varphi \leq 1$. Recall that the price-cost margin is $\mu(\varphi) = U'(\varphi) - 1$.

THEOREM 2.2.1. *A social norm $\varphi > 0$ is feasible if and only if $\pi_1 > \pi$ in which case the optimal incentive compatible punishment is $\hat{P} = \mu(\varphi)(X-\varphi)/(\pi_1 - \pi)$ and costs are given by $D(\varphi) = \varphi$, $M(\varphi) = \theta\mu(\varphi)(X - \varphi)$, and $C(\varphi) = (1 - \theta\mu(\varphi))\varphi + \theta\mu(\varphi)X$.*

PROOF. Follows from observing that if a social norm $\varphi \leq 1$ is to be violated, the optimal choice of action is to increase output to the maximum $x^i = X$ so the gain to deviating is $\mu(\varphi)(X - \varphi)$. □

As in the canonical public goods model the monitoring cost is equal to monitoring difficulty times the gain to deviating. Now, however, the gain to deviating is the gain to increasing output from the social norm to capacity X, and is proportional to the price-cost margin $\mu(\varphi) = U'(\varphi) - 1$. Notice that monitoring difficulty θ does not simply change the marginal cost of producing output - there is a second term in total cost $\theta\mu(\varphi)X$ that is not present in the simple public good model.

A simple calculation then shows that

THEOREM 2.2.2. *The optimal social norm is given by*

$$\hat{\varphi} = \min\left\{1, \frac{1}{2} + \frac{\theta}{2(1+\theta)}X\right\}.$$

PROOF. The price cost margin is $\mu(\varphi) = V(1-\varphi)$. From Theorem 2.2.1 the objective is

$$\begin{aligned}\mu(\varphi)\varphi - \theta\mu(\varphi)(X-\varphi) &= (1+\theta)V(1-\varphi)\varphi - \theta V(1-\varphi)X \\ &= (1+\theta)V(1-\varphi)\left[\varphi - \frac{\theta}{1+\theta}X\right]\end{aligned}$$

This is a concave quadratic function with zeros at $\varphi = 1$ and $\varphi = \theta X/(1+\theta)$; so if $\theta X/(1+\theta) \geq 1$ the maximum for $0 \leq \varphi \leq 1$ is 1; otherwise it is the midpoint of the two zeros. □

The takeaways from this are the following (keeping in mind that higher output is bad for the cartel):

- For a given capacity constraint X, output is increasing in the monitoring difficulty θ, ranging from the monopoly solution $\varphi = 1/2$ when $\theta = 0$ and reaching the competitive equilibrium $\varphi = 1$ when $\theta = 1/(X-1)$.
- The value of output V does not matter.
- For a given monitoring cost θ, output is increasing in the capacity constraint, reaching the competitive equilibrium when $X = (1+\theta)/\theta = 1 + 1/\theta$.

The first of these conclusions is not surprising and is the same as is the case for public goods production. Greater monitoring difficulty inhibits cartel formation as it increases output (which means it reduces public goods production). The second conclusion reflects the fact that in the cartel problem increasing the value of output increases the value of monopoly, but equally the gain to deviating, so the two effects cancel out. The third conclusion is the most important one: peer enforced cartels are less likely to be formed when capacity is large relative to the size of the market. If there are many firms and each can easily replace the output of another firm by hiring additional inputs we should not expect to see peer enforced cartels. On the other hand, when the "firms" are individual workers, they are capacity constrained by the hours and intensity with which they work - they cannot simply increase output by going out and hiring additional inputs to increase their output. Hence capacity constraints are more significant in the setting of workers and less binding in the case of firms - which coincides with the observation that peer enforced cartel behavior is less common with firms than with workers.

2.3. Subversion in Politics and Markets

Why are markets harder to subvert than political systems - or to put it differently - why is it easier to overcome the public goods problem in lobbying than in cartel formation? One possibility is that there are legal restrictions: lobbying is legal, cartel formation is not. We defer a discussion of this to the next section. Our theory directs us towards alternative answers - it focuses our attention on the value of the public good V, the monitoring difficulty θ, and the capacity constraint X. We now focus on θ and X and we set aside the role of V for a subsequent discussion in which the value of lobbying is determined by opponents of the industry as well as the proponents.

One possibility raised by the theory is that monitoring of contributions to lobbying efforts is less difficult than monitoring of output. This may be the case, but we suspect is a limited part of the story. It is not immediately obvious, for example, that farmers living in a farm community are less able to observe how many fields their neighbors plant than they are to observe whether their neighbors contribute to farm lobbying efforts. In manufacturing monitoring of prices is difficult, and perhaps even monitoring of outputs. But monitoring of inputs is not so difficult. If manufacturing firms agreed to limit themselves to one six hour shift a day - in respect of workers rights - that would not only be relatively easy to monitor but would be unlikely to violate anti-trust laws.

The key variable uncovered by the previous analysis is the capacity constraint. We find this plays no role in lobbying but is crucial in cartel formation. If large increases in output are possible then the incentives to cheat on a cartel are great and a cartel will not form. If - by contrast - possible output increases are low, the cartel formation problem is relatively similar to the lobbying problem.

We can illustrate our theory by contrasting three industries:

1. Manufacturing firms: it is relatively easy for manufacturers to observe each others activities but firms can easily expand in size by hiring more inputs.

2. Plant workers: it is relatively easy for workers on a factory floor to observe each others effort but workers are physically limited in how much they can increase individual output.

3. Hair dressers: like plant workers hair dressers are physically limited in how much they can increase individual output, but they are diffused in many locations and cannot easily monitor each other. Here we view hair dressers as representative of a class of service workers who are diffused to many locations.

The theory then predicts the pattern given in the table below: manufacturers should be effective at lobbying but not cartelization, plant workers at both, and hair dressers at neither.

industry	monitoring cost	supply elasticity	lobbying	cartel
manufacturing	low	high	yes	no
plant workers	low	low	yes	yes
hair dressers	high	low	no	no

That manufacturers are good at lobbying and better at lobbying than forming cartels is perhaps not so controversial, as is success of unionized workers in both lobbying and cartel operations. What about hair dressers and similar service workers? The U.S. Bureau of Labor Statistics report unionization by different occupational categories: in 2017-2018 only 6.6% of “Personal care and service” workers were unionized, in contrast to 20.2% of “Construction and extraction” workers. Also similar to plant workers, school teachers are heavily concentrated in particular locations, and “Education, training, and library” workers have a 37.2% unionization rate. Further with respect to lobbying, if we examine the lobbying records of the large U.S. state of California, among the top ten we find teachers, various business organizations, and one service employee organization, the California State Council of Service Employees.[3] The latter, according to their website, represents highly concentrated and unionized service workers, not more diffuse groups such as hair dressers.[4] So generally speaking we find the facts in accord with the theory.

2.4. Lobbies, Cartels, and Laws

It is conventional to think of lobbying and cartelization as determined largely by legal restrictions. In this view cartels do not form because they are prohibited by anti-trust law, while lobbying could defanged by laws against campaign contributions and corruption. While we do not doubt that these laws have some effect, in our view laws are endogenous. Our theory of fundamentals does a good job of explaining which industries lobby and form cartels. Laws reflect this reality. Cartelization by workers and lobbying are difficult to prevent because demand for them is high and so they are legal.

Consider first workers who would like to exploit their power. Since the demand for effort is downward sloping, workers as a group can take advantage of their monopsony power by reducing effort - and indeed they often do exactly that. Even without a labor union an informal agreement or social norm not to "work too hard" with social sanctions against those who are overly energetic is common in blue-color settings. Moreover, while we today think of labor unions as encouraged by and supported by the state this has not always been true: historically governments and owners have discouraged unions, often with violence. One of the first known unions were the woolcombers of Florence - the organizer Cinto Brandini was executed in 1345 for his trade union activities (Docherty and van der Velden (2012)). In the early 20th Century in the United States violence against unions was common: as, for example, in the 1927 at the Columbine mine massacre (Zieger (1994)). More recently the Solidarity Union in Poland operated in a hostile political environment. Never-the-less unions have been effective in restricting labor input - indeed there would be very little purpose in murdering union members if the unions were ineffectual. We should note as well: unions are an outstanding example of peer enforcement. The social - and even physical sanctions - taken by colleagues against workers for failing to participate in union activities and strikes are well documented - see, for example, Brinker (1985).

We question as well whether anti-trust laws really matter for large cartels. Most anti-trust activity is directed against small cartels: for example the average number of firms in a cartel pursued by the European Commission is 7.61 (see Ordonez-De-Haro et al. (2018)). Moreover, input restrictions are not so likely to run afoul of anti-trust laws - manufacturing firms can hide collusion as concern over workers rights. In a similar way if farmers got together and talked about colluding to reduce output this would be legally problematic. But if they get together - as they do - to discuss best farming practices and agree that a number of fields should be left fallow, that less fertilizers and less intensive farming is a better practice - and this could be successfully enforced as it is in the case of contributions to lobbying efforts - it seems unlikely it would run afoul of anti-trust policy. Indeed, most governments encourage farmers to discuss and adapt best farming practices - often even subsidize them to do so. Finally, even when cartels are legal, the existing empirical evidence seems to support the idea that large cartels are not very common. For example

[3]https://prd.cdn.sos.ca.gov/lobreport2005/Lobbyist_Report_2005.pdf

[4]http://seiuca.org/about/

Haucap et al. (2010) documents that the median number of members for legal cartels authorized by the German Federal Cartel Office (FCO) between 1958 and 2004 was four.[5]

Turning next to lobbying, it is argued that the corrupting influence of money in U.S. politics comes about because political campaigns are financed by rich lobbies. A common solution is that we need to have public financing of political campaigns so that politicians are not dependent on donations.[6] The problem with this is that the U.S. system of expensive and privately financed political campaigns is relative unique - yet political corruption is by no means limited to the United States. Take Ireland where political campaigns are publicly financed: in September 2008 the Irish finance minister used 64 billion Euros of taxpayer money to bail out banks that - like Goldman Sachs - had made some bad bets. Or take Italy where public financing of political campaigns has been introduced in 1973 and abolished in 1993 with a national referendum in the aftermath of "Tangentopoli" - the biggest investigation on political corruption in the Italian postwar period.[7]

One reason keeping private money out of campaigns is not likely to have much impact on lobbying is that a great deal of corruption is due to appointed or civil service officials and not elected officials. More important: bribing politicians through campaign contributions is only the tip of the iceberg. Now and historically a simple and effective incentive is to give money to the family or to give money after departing office. When he was a Senator, Chris Dodd was famous for carrying the water of the motion picture industry. If the industry wanted the internet shut down so that their films could not be pirated, he was there to fight for them. After he left office in 2011 he took a several million a year job as the CEO of the Motion Picture Association of America. When as a sleek lobbyist Chris Dodd appears in the office of one of his former colleagues, do you suppose the message he brings is "this copyright restriction is good for your constituents for the following reasons?" Or do you suppose his message is "look how rich I am - if you play ball like I did you too can one day be a rich and sleek lobbyist like me?" These issues are not limited to the USA: the Greensill scandal in the UK is one example. How many 31 year old's fresh out of school whose father is not a former President and whose mother is not a former Secretary of State are offered a $600,000 a year job as "special correspondent?" And so forth and so on.

If lobbyists take the long view it is hard to legislate against them: Do we pass a law that anyone who has ever worked in government, is likely ever to work in government or who is related to such a person is unemployable? It is a possible solution - and one that has been tested and proven effective in the past. In Imperial China and in the Ottoman Empire high ranking government officials were castrated male slaves separated from their families at an early age. This solution seems unlikely to be acceptable in the current social environment.

---

[5] Interestingly, the median number of members of illegal cartels in the same period was five.

[6] One particular proponent of this is Larry Lessig. Like one of the authors of this book he became interested in political corruption because of the brutality with which the copyright lobby has pushed aside the public interest. He was a candidate for the Democratic Party's nomination in the 2016 U.S. presidential election running on an anti-corruption platform of limiting campaign contributions, but withdrew before the primaries.

[7] Public financing of political campaigns in Italy has been reintroduced in Italy in 1993 and abolished again in 2013.

CHAPTER 3

The Problem with Pigou[1]

Our basic hypothesis - the Ostrom hypothesis - is that in a stable environment a reasonably homogeneous group of people do a pretty good job of finding a mechanism to deal with public good problems and externalities. In this chapter we ask *what happens when circumstances change unexpectedly.* A particular circumstance we consider is what happens when public goods production is supported by a social mechanism as in the previous chapter and, following best economic advice a Pigouvian subsidy is introduced. Our point of departure is that social mechanisms are costly to implement and operate. We show as a consequence subsidies may cause social incentives to collapse and that this can result in less output of a public good. This is one reflection of the fact that social norms are endogenous. We will consider other consequences of endogeneity and other applications and lessons for policy and for empirical and experimental research.

3.1. Public Good Redux

As before we consider a large organized group with many members. Each member i chooses the amount of public good $x^i \geq 0$ to produce at unit marginal cost. The average output of the group $\overline{x}$ continues to represent a public good benefiting each member. To avoid complication with introducing subsidies we now take the utility value of the public good to be linear and the cost of producing it to be quadratic. Specifically, the utility value of the public good to each group member is given by $V\overline{x}$ with the utility cost of producing the public good given by $(x^i)^2$. Hence a representative group member receives utility

$$V\overline{x} - (x^i)^2.$$

We continue to use the same model of monitoring, so the for the quota φ incentive constraint is $-\varphi^2 - \pi P \geq -\pi_1 P$ giving $P = \varphi^2/(\pi_1 - \pi)$ and the monitoring cost is $M(\varphi) = \theta\varphi^2$ and the objective function $V\varphi - (1+\theta)\varphi^2$ with the optimal quota given by $\hat{\varphi} = (1/2)V/(1+\theta)$.

3.2. Introduction of a Pigouvian Subsidy

We now suppose that a Pigouvian subsidy is introduced, that is an amount σx^i is paid to group members to encourge production. In the absence of a social mechanism individuals would produce $x^i = \sigma$. The group now faces a choice.

(1) Continue with the current quota $\hat{\varphi}$ and punishment $\hat{P} = \hat{\varphi}^2/(\pi_1 - \pi)$. We refer to this as *standing pat.*
(2) Drop the social mechanism entirely, let individuals do as they wish and produce $\overline{x} = (1/2)\sigma$. We refer to this as the *law of the jungle.*

[1]This chapter is based on Dutta et al. (2021).

(3) Design a new social mechanism φ, P – but this is costly especially in the short run.

Alternative #2 the law of the jungle shows in an important way how a group differs from an individual: an individual may either stand pat or reoptimize, but has no equivalent of the law of the jungle. We investigate this first.

Specifically, suppose that the cost of designing and agreeing to a new mechanism is too costly to justify doing so. Hence we ask is it better to stand pat or revert to the law of the jungle? The answer can be seen in figure 3.2.1 below showing how output depends upon the subsidy. As shown by the blue line law of the jungle output $\overline{x} = \sigma$ rises linearly with the subsidy. Suppose instead the original quota is maintained. The subsidy reduces the incentive for group members to decrease output, but as long as the quota $\hat{\varphi} > \sigma$ group members still prefer to undershoot the quota absent punishment, so output remains constant at $\hat{\varphi}$. Once $\sigma > \hat{\varphi}$ they prefer to produce more than the quota and there is nothing to stop them from doing so. Now consider the switchpoint where $\hat{\varphi} = \sigma$. Both standing pat and the law of the jungle result in the same output. However: the original mechanism has associated with it a monitoring cost $M(\hat{\varphi}) = \theta\hat{\varphi}^2 > 0$ that can be avoided by using the law of the jungle. Hence the group strictly prefers the law of the jungle to standing pat at $\sigma = \hat{\varphi}$. It follows that the point of indifference where it is optimal to switch from the original mechanism to the law of the jungle must take place at a lower subsidy as shown by the green vertical line at $\tilde{\varphi} < \hat{\varphi}$. We see then that as the subsidy increases output initially remains constant until $\tilde{\varphi}$ at which point it drops discontinuously and begins to rise again reaching the original output only when $\sigma = \hat{\varphi}$.

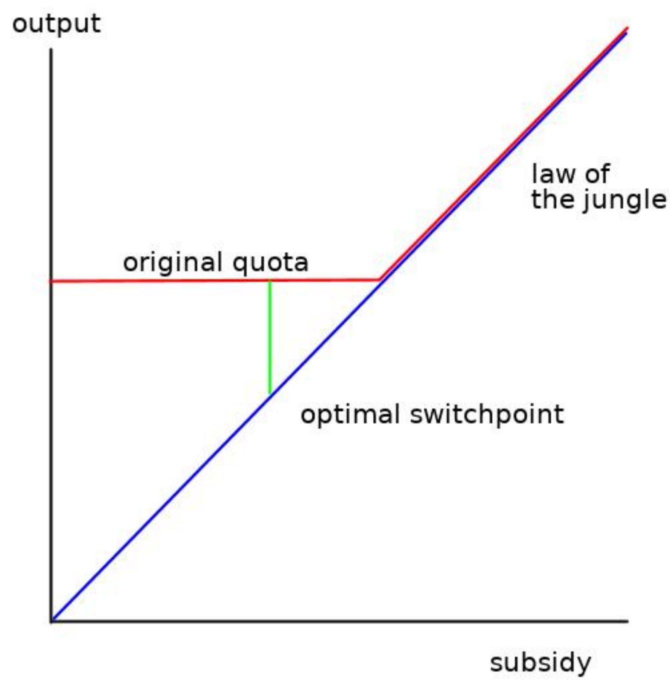
The theory applies equally well to the case where there is a negative output externality rather than a positive one. In the case of a negative externality Pigou will introduce a tax and the social norm will be an upwards quota: do not produce more than the quota. In the case of a negative externality a Pigouvian tax will have the opposite effect of a subsidy in the positive externality case: as the tax increases output will remain flat, then increase before declining again. While it is possible to introduce elaborate notation in order to treat both cases simultaneously - for example, taking the negative of output in the case of a negative externality - this can be confusing, and rather than doing everything twice we have chosen to illustrate the theory for the case of a positive externality.

Next we consider several case studies to see if in response to shocks effective social norms are abandoned for the law of the jungle and if so whether this moves output in the wrong direction.

3.3. Breakdown of Cartels

A key element of our theory is the possibility that in response to an unanticipated change in circumstances a social mechanism may be abandoned in favor of non-cooperative behavior. This can have counter-intuitive consequences: in particular, in the case of a negative externality, an adverse intervention that would ordinarily reduce output might instead increase output. Is there evidence that social mechanisms do break down in response to unanticipated changes? Is this due to bargaining costs? One type of social mechanism that has been extensively studied by economists are cartels.

FIGURE 3.2.1. Optimal Social Mechanisms



As described in the previous chapter our theory applied to cartels differs from those most common in the theory of repeated games. In Green and Porter (1984), Rotemberg and Saloner (1986) or Abreu et al. (1990) price wars are a disciplinary device and are the anticipated consequence of real or apparent cheating. In our account, as in the theoretical and empirical account of Harrington and Skrzypacz (2011), cartel discipline is achieved through modest individual penalties for real or apparent cheating. In the empirical literature our model of cartel breakdown appears to be the more relevant one. Indeed, much of the empirical literature, for example the classical study of sugar cartels by Genesove and Mullin (2001), is devoted to debunking the price war model. As an example we quote from the survey by Levenstein and Suslow (2006): "after the adoption of an international price-fixing agreement in the bromine industry, the response to violations in the agreement was a negotiated punishment, usually a side-payment between firms, rather than the instigation of a price war... As repeatedly discovered by these cartel members, the threat of Cournot reversion is an inefficient way to sustain collusion."[2]

[2]See Chapter 13 for an explanation of why this is so.

In our account, unlike in the repeated game literature, cartel breakdown occurs because of the cost of bargaining in the face of unanticipated changes in circumstances. Again this seems to be the relevant reason for cartel breakdown. Again from Levenstein and Suslow (2006) "Bargaining problems were much more likely to undermine collusion than was secret cheating. Bargaining problems affected virtually every cartel in the sample, ending about one-quarter of the cartel episodes." Their overall conclusion is "cartels break down in some cases because of cheating, but more frequently because of entry, exogenous shocks, and dynamic changes within the industry."

This evidence suggests that social mechanisms do revert to the law of the jungle because of the cost of bargaining in the face of changed circumstances. The literature has not addressed the issue of whether as a result, output increases in response to unanticipated adverse changes. Recently, however, there has been a rather striking natural experiment. In response to the unanticipated reduction in oil demand due to the covid-19 pandemic, OPEC+ attempted to negotiate reduced quotas. On March 8, 2020 bargaining broke down. Subsequently cartel members announced plans instead to increase output, and they did so. The relevant output is reported in the OPEC Monthly Oil Market Reports. During the period December 21, 2019 to March 20, 2020 while the agreement was in effect, and including the period clearly prior to the Covid-19 shock, OPEC output ranged from 27.8 to 28.6 millions of barrels per day. In the following month after the agreement was allowed to expire from March 21 to April 20 OPEC output increased to 30.4 mb/d, a more than 6% increase in output. In brief an unanticipated negative demand shock resulted in a substantial increase in cartel output.

It should be noted that the marginal cost to Saudi Arabia of extracting a barrel of oil (see knoema.com) is estimated to be less than $3 while even with the substantial price fall that took place, the price remained well above $20 so there is no issue here of a price war in the sense of producing below marginal cost.

3.4. The Trouble with Foreign Aid

Our theory shows how subsidies can reduce the provision of a public good. In the case of foreign aid, it is sometimes asserted that subsidies provided by foreign governments and NGOs do exactly this. A good case study is Bano (1973), based on extensive fieldwork in Pakistan complemented by survey data.

Bano (1973) examines public goods that were provided through voluntary efforts with socially provided incentives for contribution. These public goods were primarily welfare related and ranged from health care and education to the defense of political rights. She conducted a detailed study of three organizations, the People's Rights Movement (a political organization), the Edhi Foundation (the largest welfare organization in Pakistan), and the Jamiat ul Uloom al-Shariah, a madrasa that provides a free Islamic education to four hundred students. She documents that volunteers provided public goods not because of altruism or self-signalling but in response to an informal system of social incentives. As in our model this is based on monitoring: examples include informal observation of which ambulance service delivered most frequently, and more formal systems such as the use of receipts to monitor donations. Incentives were social in nature: those who were thought not to pull their weight received less respect and were less likely to be invited to social events such as weddings. As can be seen the narrative fits our model.

Subsequently donor organizations attempted to increase public good provision through subsidies in the form of salaries to contributors. In Bano (1973)'s case studies this led to the unraveling of the provision of social incentives and to decreased provision of the public good. She first documents this for four voluntary organizations. In one case she indicates that "[t]he Maternity and Child Welfare Association... almost collapsed with the influx of such aid." Similarly six community based organizations in Sindh engaging primarily in charity and welfare saw a substantial decrease in provision following the arrival of aid from Oxfam. Finally she discusses the collapse of the Asthan Latif Welfare Trust after the arrival of UNICEF aid. In each case she demonstrates that the reduction in public good provision came about because monitoring and social incentives were abandoned in response to formal incentives and that in the absence of these social incentives volunteer effort dried up.

The bottom line is that Bano (1973)'s evidence fits our model. A public good was provided with social incentives. A subsidy was introduced and the social incentives ended and public good provision declined - as our model predicts.

3.5. Lump Sum Taxes Matter

Net of monitoring costs when there is a subsidy that is paid by outsiders - for example, as in the case of foreign aid - the objective function of the group is

$$V\overline{x} - (x^i)^2 + \sigma x^i.$$

Alternatively it might be that the group must pay the subsidy through lump sum taxes, which results in the objective function

$$V\overline{x} - (x^i)^2 + \sigma x^i - \sigma\overline{x}.$$

This does not change individual incentive not changed, but it does change group incentives.

We consider two consequences of the group paying the subsidy. First, the optimal production of the public good absent free-riding is $V/2$. If the subsidy is set equal to the marginal social value of output $\sigma = V$, the optimal Pigouvian subsidy, then output in the law of the jungle is $\overline{x} = V/2$ and the first best is obtained. Clearly in this case there would be no reason to introduce a social mechanism. By comparison, if the group does not pay the subsidy then the optimal social mechanism will be to encourage overproduction through costly monitoring - not such a desirable result. While individual behavior is not changed by lump sum taxes, it matters to the group and the group will take account of them in designing a social mechanism.

Second, again consider the case where the group pays the subsidy and a subsidy is introduced resulting in a decline in public good output. Conventional analysis suggests this is a failure of policy. It is not: the decline in output (except at the zero measure switchover point) unambiguously represents a welfare improvement. The loss of public good due to the abandonment of social incentives is more than offset by the reduction in monitoring cost: this indeed is why the group reverts to the law of the jungle. The point is that inefficient incentives (costly monitoring) is replaced by an efficient incentive scheme in the form of a subsidy.[3] Exactly the

[3]We are assuming that there is no cost in operating a subsidy - in practice this may not be the case.

same considerations apply in the case of a negative externality and a tax: if the tax is rebated lump sum and results in an increase in output this is an unambituous welfare gain.

3.6. Day Care

An important study documenting a fall in output in response to a tax is the field experiment of Gneezy and Rustichini (2000). They studied the introduction of modest fine for picking up children late at a day-care center. They observed that this resulted in more parents picking up their children late - the opposite of the expected and intended effect. As there was no prior warning or discussion of the fine, it is reasonable to think it was unanticipated. Moreover, as the fine was introduced suddenly and without explanation it might well have been anticipated to be of short duration (as in fact it was) so that it would not be worth renegotiating to identify the re-optimal social norm reducing lateness. Hence our theory predicts the tax were chosen slightly larger than the switching point indeed more parents would pick up their children late.

Authors including Gneezy and Rustichini (2000) and Benabou and Tirole (2006) who have discussed the increased lateness have assumed that this resulted in a drop in welfare. A day-care center, however, is a closed system in which the school is supported by fees from the parents and different schools compete with each other. Implicitly, the money from fines either reduces what parents have to pay, or increases the services they receive. In other words, in this setting it makes sense to assume that the tax is rebated lump sum. If this is the case then the assumption that welfare decreased is wrong: in fact it went up. This highlights the importance of knowing whether social norms are involved and the role of the distribution.

Other theories than ours have been used to explain the increase in lateness: one of the best worked out is that of Benabou and Tirole (2006). Their idea is that in the absence of fines, picking up children on time serves a valuable self-signaling purpose of virtue. With fines, the signaling value of being on time is lowered enough that it becomes worthwhile to be a little late and pay the fine. In contrast in our account prior to the fine there was an informal system of enforcement. Teachers scolded parents who were late and complained to their peers and other parents about people who were persistently late. After the fines were introduced this stopped and parents simply paid their fines. That is, there was punishment before but not after. While this is plausible we do not know whether or not it was the case, and hence we do not have direct evidence about the merits of our theory versus that of Benabou and Tirole (2006). As the welfare analysis for the two theories is opposite it is of importance to know.

There is an important lesson here for the way in which field experiments are conducted. It was possible for Gneezy and Rustichini (2000) to have arranged the experiment to observe punishment before and after. This could have been done by direct observation of teacher behavior at the pickup point - did they scold parents before, but not after? It could also have been done by a before and after survey instrument asking parents and teachers about their expectations of the response to late pickup. In other words: it would be desirable if field experiments where social norms might be involved attempted to ascertain the presence of informal punishments and if this was changed by intervention.

In existing analyses an upward jump in output in response to a Pigouvian tax is regarded as a failure of policy. The goal of the policy is to reduce output in the face of an externality. But that analysis may miss the mark. If there are informal punishments and taxes are rebated lump sum, increased output is an indication that the policy has a desirable effect. While the increase in output has a negative consequence for welfare, overall welfare goes up because by switching to the law of the jungle the cost of monitoring is avoided and this more than makes up for the loss from increased output.

3.7. Persistence is Important

As indicated there are psychological theories that predict reduction in output in response to subsidies. As mentioned the self-signaling theory of Benabou and Tirole (2006) is one such. Other theories involve the subsidy signaling the "importance" of the externality. A small subsidy may lead people to conclude the externality was less important than they thought and they may then reduce their output.

Neither of these theories are of great use in explaining either the behavior of OPEC or what happened with NGOs in Pakistan. Never-the-less the theories are not so obviously wrong in the case of fines for picking up late at school, or other circumstances where people voluntarily contribute to good causes. One thing the psychological theories have in common is that they suggest that output reductions are likely to persist. This is not the case with social norms.

If the benefit of a new social norm is low - the extreme case being one in which lump sum subsidies are paid by the group and the subsidy it the optimal Pigouvian one - then there will be little reason to introduce a new social mechanism and indeed the reduction may indeed persist. However, when the gains are greater, it is generally easier to negotiate and find a new mechanism over a longer period of time. In this case we expect that eventually a new social mechanism may be introduced and output may increase to exceed the original level. In other words, the initial drop in output may not persist. Indeed: this is exactly what happened with OPEC: after a few months they reached a new agreement and (as the externality was negative) output dropped below pre-Covid levels

3.8. Milking the Cow

We now consider some further implications of lump sum taxes, or more specifically, their absence. Consider the case of a negative externality in which revenue from the tax is taken by an outside agency imposing the tax. We know that if the tax is low enough behavior will not change: the original quota will be maintained and binding. Hence the only effect of the tax is to enrich the outside agency at the expense of the group - in the French vernacular, milking the cow. In such a case the policy maker may claim to be Pigouvian but but it is unlikely that anyone will believe them.

This is an especially important consideration because if the group decides to negotiate and organize a new social mechanism they have options not available to individuals. In particular, they can engage in political action, including rebellion, protest, and tax repeal.

An interesting example of a group responding to the naive imposition of Pigouvian taxes by engaging in tax repeal is the case of the French "yellow vests." In this instance output is driving speed while the tax was imposed by lowering the speed limit. Specifically, on July 1, 2018 the French Federal Government lowered the speed limit on secondary highways from 90 km/h to 80 km/h ostensibly to reduce highway accidents. The bulk of the impact fell on rural communities where there are no primary highways. Although driving is to an extent anonymous, there are informal social norms, and drivers who are perceived to drive excessively fast are often punished.[4] Moreover revenue from the speed camera revenue is not returned to rural drivers or communities. The cost of organization appears to have been fairly low due to the advent of social media: Facebook played a key role in the organization of the yellow vests. Hence our theory says that if they could do so at low cost they would organize not only a new driving speed norm, but also eliminate the tax. In fact the yellow vests did act to "repeal" the tax. The rate of traffic camera destruction jumped by 400% and in the year following about 75% of all traffic cameras in France were destroyed.[5]

3.9. Aggregate Output Subsidies[6]

Pigouvian subsidies are generally aimed at individuals. This does not have to be the case. Rather than paying the salary of employees as NGOs did in Pakistan they could provide a health charity with medical training and equipment such as ambulances. In other words, instead of a subsidy $\sigma > 0$ resulting in a group objective function

$$V\overline{x} - (x^i)^2 + \sigma x^i$$

they could introduce a subsidy $\gamma > 0$ on output resulting a group objective function

$$V(1+\gamma)\overline{x} - (x^i)^2.$$

An aggregate subsidy of this type has no implication for individual behavior and in standard theory would probably be regarded as a waste - resulting in no output. In our setting of social norms the situation is quite different: such a subsidy does not substitute for a social mechanism as does a Pigouvian one. In particular, γ can never cause reversion to the law of the jungle, and can only increase output. This suggests that such a subsidy has advantages over the standard Pigouvian one.

Is it really true that the Pigouvian subsidy reduces output while the aggregate subsidy increases it? The US government recently tested the theory. In Afghanistan referring to the total collapse of the Afghan army - the lack of provision of the public good of national defense - according to President Biden "We paid their salaries...What we could not provide was the will to fight..." We would say "they did not have the will to fight because we paid their salaries." By contrast in Ukraine the aggregate subsidy - the provision of military equipment and training - does not appear to have undermined the will of the Ukrainians to fight.

3.10. Scaling

It is not uncommon these days to carry out a field experiment or randomized control trial and if is successful suggest that it be adopted on a wide scale. The fact is that a small scale - and temporary - experiment is unlikely to result in the

[4]While fictional, the Damián Szifron film "Relatos Salvajes" illustrates the idea well.

[5]Private communication from Pierre Boyer. Our account is based on Boyer et al. (2019) who documents both the link between the change in speed limit and the yellow vest movement, as well as the systematic way in which that group organized itself.

[6]Based on Dutta et al. (2022).

development of new social norms. A wide scale adoption on a permanent basis is more likely to result in changes in social norms - and this may have unanticipated consequences. A story from the 19th century should be cautionary.

In the early 19th century Britain had great success in extracting trade concessions from China - free trade in opium primarily. It did so by sending a few gunboats and either threatening to blow stuff up or actually blowing it up. China is a big country and these attacks while disliked were essentially pinpricks and had little impact on social norms or institutions.

Inspired, perhaps, by this example the United States in 1853 sent gunboats under the command of Commodore Perry to the much smaller Japan to extract trade concessions. While equally as successful as the British this policy intervention did change social norms and dramatically so. As a direct consequence of the Commodore Perry's opening of Japan was the Meiji restoration.[7] During the next 40 years social norms and institutions in Japan changed in every dimension from the form of government to the mode of dress. This change was at substantial cost as it involved a civil war. As a consequence Japan changed from a medieval peasant economy to a modern industrial economy.

Fifty one years after Perry's policy intervention when Russia sent a great number of gunboats - an entire Navy in fact - it met a rather different fate than Commodore Perry's gunboats: the Russian ships were sent directly to the bottom of the sea by the Japanese Navy. Indeed from an American perspective it is hard to see how the Japanese attack on Pearl Harbor 89 years after Perry's intervention could have occurred without the dramatic change in norms it occasioned. And if 89 years seems a long time to worry about policy consequences, bear in mind that most of the bad effects of climate change are predicted to occur in this timeframe.

3.11. How Sticky are Social Norms?

There is a fundamental puzzle in the empirical literature on the political economy of culture and institutions. It concerns the persistence of dysfunctional cultures. On the one hand there is a substantial literature indicating that these can be quite persistent. Acemoglu and Robinson (2001) give evidence for persistence on the order of four centuries. Bigoni et al. (2013) have evidence of a similar effect in Italy over nearly nine centuries and Belloc et al. (2016) point to persistence In Italy that also lasts centuries. Dell et al. (2018) have highly persuasive evidence for persistence in Vietnam on the order of a century and a half.

On the other hand it cannot be that it is simply impractical to change social and cultural norms: side by side with the survival of dysfunctional norms we see abrupt change over periods of a few decades. The most dramatic example is that of Japan occasioned by Commodore Perry's intervention. This is not a unique example. With respect to social norms - presumably more mutable than cultural norms - three cases make the point:[8] the rapid change in social norms (measured in minutes) concerning the treatment of airplane hijackers that took place on September 11, 2001; the change in social norms (measured in months) concerning public protest that took place in East Germany following the commitment by Gorbachev that military intervention in East Europe was off the table; and the rapid and organized

[7]See, for example, Jansen (2002).

[8]See the discussion in Levine (2012) for details of these three cases.

change in social norms (following a debate that lasted over 12 years) that took place in Sweden when the change was made from left-side to right-side of road driving.

Cultural norms are broader and deeper than social norms such as which side of the road on which to drive. Two central aspects of culture are religion and language. Yet we observe that even these fundamental aspects of society change over short periods of time. Prior to 1990 the country of Ireland could well be described as Catholic. Yet by the end of the decade the church lost its central place in Irish life and the country could be better described as secular.[9] With respect to language we may point to the remarkable example of Hebrew. In 1880 Hebrew was not a conversational language. In 1903 there were perhaps a few hundred Hebrew speakers. Within fifteen years more than 30,000 Jews in Palestine claimed Hebrew as their native language.[10]

Our basic hypothesis is that the group designs an incentive compatible mechanism for itself that is mutually beneficial for members. That is, we do not assume that social norms are left-over from some past meaningful equilibrium - we reckon that groups choose them and change them in response to changed circumstances. This well explains why norms sometimes respond quickly but seems inconsistent with the idea that they sometimes change slowly. In thinking about possible ways of reconciling slow and fast change it is important first to realize that when change was rapid the incentives for change were large. Second, in at least in some cases where change was slow (in Italy) the dysfunctional social norms involved poor treatment of outsiders became dysfunctional because it inhibits trade and investment. This suggests two complementary theories of why change may sometimes be fast and sometimes slow. The fact that large incentives do seem to lead to rapid change suggests that adjustment costs may play an important role: only if the benefit of change exceeds the cost of adjustment would we observe a change. That is the avenue explored in this chapter.

The fact that outsiders may play a role suggests a possible reputational model in which a bad reputation once acquired may be hard to lose. We examine this possibility in Chapter 15. Finally, it may be that different social or cultural norms are more resistant to change than others: for example, market institutions that are highly decentralized may be harder to change than centralized institutions where planning and consensus are the norm.[11] This idea must await future research because we do not currently have a good theory of why some institutions maybe be harder to change than others.

It is trite to observe that if their is a cost of changing social norms then they will remain the same until the benefit of reoptimizing exceeds the cost. This, in some sense, is the message of the menu cost model in macroeconomics following Calvo (1983). In this chapter our basic presumption is that reverting to the non-cooperative norm is costless while designing a new social norm is costly. Reverting to the law of the jungle is a decentralized decision: if it is evident that the non-cooperative social norm is superior to the alternatives there is no need to get together to discuss this and reach an agreement, implicitly everyone has agreed in advance that in this case they will all go their own way. By contrast developing

[9] See, for example, Donnelly and Inglis (2010).

[10] See, for example, Bar-Adon (1972).

[11] We are grateful to Melissa Dell for raising this point in the context of the Vietnam villages studied in Dell et al. (2018).

a new social norm cannot be decentralized and the group must be reconvened to agree upon a new social norm.[12]

Of course the question arises whether the costs of changing a menu is really so great that it does not pay to raise prices. A similar question arises with respect to changing social norms. In one case we have an idea about the rough benefit of the change: Dell et al. (2018) estimate that if Khmer villages were to switch to Dai Viet institutions it would lead to roughly a 33% income equivalent improvement. However, a third of that gain is due to human capital differences, so the medium term gain would only be about 20%. It seems plausible that the cost of the substantial institutional change involved could be this great.

[12]It is important here that change be unanticipated: otherwise social norms should be designed with contingency planning. Of course it is costly to do so for infrequent events. This follows the literature on incomplete contracting such as Hart and Moore (1988) and rational inattention such as Sims (2003). Our formal model is similar to those of unawareness as in Modica and Rustichini (1994) and in the spirit of Tirole (2009) and Dye (1985) or costly contemplation such as Ergin and Sarver (2010).

CHAPTER 4

Binary Choice: Participation with Hidden Cost[1]

In the political arena individual effort is often indivisible: for example, either to vote or not to vote, or whether to participate in a demonstration or not. In lobbying effort may be more continuous, but often the group asks for a fixed levy of time, effort, or money. However, even when group members are *ex ante* identical, typically people will face different participation costs at the time the participation decision is made. To take an example: it may be that on election day a party member is in the hospital and so it is very costly for the member to vote that day. In Chapter 2 we examined a model of common cost in which output was continuous but private. Here we turn to a model of private cost in which output is discrete (binary in fact) and public.

We adopt the standard model, for example the Palfrey and Rosenthal (1985) model of voter turnout. Here group members independently draw types y^i uniformly distributed on $[0, 1]$ and may contribute zero effort at zero cost (not participate) or contribute a single unit of effort (participate). The cost of participation is $c(y^i)$, where we assume that types are ordered so that this is a non-decreasing function: higher types have higher cost. Furthermore, we assume that cost is linear $c(y_i) = c_0 + y^i$. We allow c_0 to be either positive or negative, and that sign plays a key role in our analysis and will be the subject of subsequent discussion.

In this setting a social norm is a threshold φ for participation: those types with $y^i < \varphi$ are expected to participate and those with $y^i > \varphi$ are not. If the social norm is followed, the expected fraction of the group that will participate is φ and in a large group we may assume that since we are averaging over many independent draws the realized participation fraction is approximately equal to its expected value.

The action of a member, whether she has participated or not, is observable by everyone, but for those who did not participate there is only a noisy signal of their type $z^i \in \{0, 1\}$, where 0 means “good, was not supposed to participate, so followed the social norm” and 1 means “bad, was supposed to participate, so did not follow the social norm.” Specifically, if the social norm was violated, that is the member did not participate but $y^i < \varphi$, the bad signal is generated with probability π_1; if i did not participate but $y^i > \varphi$ so that she did in fact follow the norm, there is nevertheless a chance π of the bad signal where $\pi \leq \pi_1$. A bad signal is punished with a utility cost of P.

[1]This chapter and the next are based on Levine and Mattozzi (2020) and presentations in 2016 at the Young Economists Workshop and the Maastricht Games Conference, in 2017 at Royal Holloway, at a public lecture at Washington University in St. Louis, at the Robert Schuman Centre for Advanced Studies of the European University Institute, and at the University of the Pacific (Lima), and in 2020 at Erasmus.

4.1. Direct Cost and Committed Members

The total direct cost of a participation rate of φ is $\int_0^\varphi c(y)dy$. Recall that $c(y_i) = c_0 + y^i$ and that we allow $c_0 < 0$. In this case members with types $c(y_i) < 0$ that is $y^i < -c_0$ derive a benefit from participation - why this might be we discuss later. We call them *committed members* - those who find it individually optimal to participate because they have negative participation costs. As we are dealing with the production of public goods rather than public bads it would make no sense for the group to deprive these members of the benefit of participation - and indeed no incentive need be provided to them. We call $\underline{\varphi}$ the fraction of committed members; so if $c_0 + 1 < 0$ - everybody enjoys participation - then $\underline{\varphi} = 1$; if $c_0 > 0$ - everybody has positive participation cost - then $\underline{\varphi} = 0$; otherwise, if $-1 \leq c_0 \leq 0$, we have $\underline{\varphi} = \max\{y \mid c(y) \leq 0\} = -c_0$. Just as in the case of cartels where we ruled out $\varphi > 1$ because it would be senseless to generate negative profits, here we restrict attention to norms $\varphi \geq \underline{\varphi}$ as it would be senseless to deny committed members the opportunity to participate. To avoid trivialities we assume that the participation cost of the most reluctant member is non-negative, that is $c(1) = c_0 + 1 \geq 0$, otherwise incentives are not a problem for the group. With this in mind we define (incremental) *direct cost* $D(\varphi) = \int_{\underline{\varphi}}^\varphi c(y)dy$ for $\varphi \geq \underline{\varphi}$ and $D(\varphi) = 0$ for $\varphi < \underline{\varphi}$.

4.2. Monitoring Cost and Incentive Compatibility

What is the cost of inducing participation $\varphi > \underline{\varphi}$? The direct cost is $D(\varphi)$. However, members with $c(y^i) > 0$ and $y^i \leq \varphi$ must be given incentives for participation. The *monitoring cost* of doing so is the cost of punishing the innocent $y^i > \varphi$: this is $M(\varphi) = \int_\varphi^1 \pi P dy$. The (incremental) *total cost* is then $C(\varphi) = D(\varphi) + M(\varphi)$.

THEOREM 4.2.1. *$D(\underline{\varphi}) = M(\underline{\varphi}) = C(\underline{\varphi}) = 0$. When $\varphi > \underline{\varphi}$, the optimal incentive compatible punishment is $\hat{P} = c(\varphi)/\pi_1$. Defining the* monitoring difficulty *as $\theta \equiv \pi/\pi_1$ and the* fixed cost *F and* marginal cost *γ by*

$$F \equiv \max\{0, \theta c_0\} \quad \textit{and} \quad \gamma \equiv [(1/\theta) - 1]F + \theta(1 - \underline{\varphi}) \geq 0,$$

then the (incremental) costs are:

$$\begin{aligned} D(\varphi) &= (F/\theta)(\varphi - \underline{\varphi}) + (1/2)(\varphi - \underline{\varphi})^2 \\ M(\varphi) &= F + (\theta(1 - \underline{\varphi}) - F)(\varphi - \underline{\varphi}) - \theta(\varphi - \underline{\varphi})^2 \\ C(\varphi) &= F + \gamma(\varphi - \underline{\varphi}) + (1/2)(1 - 2\theta)(\varphi - \underline{\varphi})^2. \end{aligned}$$

PROOF. Take first the case $\varphi = \underline{\varphi}$. In this case there is no $y^i > \underline{\varphi}$ who fails to participate, since nobody is supposed to. Consequently $M(\underline{\varphi}) = 0$. Since $D(\underline{\varphi}) = 0$ as well, the total cost is $C(\underline{\varphi}) = 0$.

Now let $\varphi > \underline{\varphi}$. Notice that $\gamma \geq 0$ follows from $\theta = \pi/\pi_1 \leq 1$. First, consider members with positive costs $c(y^i) > 0$ who are supposed to participate, that is, for whom $\underline{\varphi} < y^i < \varphi$. If they participate they pay the cost $c(y^i)$, if they fail to do so they pay the punishment $\pi_1 P$. Hence they are willing to participate if and only if $c(y^i) \leq \pi_1 P$. If $c(\varphi) \leq \pi_1 P$ then the social norm φ is incentive compatible for those who are supposed to participate.

Second consider members with positive costs $c(y^i) > 0$ who are not supposed to participate, that is, for whom $y^i > \varphi$. If they participate they pay the cost $c(y^i)$, if they fail to do so they pay the punishment πP. Hence they are willing not to participate if and only if $c(y^i) \geq \pi P$. If $c(\varphi) \geq \pi P$ then the social norm φ is incentive compatible for them.

As cost is increasing in P we should minimize P subject to the constraint $\pi P \leq c(\varphi) \leq \pi_1 P$ that is $c(\varphi)/\pi_1 \leq P \leq c(\varphi)/\pi$. Since it is the lower constraint that binds, we find $\hat{P}$ as indicated.

We now compute $D(\varphi), M(\varphi)$ and $C(\varphi)$. By integration we get

$$D(\varphi) = \int_{\underline{\varphi}}^{\varphi} c(y)dy = c_0(\varphi - \underline{\varphi}) + (1/2)(\varphi^2 - \underline{\varphi}^2)$$

$$M(\varphi) = \int_{\varphi}^{1} \pi \hat{P} dy = \theta(1-\varphi)c(\varphi) = \theta(1-\varphi)(\varphi - \underline{\varphi} + \max\{0, c_0\})$$

where in the last expression we have used the fact that $c_0 \geq -1$ implies $c_0 + \underline{\varphi} = \max\{0, c_0\}$. The remainder of the proof is algebra given in the Appendix 4.4 to this chapter. □

To discuss the implications of this result we make the following observation. Recall that a concave function has declining average cost, while a convex function has increasing average cost. Since a smaller group must have a greater participation rate than a larger group in order to mobilize the same number of members, this means that when a small group mobilizes as many members as a large group it will face lower average cost when cost is concave and higher average cost when cost is convex. In this sense *concavity of cost favors the small while convexity of cost favors the large*. We will make this intuition precise in the next chapter when we study political contests between small and large groups.

More broadly, for $\theta \geq 0$ the implications of Theorem 4.2.1 concerning the crucial issue of concavity are:

- In the range $\varphi \geq \underline{\varphi}$ direct cost is always convex, while monitoring cost is always concave.
- In the range $\varphi \geq \underline{\varphi}$ total cost is convex if $\theta < 1/2$ and concave if $\theta > 1/2$.
- Total cost is globally convex if $F = 0$ (so $\underline{\varphi} \geq 0$) and $\theta \leq 1/2$ and globally concave if $F \geq 0$ (so $\underline{\varphi} = 0$) and $\theta \geq 1/2$.

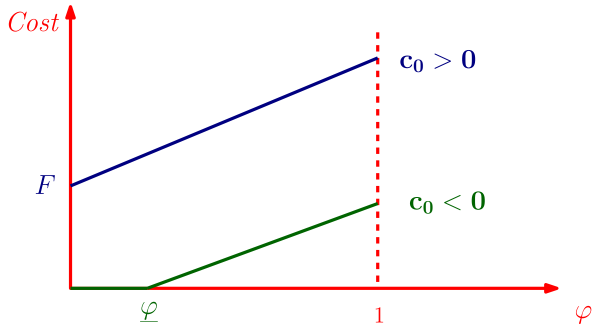
In light of this, the overall take away is that *committed members and low monitoring difficulty favor large groups while positive fixed costs and high monitoring difficulty favor small groups*. A case of particular interest is when monitoring difficulty is such that $\theta = 1/2$ and so incremental cost is linear. In this case we have a simple relationship between c_0 and convexity and concavity: $c_0 < 0$ corresponds to convex costs, while $c_0 > 0$ corresponds to concave costs. This is illustrated in Figure 4.2.1.

In this case total cost is linear and convexity occurs exactly when there are committed members and concavity exactly when there are positive fixed costs. We refer to this important special case as the *linear case.*

4.3. The Intuition of Cost

As the concavity or convexity of cost plays a crucial role in our analysis it is important to have a strong intuition about its source. The overall shape of costs has two parts: the concavity or convexity above $\underline{\varphi}$, which depends on the relative strength of (convex) direct costs and (concave) monitoring costs, and the presence or absence of fixed costs and committed members.

FIGURE 4.2.1. Total Cost of Participation



In the case of $c_0 > 0$ the cost is $F + \gamma\varphi$ for $\varphi > 0$. In the case of $c_0 < 0$, $F = 0$ and the cost is zero up to $\underline{\varphi}$ and then it starts increasing with slope γ.

Why are direct costs convex? Because the group draws first on low cost types, bringing in only higher cost types when higher participation is needed. As the low cost types enter first, there is a disproportionate increase in cost as participation increases. This is the usual intuition behind diminishing returns.

Why are monitoring costs concave? As we have seen, if only committed types participate there is no monitoring cost. However, if all types participate, that is $\varphi = 1$, then there are no non-participants who need to be punished so there are also no monitoring costs. Intuitively a function which is zero at both ends and positive in between is "concavish."

Clearly group members have different participation costs - for example, some members might have a compelling reason to be out of town when there is a vote or demonstration - and it seems intuitive that these costs should be positive for all members. Participation surely involves time and effort and indeed, in the case of lobbying we believe this is so. However, this does not need to be the case in the case of voting. In fact, two non-strategic reasons for voting given by political scientists are civic duty and expressive voting. Both can be interpreted as a negative cost of voting, although for different reasons. Civic duty means voting out of a sense of obligation to society. Expressive voting is more akin to low stakes sport betting on a favored team: a way to show solidarity with or support for a favored candidate cause - not out of expectation of winning but as a symbolic gesture. In both cases the group member (voter) derives a benefit from participating - and this may more than offset the cost of the time and effort required to vote.[2] Hence the possibility of committed members (voters).

While the presence of committed members favoring convexity when $c_0 < 0$ is clear enough, the role of the fixed cost $F = \max\{0, \theta c_0\}$ is less intuitive. Notice first the key fact that if $\theta = 0$ - so monitoring is easy - then there is no fixed cost. But since mobilizing nobody costs nothing, why should mobilizing even a few involve a

[2]Note that voting is not the only type of participation in which there may be a benefit as well as a cost. Attending a political rally on a beautiful spring day (protests tend to take place in nice weather) can be an enjoyable experience, and indeed can be an opportunity not only to enjoy friends but to chant, march, sing and indeed to meet new like-minded people as well. For others the opportunity to participate in a violent rally and physically attack evil-doers with different beliefs may be an enjoyable experience.

discrete increase in cost? The answer lies in the fact that to give incentives to even a few participants with positive costs to participate is it necessary to punish many voters who fail to participate. That is, when $\pi > 0$ everyone has to be punished to provide incentives at all. This is a second dimension in which monitoring cost is favorable to the small.

To capture these ideas we refer to the case $c_0 < 0$ as a *duty* - reflecting the idea that voters feel a duty to vote and consequently get utility from discharging that duty. By contrast we will refer to the case $c_0 > 0$ as a *chore* - the lobbying arena.

In the "conventional" analysis of voting in Feddersen and Sandroni (2006) or Coate and Conlin (2004a) in which voters are ethical or rule utilitarian there are no monitoring costs, voters do the "right thing" out of ethical concerns and require no incentives. This is formally equivalent to $\theta = 0$ in our model. In this case the fixed cost is zero and there is only direct cost which is convex so overall cost is necessarily globally convex. This theory has a great deal of difficulty in explaining either why small groups sometimes do well in elections - and very well at lobbying. As we shall see adding monitoring cost to the mix greatly increases the explanatory power of the theory.

4.4. Appendix: The Algebra of Participation Cost

We consider the more general technology $c(y_i) = c_0 + c_1 y^i$ because we will have $c_1 \neq 1$ later on. Define as before $F \equiv \max\{0, \theta c_0\}$ and observe that $\underline{\varphi} = \max\{0, -c_0/c_1\}$. Finally, define

$$\gamma \equiv [(1/\theta) - 1]F + \theta c_1(1 - \underline{\varphi}).$$

LEMMA 4.4.1. *We have* $D(\varphi) = (F/\theta)(\varphi - \underline{\varphi}) + c_1(1/2)(\varphi - \underline{\varphi})^2$, $M(\varphi) = F + (\theta c_1(1-\underline{\varphi}) - F)(\varphi - \underline{\varphi}) - \theta c_1(\varphi - \underline{\varphi})^2$, *and* $C(\varphi) = F + \gamma(\varphi - \underline{\varphi}) + c_1(1/2)(1 - 2\theta)(\varphi - \underline{\varphi})^2$.

PROOF. By integration

$$\begin{aligned}
D(\varphi) &= \int_{\underline{\varphi}}^{\varphi} c(y)dy = c_0(\varphi - \underline{\varphi}) + c_1(1/2)(\varphi^2 - \underline{\varphi}^2) \\
&= c_0(\varphi - \underline{\varphi}) + c_1(1/2)(\varphi^2 - \underline{\varphi}^2) + c_1(1/2)(\varphi - \underline{\varphi})^2 - c_1(1/2)(\varphi^2 - 2\underline{\varphi}\varphi + \underline{\varphi}^2) \\
&= c_0(\varphi - \underline{\varphi}) - c_1\underline{\varphi}^2 + c_1(1/2)(\varphi - \underline{\varphi})^2 + c_1\underline{\varphi}\varphi.
\end{aligned}$$

From $c_0\underline{\varphi} = c_0 \max\{0, -c_0/c_1\} = -c_1\underline{\varphi}^2$

$$D(\varphi) = (c_0 + c_1\underline{\varphi})\varphi + c_1(1/2)(\varphi - \underline{\varphi})^2.$$

From $c_0 + c_1\underline{\varphi} = c_0 + \max\{0, -c_0\} = \max\{c_0, 0\} = F/\theta$

$$D(\varphi) = (F/\theta)\varphi + c_1(1/2)(\varphi - \underline{\varphi})^2$$

and from $F\underline{\varphi} = 0$ this is

$$D(\varphi) = (F/\theta)(\varphi - \underline{\varphi}) + c_1(1/2)(\varphi - \underline{\varphi})^2.$$

This is the expression for $D(\varphi)$ in the Lemma.

From $\hat{P} = c(\varphi)/\pi_1$

$$\begin{aligned} M(\varphi) &= \int_{\varphi}^{1} \pi \hat{P} dy = \theta(1-\varphi)c(\varphi) = \theta(1-\varphi)(c_0 + c_1\varphi) \\ &= \theta(1-\varphi)(c_1(\varphi - \underline{\varphi}) + \max\{0, c_0\}) \\ &= F(1-\varphi) + \theta c_1(1-\varphi)(\varphi - \underline{\varphi}) \\ &= F(1-\varphi) + \theta c_1(\varphi - \underline{\varphi}) - \theta c_1 \varphi(\varphi - \underline{\varphi}). \end{aligned}$$

Using $F\underline{\varphi} = 0$

$$\begin{aligned} M(\varphi) &= F - F(\varphi - \underline{\varphi}) + \theta c_1(\varphi - \underline{\varphi}) - \theta c_1 \varphi(\varphi - \underline{\varphi}) \\ &= F + (\theta c_1 - F)(\varphi - \underline{\varphi}) - \theta c_1(\varphi - \underline{\varphi})^2 - \theta c_1 \underline{\varphi}(\varphi - \underline{\varphi}) \\ &= F + (\theta c_1 - F - \theta c_1 \underline{\varphi})(\varphi - \underline{\varphi}) - \theta c_1(\varphi - \underline{\varphi})^2 \\ &= F + (\theta c_1(1 - \underline{\varphi}) - F)(\varphi - \underline{\varphi}) - \theta c_1(\varphi - \underline{\varphi})^2. \end{aligned}$$

This is the expression for $M(\varphi)$ in the Lemma.

We now compute total cost $C(\varphi) = D(\varphi) + M(\varphi)$ as

$$\begin{aligned} C(\varphi) = (F/\theta)(\varphi - \underline{\varphi}) + c_1(1/2)(\varphi - \underline{\varphi})^2 + F \\ +(\theta c_1(1 - \underline{\varphi}) - F)(\varphi - \underline{\varphi}) - \theta c_1(\varphi - \underline{\varphi})^2. \end{aligned}$$

Collecting terms this is

$$C(\varphi_k) = F + \left([(1/\theta) - 1]F + \theta c_1(1 - \underline{\varphi})\right)(\varphi - \underline{\varphi}) + c_1(1/2)(1 - 2\theta)(\varphi - \underline{\varphi})^2$$

and the final result of the Lemma follows from the definition $\gamma \equiv [(1/\theta) - 1]F + \theta c_1(1 - \underline{\varphi})$. $\square$

CHAPTER 5

Competing for a Prize: Political Auctions

Economics is focused on mutual gains to trade: you have a banana that I want, I have an apple you want, we trade and are both better off. Politics is different. If we raise taxes to pay subsidies to farmers it doesn't make both farmers and urbanites better off: it is money out of the pocket of the urbanite and in the pocket of the farmer - with a little money falling along the side of the road. That is: a great deal of politics is not about trade but is about transfers between groups, hence conflict.

Political conflict is complicated. Different groups compete - lobbying groups, trade unions, political parties - and provide effort in the form of money for bribes and advertising, votes, time, demonstrations, strikes, internet activity and so forth and so on. What they are competing for is complex: a party may lose a national election but increase the number of regional governments it controls; legislation may be passed into law with more or less favorable amendments. On the grounds that it is better to walk before you can run, we are going to start with the simplest case.

To make things concrete, think of a country like Greece where the political party that wins the election gets a lot of government jobs to reward its followers. There are just two groups, the large L and the small S. The government jobs are worth V - this is the prize. The key to understanding elections - as every political scientist knows - is turnout. Polls do a good job predicting how people are going to vote. When we see an unexpected outcome like Brexit or Trump it isn't because the polls were wrong in predicting how people were going to vote - it is because *polls do a poor job of predicting whether people are going to vote or not.* A good example of this is the Spanish national election that took place on March 14, 2004. The incumbent People's Party (*Partido Popular*) was favored to win by around 6 percent. However, three days before the election 191 people were killed in a terrorist bomb attack on four commuter trains approaching the Atocha station. The People's Party responded to the attack by lying and blaming the attack falsely on Basque terrorists despite the evidence that it was conducted by al-Qaeda. Three days after the attacks the election was held and furious voters voted the People's Party out of office. What happened? Did People's Party supporters vote for other parties? No. What happened is that opposition voters turned out in much greater than expected numbers.

It makes sense, then, to use our model of groups with a fixed set of members who support the party. Let's say that relative size of the two groups is $\eta_L > \eta_S > 0$ with $\eta_L + \eta_S = 1$. Participation by group $k \in \{L, S\}$ is the fraction φ_k of its members it sends to the polls, where $0 \leq \varphi_k \leq 1$. We assume both groups have the same monitoring technology so that the cost of participation is given by $C(\varphi_k)$. We have now to account for the groups different sizes, so we scale costs: to bring a total of $\eta_k \varphi_k$ voters to the polls costs $\eta_k C(\varphi_k)$. The group that has the most participants (voters) wins the prize, and if there is a tie the prize is split. This model of parties that win a prize by sending voters to the polls is basically that proposed by Shachar and Nalebuff (1999).

Seen this way the election is a game between two players - the parties. Party k wins if it gets the most votes: $\varphi_k \eta_k > \varphi_{-k} \eta_{-k}$. The payoff to party k for winning is $V - \eta_k C(\varphi_k)$ while it gets $-\eta_k C(\varphi_k)$ if it loses and $V/2 - \eta_k C(\varphi_k)$ if there is a tie. This is a model of competition that economists are familiar with - it is called the *all-pay auction*. We can think of the number of members who participate $b_k = \varphi_k \eta_k$ as a *bid* - and the highest bidder wins the prize. It is, however, not a standard winner-pays form of auction - it is called an all-pay auction because you have to pay your bid even if you don't win. In fact, the cost of turning out members is sunk no matter what the outcome of elections is.

Notice - and this is one of the strengths of game theory and of the beauty of simple formal models - that the same game could be a lobbying game. That is, two lobbying groups might compete over a piece of legislation by bribing a politician. Successfully getting your own agenda passed into law is worth V to the winning group, and now each group member is endowed with a unit of resources, and φ_k represents the fraction of members of group k contributing to bribe the politician. This general conceptual framework in which several groups compete in a game for a prize by providing effort is the workhorse model political economists use to study voting, lobbying and other political conflicts including warfare. However, while in the case of voting it makes sense that both parties expend effort voting, in the case of lobbying it is more likely that the politician collects only from the winning group. That is - in the case of lobbying the auction might be an ordinary winner-pays auction rather than an all-pay auction. So here we have one possible difference between voting and lobbying - perhaps small groups are more effective in winner-pays auctions and large groups more effective in all-pay auctions. But as we shall see it is not the form of auction that matters.

5.1. The All-Pay Auction

To explain how the all-pay auction works we are going to start with the simplest case: the linear case $\theta = 1/2$ together with the neutral assumption that participation is neither a duty nor a chore $c_0 = 0$.[1] This implies $C(\varphi_k) = \gamma \varphi_k$ so that $\eta_k C(\varphi_k) = \gamma b_k$, and if we ignore the fact that the groups are constrained by $\varphi_k \leq 1$ we have the classical all-pay auction in which cost is proportional to the bid, first studied by Hillman and Riley (1989). They showed there is a unique Nash equilibrium in which each party chooses an optimal turnout given the turnout of the other party. This equilibrium has two key characteristics:

- The equilibrium is not in pure strategies so the outcome of the contest is necessarily unpredictable.
- The large group never does worse than the small group in expected utility and sometimes does better, with higher stakes favoring the large group.

This makes sense for elections - in fact elections are dominated by large parties. The first point is crucial in understanding real elections: there cannot be a pure strategy equilibrium - the outcome of the election cannot be predicted in advance, it must be uncertain. Upsets such as Brexit or Trump are to be expected - and

[1] We take up the $c_0 \neq 0$ case in Section 5.4.

there is nothing any pollster or political scientist can do to make it otherwise. We call this the *uncertainty principle for elections*. Here already the theory tells us something: it tells us why pollsters are often wrong. This theory is pretty good for elections. It also works for wars, strikes, public demonstrations and other conflicts in which both sides pay regardless of whether they win or lose.

We will run through analysis of the all-pay auction taking account of the constraint $\varphi_k \leq 1$.

How It Works.

The lack of a pure strategy equilibrium is easy to establish.

THEOREM 5.1.1. *In equilibrium there cannot be a positive probability of a tie and there is no equilibrium in pure strategies.*

PROOF. Why can there not be a positive probability of a tie? Noting that the cost of a bid b is γb, if the tie is at $\gamma b < V$ then each party would wish to break the tie by shading its bid a little higher raising its probability of winning by 1/2 at trivial extra cost. If there is a positive probability of a tie at $\gamma b = V$ neither party wins with probability 1 when bidding V so each party strictly prefers to bid zero.

We can then make use of the absence of ties to show that there is no pure strategy equilibrium. With pure strategies and no tie one party must lose with probability 1 and so must be bidding 0. But if one party bids 0 the other party should bid the smallest number bigger than zero and there is no such number. □

To develop a deeper understanding it is convenient to introduce the concept of *desire to bid* and *willingness to bid*. To get a prize V_k for sure group k would bid up to $B_k = V_k/\gamma$ since this would cost V_k; this we call the *desire to bid*. On the other hand any bid $b_k = \varphi_k \eta_k$ by group k is bounded above by η_k since $\varphi_k \leq 1$. Therefore the highest bid W_k a party is *willing and able* to provide is the smallest of the two: $W_k = \min\{B_k, \eta_k\}$, which we call *willingness to bid* of group k. In the present context $V_k = V$ for both groups so $W_k = \min\{V/\gamma, \eta_k\}$. It is convenient to distinguish between a *medium stakes election* where $V \leq \gamma\eta_S$ in which both parties have the same willingness to bid, $W_S = W_L = V/\gamma$, and a *high stakes election* where $V > \gamma\eta_S$ and the large party has a higher willingness to bid, $W_S = \eta_S < W_L$. In the former case both groups are willing to spend V, in particular $\gamma W_S = V$; with large stakes the small group is constrained by size and $\gamma W_S < V$. Now we can be more specific about the equilibrium.

THEOREM 5.1.2. *The large party has an expected equilibrium payoff of $V - \gamma W_S$, while the small party gets 0. The small party bids 0 with probability $1 - \gamma W_S/V$ and the large party bids W_S with probability $1 - \gamma W_S/V$. All remaining probability of either party is a uniform density on $(0, W_S)$ of height γ/V.*

PROOF. To prove the theorem we will start by showing that one party must get 0 and both parties must bid arbitrarily close to W_S. The second fact will imply that it is the small party that gets 0. That the large party gets $V - \gamma W_S$ then follows easily.

One party must get 0. The argument is the following, we make it precise in the next boxed lemma. Suppose $\underline{b}$ is the lowest bid by either party. It cannot be that bidding $\underline{b}$ leads to a tie with positive probability. So one party k must face an opponent who has zero probability of bidding $\underline{b}$ or less. That means that k must be almost certain to lose if it bids near $\underline{b}$ so if it is bidding near $\underline{b}$ it must be getting

0 in equilibrium. If k is not bidding near $\underline{b}$ then the other party $-k$ must be and these are losing for sure so $-k$ must be getting 0 in equilibrium.

LEMMA 5.1.3. *If k's opponent $-k$ has zero chance of playing the lowest bid $\underline{b}$ then k must be bidding near $\underline{b}$ and these bids are almost certain to lose; in particular k must get* 0 *in equilibrium.*

To prove this we must first formally define a mixed strategy. This is a probability distribution represented by a cumulative distribution function over bids, that is, a *mixed strategy* G_k is a non-decreasing function on $(-\infty, \infty)$ with $G_k(b) = 0$ for $b < 0$ and $\lim_{b\to\infty} G_k(b) = 1$. It is right continuous and if it fails to be left continuous at a bid b, the height of the jump at b is the probability with which b is bid - it is an *atom* in the probability distribution. At points of continuity of G_k the probability of the bid is zero.
When we speak of a "lowest bid" $\underline{b}$ we mean that for $b < \underline{b}$ we have $G_k(b) = 0$ for both parties while for $b > \underline{b}$ we have $G_k(b) > 0$ for at least one of the parties.

PROOF OF THE LEMMA. By assumption party $-k$ has a continuous G_{-k} at $\underline{b}$. If it was the case that k is not bidding near $\underline{b}$ then for some $b > \underline{b}$ we have $G_k(b) = 0$. Hence it must be that $G_{-k}(b) > 0$ (as we defined $\underline{b}$ they cannot both be zero) and since bids by $-k$ in $(0, b]$ lose for certain, they are not made. Hence we have $G_{-k}(0) = G_{-k}(b) > 0$. This implies that $\underline{b} = 0$ and that G_{-k} is discontinuous there, a contradiction.
We may assume, then, that for $b > \underline{b}$ we have $G_k(b) > 0$. Since G_{-k} is continuous at $\underline{b}$, for $b \to \underline{b}$ we have $G_{-k}(b) \to 0$. That is to say that bids by k in the range $(\underline{b}, b]$, which we know have positive probability, lose with probability at least $1 - G_{-k}(b) \to 1$ and earn at most $G_{-k}(b)V \to 0$. □

Both parties must bid arbitrarily close to W_S: If the highest bid is less than W_S the party getting an expected payoff of zero should bid a shade higher because this would turn a profit. Moreover, one party cannot have a higher highest bid than the other, since the party with the higher highest bid could lower its bids, saving cost and still winning with probability 1. Hence both parties must bid near W_S.

Equilibrium payoffs: We can conclude that the small party must get 0 in equilibrium and the large party gets $V - \gamma W_S$, as follows. In a medium stakes election, as bids approach W_S, the most either party can earn approaches 0; hence - since expected equilibrium payoffs must be non-negative - both parties' expected payoff must be zero, and for the large party this is also equal to $V - \gamma W_S$. In high stakes elections the small party must get zero by the same argument above, which also implies that the large party cannot get more than $V - \gamma W_S$; on the other hand the large party cannot get less than $V - \gamma W_S = V - \gamma \eta_S > 0$ since it can always bid a bit more than η_S and win for sure; thus again its expected payoff is $V - \gamma W_S$.

Equilibrium strategies: since the large party gets $V - \gamma \eta_S$ in a high stakes election and bids close to zero it must still get $V - \gamma \eta_S$ for those low bids meaning that the probability it wins must be close to $1 - \gamma \eta_S / V$. For this to be the case the small party must bid zero with that probability.

To find rest of the equilibrium strategies we need to know that the probability that party k bids less than or equal to b denoted by $G_k(b)$ is continuous and strictly increasing on the open interval between 0 and W_S.

LEMMA 5.1.4. *$G_k(b)$ is continuous and strictly increasing on* $(0, W_S)$.

PROOF. If G_k is not strictly increasing this means there is a gap where party k does not bid. Notice that if there is a gap for one party the other party must have the same gap as there is absolutely no point bidding in a range where the other party does not bid: better to bid at the bottom. At the top hypothetical gap $b < W_S$ we know, since there is no positive probability of a tie, that one party $-k$ does not have an atom. Hence party k should not bid above but close to b: it would do better to bid at the bottom. Since we are assuming $b < W_S$ this contradicts the fact that b is the top of the gap. Hence there are no gaps.
If G_k is discontinuous at b then there is an atom there. Suppose this is the case. Then party $-k$ should not bid just below b: it would be better to bid just a bit above, increasing substantially the probability of winning while increasing cost only a shade. So there would have to be a gap below b and we just showed that is impossible. □

With G^k continuous the utility for party k from the bid b is $G_{-k}(b)V - \gamma b$. Since G_k is continuous and strictly increasing, utility must be constant for any b. As we already know the equilibrium payoffs we may directly compute that $G_L(b) = \gamma b/V$ and $G_S(b) = 1 - \gamma W_S/V + \gamma b/V$. □

5.2. The Tripartite Auction Theorem

As Mancur (1965) and many others since have argued, smaller lobbying groups often seems to have much greater success than the larger ones. Why is lobbying different than voting? One obvious difference is that we do not think that lobbying is an all-pay auction. Think of the bids $b_k = \eta_k \varphi_k$ as bribes offered to a politician who decides which group gets V. Politicians do not generally collect bribes from each group, rather they typically sell themselves to the highest bidder - taking a bribe only from the group that offers the better bribe. That is: lobbying is typically a winner-pays rather than all-pay auction.[2]

There are two important kinds of winner-pays auctions. One is a first-price sealed bid auction: each of the two lobby groups offers a bribe in a sealed envelope, and the politician returns the envelope holding the smaller bribe. The other is an English auction in which the lobby groups compete with each other increasing their offers until one drops out of the bidding. In this case the winner winds up paying just a shade more than the losers last bid - so from a game theoretic point of view it is pretty much the same as if each group put their best offer in an envelope with the high bid winning - but paying only the losing bid. This is called a second-price sealed bid auction and while less descriptively realistic, it captures the right idea and is easier to analyze.

[2]See Baye et al. (1993) and Che and Gale (2000), however, who argue that when payments are up front in the form of campaign contributions or "wining and dining" lobbying is in fact an all-pay auction. As the most significant payments - post-retirement jobs and jobs for relatives - are after the fact, we see that winner-pays is also important. A number of papers analyze lobbying using menu auctions. In the two bidder case this is the same as a second-price winner-pays auction.

We now have three different kinds of auctions: all-pay, first price sealed bid and second price sealed bid.[3] What difference does it make?

The second price sealed bid auction is a classic illustration of the idea of dominant strategies. The price you pay if you win does not depend on your bid, only on the other bidders bid. That means the only thing your bid does is determine whether you win or lose. As a consequence you can do no better than bidding your willingness to bid - in that case you win whenever it is advantageous to do so (for the amount you have to pay for winning is less than your willingness to bid) and lose whenever it is advantageous to do so. So each group bids their willingness to bid. If the stakes are high, $V > \gamma\eta_S$, we know $W_L > W_S = \eta_S$ hence the large group wins and gets $V - \gamma W_S > 0$ (prize value minus cost of second highest bid) while the small group gets zero. If on the other hand $V \leq \gamma\eta_S$ since $W_S = W_L$ both groups get zero whichever group gets the prize. Hence the amount that the groups earn is exactly the same as in the all-pay auction. The description of what happens is rather different, however: while in the all-pay auction it is necessarily uncertain which party wins, in the second price sealed bid auction if the stakes are high the large party wins for sure.

In the first price sealed bid auction if $V \leq \gamma\eta_S$ - so that both groups have the same willingness to pay - it is pretty obvious that the only pure strategy equilibrium is for both groups to bid their willingness to pay. Indeed, if the winning bid is less than that, it would pay the other group to bid a shade more, while if the winning bid is that amount and the losing group bid less, the winner would want to bid less. This is the same as for the sealed bid second price auction. On the other hand, when $V > \gamma\eta_S$ the small group cannot bid more than η_S, so the only equilibrium is for the large group to bid this amount and win for sure. Notice that here the tie-breaking rule must be endogenous: it must be that in case of a tie the large group wins. If we tried to say that the prize is split equally in case of a tie, the large group would always try to bid the smallest number bigger than the tie and there is no such number. The fact that the large group wins in equilibrium reflects the fact that it is the group willing to bid a bit more in order to win. So we see that it does not matter whether the winner pays auction is a first price or second price auction.

This result - the tripartite auction theorem - says that with a certain prize the utility of the bidders in an all-pay, first price sealed bid and second price sealed bid auction is exactly the same. The result is quite robust - it does not require the two bidders to have the same costs for providing effort, nor does it require that they value the prize the same way.[4]

So: the theory presented so far about political auctions - where recall we have restricted attention to the case $c_0 = 0$, that is no committed voters and no fixed costs - can explain why elections are uncertain and lobbying much less so. But

---

[3]For the first price auctions the solution concept is Nash equilibrium in weakly undominated strategies. This forces bidding the value of the prize in sealed bid second price auctions and eliminates bidding higher than the value in sealed bid first price auctions.

[4]Notice that the tripartite auction theorem has nothing to do with the better-known revenue equivalence theorem - the tripartite auction theorem is about the bidders utility in an auction with a commonly known value, while the revenue equivalence theorem is a theorem about the sellers utility in an auction with private values. In fact, while it is possible to show that a first price sealed bid and second price sealed bid auction generates the same revenue to the auctioneer, this is certainly not the case for an all-pay auction.

it certainly does not explain why small lobbying groups are effective - quite the opposite, it says that they should be ineffective. Whatever it is that explains the difference between elections and lobbying it is not the fact that one is an all-pay auction and the other a winner-pays auction.

5.3. Unequal Prizes

The tripartite auction theorem is valid even if the value of the prize for the two groups is unequal: for example, if the small group gets V_S if it wins and the large group gets V_L. There are two issues that arise: first, in this case efficiency demands that the prize go to the group with the higher value. Second, if the bids are a bribe paid to a politician, how does the politician view the different auction formats? What does the tripartite auction theorem tell us about these?

Willingness to bid is $W_k = \min\{V_k/\gamma, \eta_k\}$ and we refer to the group with the greater willingness to bid as *advantaged* $(-d)$ and the group with the lesser willingness to bid as *disadvantaged* (d). From consideration of the second price auction we see that the result is that the disadvantaged party gets nothing and the advantaged group gets the difference between their prize value and the cost of matching the willingness to bid of the disadvantaged group: $V_{-d} - \gamma W_d$.

To see which group is advantaged observe that if $V_S < V_L$ then clearly $W_S < W_L$ and the large group is advantaged. If $V_S > V_L$ there are two cases. If $V_L < \gamma\eta_S$ then $W_L = V_L/\gamma < \min\{V_S/\gamma, \eta_S\} = W_S$ and the small group is advantaged. If $V_L > \gamma\eta_S$ then $W_L = \min\{V_L/\gamma, \eta_L\} > \eta_S = W_S$ and the large group is advantaged. Hence the group with the higher value of the prize is advantaged unless $\gamma\eta_S < V_L$, in which case the small group is unable to compete and so the large party is advantaged. To visualize: this typically happens for small γ, hence with $W_k = \min\{V_k/\gamma, \eta_k\} = \eta_k$ the large party wins.

5.3.1. Efficiency and Direct Democracy. If $V_L < \gamma\eta_S$ the result is a good one for efficiency: efficiency demands the prize go to the group with the higher value and indeed in both winner-pays auctions this is the case. Regardless the surplus accrues to the group with the higher value.

If on the other hand $V_L > \gamma\eta_S$ the large group wins regardless of efficiency. For fixed parties sizes, for γ small enough this is what happens. Thus with small γ there is a large set of contexts - that is (V_S, V_L) pairs - where the large group wins even if $V_L < V_S$.

In this vein what about lowering participation costs, for example as many populist parties propose, through internet voting and the like? The effect is to lower γ, the cost of turning out participants. Lowering participation cost, in other words, reduces efficiency, allowing the large group to win even when it is inefficient to do so. This is bad. The worship of the 51% majority - the idea that if a group wins by one vote they have some unique moral claim to do whatever they want - flies in the face of efficiency. Efficiency aside, it is hard to see the morality of dispossessing a minority in order to provide a trivial gain to a majority. Perhaps not surprisingly we observe that the same populist parties that worship the 51% majority and are eager to lower the cost of voting have little regard for minority rights and exhibit little concern for anyone but themselves. Their plans, however, may be shortsighted in that other more costly methods of resolving political conflict are always available - such as demonstrations and civil war.

5.3.2. Bribery and Lobbying. In the winner pays auctions the politician gets W_d - the willingness to bid of the disadvantaged group, which is the bid of the winning advantaged group. How does this compare with the all-pay auction?

Let C_k be the expected bid for group k in the all-pay auction. The politician gets $C_S + C_L$. From the surplus result we can compute this relatively easily. The expected utility of a group is the difference between the expected value of winning and the expected cost of bidding: $\pi_k V_k - \gamma C_k$. We know this is $V_{-d} - \gamma W_d$ for the advantaged group and 0 for the disadvantaged group. That is $\pi_{-d} V_{-d} - \gamma C_{-d} = V_{-d} - \gamma W_d$ and $\pi_d V_d - \gamma C_d = 0$; adding up using $\pi_d + \pi_{-d} = 1$ we find the profit of the politician: $C_{-d} + C_d = W_d - \pi_d(V_{-d} - V_d)/\gamma$. When $V_L < \gamma \eta_S$ or more generally, if the advantaged group is the one with the greater value of the prize, we conclude that the politician gets less with the all-pay auction than with the winner pays auctions.

Thus while the groups are indifferent to the auction format the politician is not. This means both that the auction preferred by the politician is likely to be implemented and that it is the most efficient. It makes sense then that lobbying generally involves winner pays auctions: this is the format generally preferred by the politician.

5.3.3. Private and Non-rival Prizes. Our benchmark case is that of an equal or *private* prize: the winning group gets V.[5] This makes sense when the prize involves a transfer payment between groups. Examples include control over natural resources, the division of government jobs, the division of a fixed budget, taxes and subsidies (such as farm subsidies), limitations on competition such as trade restrictions, occupational licensing and generally speaking a prize involving money, goods or services. These prizes are studied, for example, by Shachar and Nalebuff (1999), Herrera et al. (2015) and Levine and Mattozzi (2020). As we have noted, a private prize is efficiency neutral: it does not matter who wins, so this is a useful benchmark case.

A useful contrast is the case of a *non-rival* prize $V_k = \eta_k v$, where each individual receives a fixed utility benefit from winning independently of group size. Examples include civil rights, laws concerning abortion, the right to bear arms, to marry, to sit at the front of the bus, criminal law, defense expenditures, non-trade foreign policy, and policies concerning monuments. These prizes are studied, for example, by Palfrey and Rosenthal (1985) and Feddersen and Sandroni (2006).

The key thing to note is that a non-rival prize always advantages the large group which has more members to enjoy the per capita value. However, the two extreme cases of private and non-rival prize ignore a number of factors that are important. For private prizes transfers may be efficient (lowering taxes to one group and subsidies to the other) or inefficient (the reverse), so that the prizes will not be equal. Moreover, when civil rights and law changes are at stake it may make sense to assume that the benefit of winning is the same for all members within a group. It is less certain that both groups should receive the same benefit - is the benefit of depriving another the right to sit in the front of the bus equal to cost of being deprived of that right? Indeed in the context of liquor referenda Coate and

[5] The terminology is due to Mancur (1965). Private here does not mean that it is possible to transfer the prize between group members, but rather the members of the winning group share equally in the prize.

Conlin (2004b) quite naturally consider a non-rival prize in which the value to the two groups may be different.

In all likelihood reality lies in between a private prize and a non-rival prize: typically elections involve a mix of issues, some involving taxes and transfers, other involving rights. Esteban and Ray (2011) consider a mix in the context of ethnic conflict and Esteban et al. (2012) have empirical results indicating that in this context private prizes are roughly five times more important than non-rival prizes (see Section 16.4).

5.4. Voting versus Lobbying: Duties or Chores?

According to the results presented so far (concerning the case $c_0 = 0$) the large group should not do worse than the small group in either voting or lobbying. In Table 5.4.1 we give some data about farm subsidies and the size of the agricultural sector. In these advanced highly urbanized countries, agriculture is a tiny fraction of GDP, less than 3%. Yet the amount of time annually that the average person must work to pay these subsidies is as high as half a week. More importantly, there seems to be a systematic relationship: the less important is agriculture the more time non-farmers have to work in order to support them. It really does seem that smaller groups are more effective in lobbying.

TABLE 5.4.1. Farm Subsidies

country	% agriculture	farm subsidy hours
Switzerland	0.8	23
Japan	1.2	19
U.S.	1.3	11
Norway	1.6	17
EU	1.7	14
Canada	1.7	8
Australia	2.4	2

This is a subset of OECD countries of similar development characteristics, size and stable democratic institutions for a relatively long period of time. As such, Iceland and New Zealand are excluded as they are much smaller, Mexico and Turkey are excluded because they are much poorer, and Israel is excluded because of the widespread use of agricultural cooperatives (the Kibbutz system). The % agriculture is the share of agriculture in value added in 2014 from http://data.worldbank.org/indicator/NV.AGR.TOTL.ZS. The farm subsidy hours is total agricultural support as a percent of GDP in 2014 from http://stats.oecd.org/viewhtml.aspx?QueryId=70971&vh=0000&vf=0&l&il=&lang=en multiplied by 2000 working hours per year.

By contrast, smaller groups rarely do well in voting. Between 2000 and 2018 there were nine special elections for U.S. House seats in California vacated due to death or resignation to take another job. These are one issue elections so turnout is not influenced by other ballot items, and the platform is determined by the state or national party. In every one of these elections the party with the largest number of registered voters in the district won the election. More generally it is an extremely rare event that a small single-issue party can prevail in elections.

Since the farmers are obviously not winning elections, it must be that they are successful at lobbying. But why? Why does not the 90% plus of the people in the economy who are not farmers form an anti-farm lobby and prevent the farmers from picking their pocket? The question seems to answer itself. Take the United States. Is it worth it to take the time and effort to find, learn about, join and support an anti-farm lobby in hopes of getting an extra 11 hours a year? It is hardly worth it to the lobby to vet me, process my application and so forth if I am only going to contribute the equivalent of a few hours a year. There is a substantial fixed cost in joining an organization: you cannot simply write a check for 32 cents to the "anti-farm lobby" as an effective way to lobby against them - it would cost more than 32 cent to process my check. Considerable cost would be incurred even as I contributed absolutely nothing to the lobbying effort. That is, lobbying is a chore.

Elections provide a contrast: if there is a referendum for example, while it is costly to go to the polling place and take time to vote, some people may view it as their civic duty, so they vote and the satisfaction of having discharged their duty might more than offset the direct cost of participating. As we argued earlier, while lobbying is a chore voting is a duty. We will now see that this makes the difference.

5.4.1. Duties Versus Chores. We are going to maintain the linearity assumption that $\theta = 1/2$ but drop the assumption that $c_0 = 0$ to allow for duties and chores. In the case of duties ($c_0 < 0$, the voting case) there will be committed members $\underline{\varphi}$, in the case of chores ($c_0 > 0$, lobbying) a fixed cost F (Figure 4.2.1 on page 27 illustrates). Here we have that the cost $\eta_k C(b_k/\eta_k)$ leads to desire to bid B_k given by $\eta_k C(B_k/\eta_k) = V$. As before the willingness to bid is $W_k = \min\{B_k, \eta_k\}$, except that now it might be $B_k \leq 0$ in which case $W_k = 0$. As before the group with the highest willingness to bid will be called the *advantaged* group and the other group will be called *disadvantaged*.

We consider first the interesting case in which $V > \eta_S F$.[6] Here we define the stakes as *medium* if $V < F\eta_L + \gamma\eta_S$ and *high* if $V > F\eta_L + \gamma\eta_S$. There are three key results on auctions with prize V:

- The level of utility of the two groups is the same regardless of whether the prize is allocated by an all-pay, first-price or second-price auction (Section 5.2 above).
- Only an advantaged group can receive a positive level of utility and always does so (Section 5.3 above).
- The small group is advantaged for a chore with medium stakes; the large group is advantaged for a duty, and for a chore with high stakes (Theorem 5.4.1 below).

We indicated that our earlier theory with $F = \underline{\varphi} = 0$ worked well for voting. If we think of voting as a duty the result here strengthens that: the large group is advantaged and while with $\underline{\varphi} = 0$ and $V \leq \gamma\eta_s$ the large party earned no utility (Theorem 5.1.2), with $\underline{\varphi} > 0$ the large party always earns something. For a chore such as lobbying we get a different result: for medium stakes it is the small group that is advantaged. Roughly speaking, with a fixed cost per person of organization, a large group faces a greater fixed cost so it is less willing to bid. However, if the prize is big enough they will take advantage of their greater resources to get the prize.

[6]This rules out the case in which, for a chore, we may have $B_k < \eta_k\underline{\varphi} = 0$ for both k in so that neither group submits a bid: neither is willing to pay the fixed cost even for a certainty of getting the prize.

Notice that in Table 5.4.1 the stakes are indeed relatively modest. While farmers are successful at getting subsidies, they are not imposing a very great tax on the non-farmers. If, for example, the numbers for the amount of time spent paying for farm subsidies in Table 5.4.1 corresponded to months rather than hours, it seems likely that the non-farmers would lobby and lobby effectively. Indeed the defeat in the U.S. Congress of the "Stop Online Piracy Act" seems to be a case in point. The act was put forward by the pro-copyright lobby. More modest efforts to impose broad internet restrictions on general internet users to protect a few holders of copyrights had passed the U.S. Congress relatively easily. This more ambitious act was sponsored by a majority of the U.S. Congress, but the drastic nature and the non-negligible consequences of the act led to a broad grass roots lobbying effort against it. As a result many of the sponsors dropped out and the act was quietly shelved.

We should point out that for some prizes elections are chores. This means small groups may want small stakes, but may also want to select issues for which there are no committed voters. This may be the case in the school board elections we consider in Section 5.8.

5.4.2. Auctions with Duties and Chores: Main Result. We now establish the main new result: that the small group is advantaged for a chore with medium stakes; the large group is advantaged for a duty, and for a chore with high stakes.

Recall from Theorem 25 that $C(\underline{\varphi}) = 0$, while with $\theta = 1/2$ the cost of norm $\varphi_k > \underline{\varphi}$ is given by $C(\varphi_k) = F + \gamma(\varphi_k - \underline{\varphi})$ where $F = \max\{0, \theta c_0\}$. With $\mathtt{1}\{\cdot\}$ denoting indicator, for $\varphi_k \geq \underline{\varphi}$ we may then write

$$C(\varphi_k) = F \cdot \mathtt{1}\{\varphi_k > \underline{\varphi}\} + \gamma(\varphi_k - \underline{\varphi})$$

where $F = \max\{0, \theta c_0\}$ so that if $c_0 > 0$ (chores) it is $F > 0$ and $\underline{\varphi} = 0$, while if $c_0 < 0$ (duties) we have $\underline{\varphi} > 0$ and $F = 0$.

The next result, which establishes which group is advantaged, hinges on the following observation. Since $\eta_k C(b_k/\eta_k) = b_k \cdot AC(b_k/\eta_k)$, where AC denotes as usual average cost, then for $\varphi_k \geq \underline{\varphi}$: if AC is increasing - as in the case of a duty - then for any b we have $b \cdot AC(b/\eta_L) < b \cdot AC(b/\eta_S)$ whence $B_L > B_S$; if AC is decreasing - as in the case of a chore - then analogously we deduce $B_L < B_S$. Here is the key role of convexity and concavity.

THEOREM 5.4.1. *Assume $V > \eta_S F$. Then the small group is advantaged in a chore with medium stakes. Otherwise with medium or high stakes, the large group is advantaged.*

PROOF. For $\varphi_k > \underline{\varphi}$ we have

$$C(b_k/\eta_k) = \begin{cases} \gamma b_k/\eta_k - \gamma\underline{\varphi} & \text{duty} \\ F + \gamma b_k/\eta_k & \text{chore} \end{cases}$$

In the case of a duty $B_k = V/\gamma + \eta_k\underline{\varphi} > 0$. Hence, $B_L > B_S$. Since $W_k = \min\{B_k, \eta_k\}$ clearly $W_L > W_S$. So the large group is advantaged.

In the case of a chore $B_k = V/\gamma - \eta_k F/\gamma$ so medium stakes $V < F\eta_L + \gamma\eta_S$ means $B_L < \eta_S$, and analogously high stakes means $B_L > \eta_S$. The assumption $V > \eta_S F$ says $B_S > 0$, so $W_S = \min\{B_S, \eta_S\} > 0$. If $B_L \leq 0$ then $W_L = 0 < W_S$. On the other hand if $B_L > 0$ then $W_L = \min\{B_L, \eta_L\}$. In this case, recalling that

$B_L < B_S$, if $B_L < \eta_S$ we get $W_L = B_L < \min\{B_S, \eta_S\} = W_S$; if $B_L > \eta_S$ then $W_L > \eta_S \geq W_S$. $\square$

The key idea is that the convexity or concavity of $C(\varphi_k)$ determines whether average costs of bidding are increasing or decreasing. The smaller group must always choose a higher fraction φ_k to match the bid of the larger group. In the convex case this implies a higher average cost disadvantaging the smaller group, and conversely in the concave case, provided that the small group is able to bid that high. Note that the fixed cost plus constant marginal cost is not important here, merely the overall convexity or concavity of the function $C(\varphi_k)$.

Let d be the disadvantaged group. Since $\eta_k C(\varphi_k) = \eta_k F \cdot \mathbb{1}\{\varphi_k > \underline{\varphi}\} + \gamma(b_k - \eta_k \underline{\varphi})$, if $W_d > \eta_{-d}\underline{\varphi}$ it costs the advantaged group $\eta_{-d}F + \gamma(W_d - \eta_{-d}\underline{\varphi})$ to match the bid of the disadvantaged group, while if $W_d \leq \eta_{-d}\underline{\varphi}$ since $\eta_{-d}C(\underline{\varphi}) = 0$ it costs nothing to overbid the disadvantaged group. We define the *surplus* as the difference between the value of the prize and the cost of matching the bid of the disadvantaged group if this is positive, zero otherwise.

THEOREM 5.4.2 (Tripartite Auction Theorem). *In the second-price, first-price and all-pay auction, the disadvantaged group gets 0 and the advantaged group gets the surplus.*

PROOF. The argument given in Section 5.2 shows that for the winner-pays auctions the disadvantaged group gets 0 and the advantaged group the value of the prize minus the cost of matching the willingness to bid of the disadvantaged group. For the all-pay auction the computation of utilities follows the lines of that for the simple all-pay auction in Theorem 5.1.2. $\square$

5.5. Why not Split a Large Group?

In the case where effort is a chore it is intuitive that the smaller group has an advantage: it must pay the fixed cost for a smaller number of members. A natural question is why the larger group does not just "act like a smaller group" by appointing a smaller subgroup to act on its behalf. The problem is that the prize is evenly split among the entire group. For example, for the non-farmers the benefit of eliminating farm subsidies is lower taxes and lower prices for food. This is shared by all non-farmers regardless of who bears the cost of lobbying. If, for example, the urbanites of Manhattan were appointed to do the anti-farm lobbying they would care only about the reduction of their own taxes and food prices, not the reduction in Los Angeles.

To see how this works, suppose a subgroup of size $\mu_k < \eta_k$ is appointed to act on behalf of the group. The prize is only worth $(\mu_k/\eta_k)V$ to the subgroup. Recall that willingness to bid is a non-decreasing function of the desire to bid

$$B_k = \eta_k \underline{\varphi} + V/\gamma - \eta_k F/\gamma.$$

For the subgroup this is

$$B_k^{\mu_k} = \mu_k \underline{\varphi} + (\mu_k/\eta_k)V/\gamma - \mu_k F/\gamma = \frac{\mu_k}{\eta_k}\left(\eta_k \underline{\varphi} + V/\gamma - \eta_k F/\gamma\right) = \frac{\mu_k}{\eta_k} B_k.$$

The desire of the subgroup to bid is always a fraction μ_k/η_k of the desire to bid of the entire group. Hence if the entire group is disadvantaged - the subgroup is even more so.

5.6. Vote Suppression

The 1993 Ed Rollins scandal suggests that sometimes party effort can be directed to suppress the votes of the opposition. More recently, it has been argued that voter identification laws increase voting costs for relatively poor Democratic voters (in particular Hispanics, Blacks, and mixed-race Americans) with relatively little effect on whites and on the political right.[7] Our model can be used to investigate the strategic use of vote suppression. Suppose that each party can slightly increase the participation cost of the opposing party from $c(y)$ to $\tilde{c}(y) = c(y) + h$ where $h > 0$, by incurring a cost of κ. As we are analyzing voting, we assume $F = 0$. We analyze the case in which the stakes are low enough that the small party is not willing to turn out all their voters.

THEOREM 5.6.1. *[Cesar Martinelli] If h is sufficiently small then only the advantaged party may want to suppress votes; if κ is sufficiently small it will do so.*[8]

PROOF. We first show that voter suppression raises total cost. Let $\tilde{C}(\varphi_k)$ be the total cost after voter suppression. Since $c(\underline{\varphi}) = 0$, and $\tilde{c}(\underline{\varphi}) = h > 0$, then $\underline{\tilde{\varphi}} < \underline{\varphi}$. For $\underline{\tilde{\varphi}} < \varphi_k \leq \underline{\varphi}$ we have $\tilde{C}(\varphi_k) > 0 = C(\varphi_k)$. For $\varphi_k > \underline{\varphi}$, we have that

$$\tilde{C}(\varphi_k) = \tilde{C}(\underline{\varphi}) + \int_{\underline{\varphi}}^{\varphi_k} [c(y) + h]dy + \theta(1 - \varphi_k)[c(\varphi_k) + h] >$$

$$> \int_{\underline{\varphi}}^{\varphi_k} c(y)dy + \theta(1 - \varphi_k)c(\varphi_k) = C(\varphi_k)$$

Next, we show that raising total cost leads to the result. If h is sufficiently small then the disadvantaged party cannot suppress enough votes to become advantaged, so vote suppression never changes which party is advantaged. The disadvantaged party therefore gets zero payoff regardless of whether it suppresses votes or not, hence it will not pay a positive cost to do so. On the other hand, if h is sufficiently small and since the willingness to pay of the small party is given by B_k, increasing the cost of the disadvantaged party must strictly decrease its willingness to bid. Since the cost function of the advantaged party remains unchanged when the disadvantaged party does not suppress votes, its surplus therefore goes up. Hence if κ is small enough it is worth paying. □

Our theory suggests that when the Republican party is advantaged it will be tempted to introduce voters suppression laws in the form of strict voter identification requirements. It will do so in order to hold down its own turn out cost. The rationale discussed in the popular press revolves instead about taking strategic advantage: in a closely contested state a short term Republican victory may be translated into long term advantage by introducing a voter suppression law.[9] To explore this we take a strong measure of what has been argued to be voters suppression - strict photo id laws. The table below reports the states with these laws and the extent of Republican advantage in those states.

---

[7] See Hajnal et al. (2016) and Ingraham (2016).

[8] This theorem was suggested to us by Cesar Martinelli during the 2015 Priorat Workshop.

[9] For example, from Levy (2016) writing in *Mother Jones* "In order to mitigate their waning political popularity, Republicans have ... passed an unconstitutional voter suppression law to weaken the voting power of African Americans and other Democratic-leaning voters."

TABLE 5.6.1. Voter ID Laws in the USA

State	Year	Republican Advantage
Alabama	2014	+23
Kansas	2011	+22
Tennessee	2011	+20
Texas	1990	+16
Mississippi	2011	+12
Indiana	2005	+10
Georgia	1977	+8
Virginia	1996	+3
North Carolina	2013	+3
Wisconsin	2011	-7

The table reports the states with strict photo identification laws, the year the law was introduced and the 2012 vote differential between Romney and Obama as a measure of which party is advantaged. Data from https://en.wikipedia.org/wiki/Voter_ID_laws_in_the_United_States.

Interestingly, the GOP holds an overwhelming electoral advantage in most of the states with strict photo id laws: the median Republican advantage in these states is more than 10 points. This is consistent with our theory and not with the strategic advantage theory. Only in the case of Virginia, North Carolina and Wisconsin - all of which have had both Republican and Democratic governors in recent years, and two of which only recently have introduced voter suppression laws - does the strategic theory seem to have merit.

5.7. Appendix: Types of Equilibria in the All-Pay Auction

In the all-pay auction there are qualitatively different equilibria depending on the size of the prize. We categorize this by the level of stakes, running from high to very low and summarize the situation in the table below.

stakes	condition	advantaged group
high	$W_S = \eta_S, W_L > \eta_S$	large
intermediate	$\eta_S > W_S, W_L > \eta_L \underline{\varphi}$	duty: large; chore: small
low	$W_S < \eta_L \underline{\varphi}$ or $\eta_S F < V < \eta_L F$	duty: large; chore: small
very low	$V < \eta_S F$	duty: impossible; chore: none

In describing the equilibrium it is useful to introduce the concept of bidding *at the bottom*. For a duty this means bidding only the committed members $b_k = \eta_k \underline{\varphi}$. For a chore it means bidding zero and for the small group paying the fixed cost $q_S(0) = 1$ and for the large group not paying the fixed cost $q_L(0) = 0$. Recall also that $G_k(b_k)$ is the cumulative probability that group k bids at or below b_k.

5.7.1. High stakes.

The constraint that the greatest effort group k can provide is η_k cannot bind on the large group since if the large group is willing to bid η_L and the small group can bid at most η_S: if the constraint was binding on the large group it would already bind on the small group. If it does bind on the small group the small group must be disadvantaged and the large group does not need to bid more than η_S. We refer to this as the *high stakes* case. It occurs when both groups desire to pay exceeds the ability of the small group to pay and the ability of the small group to pay exceeds the committed bid of the large group. For duties this is the same as the desire to pay of the small group exceeding its ability to pay. For chores this as the willingness to pay of the large group exceeding the ability of the small group to pay. In all cases the large group is advantaged.

Description of Equilibrium: Both groups bid using the same uniform distribution on $[\eta_L \underline{\varphi}, \eta_S]$. Both groups have positive probability of bidding at the bottom. The large group also has a positive probability of bidding η_S winning for sure if there is a tie.

The probability of the large group bidding at the bottom is determined by the small group earning zero. For a duty the probability the small group wins by bidding the large group committed effort times the value of winning for sure must equal the cost of that bid $G_L(\eta_L \underline{\varphi})V = \gamma(\eta_L - \eta_S)\underline{\varphi}$. For a chore the probability the small group wins by incurring the fixed cost times the value of winning for sure must equal the fixed cost to the small group $G_L(0)V = \eta_S F$. The remaining probability is determined by computing the probability p_L "left over" from the uniform. The height of the uniform is γ/V so the probability of the uniform is

$$1 - p_L = \gamma \frac{\eta_S - \eta_L \underline{\varphi}}{V}$$

The small group bids at the bottom with exactly probability p_L while the large groups bids the top η_S with probability $p_L - G_L(\eta_L \underline{\varphi})$, that is, in the case of a duty

$$\frac{V - \gamma(\eta_S - \eta_L \underline{\varphi}) - \gamma(\eta_L - \eta_S)\underline{\varphi}}{V} = \frac{V - \gamma\eta_S(1 - \underline{\varphi})}{V}$$

and in the case of a chore

$$\frac{V - \gamma(\eta_S - \eta_L \underline{\varphi}) - \eta_S F}{V} = \frac{V - \eta_S(\gamma + F)}{V}$$

5.7.2. Intermediate stakes. If the willingness to pay of both groups is less than the ability of the smaller group to pay and exceeds the committed bid of the large group then the constraint does not bind on the small group and both groups are active in bidding. For duties this is the same as the desire to pay of the small group being less than its ability to pay but greater than the committed bid of the large group. For chores this is the same as the willingness to pay of the large group less than the ability of the small group to pay but positive. This case also be described as the *interior case*. Here the advantage depends on whether cost are a duty or a chore: in the case of a duty the large group is advantaged and in the case of a chore the small group.

Description of equilibrium: Both groups bid using the same uniform distribution on $[\eta_L \underline{\varphi}, W_d]$. Both groups also bid with the same positive probability at the bottom. The probability of a bottom bid is determined by the disadvantaged group earning zero. For a duty the probability the small group wins by bidding the large group committed effort times the value of winning for sure must equal the cost of that bid $G_L(\eta_L \underline{\varphi})V = \gamma(\eta_L - \eta_S)\underline{\varphi}$. For a chore the probability the large group wins by incurring the fixed cost times the value of winning for sure must equal the fixed cost to the large group $G_S(0)V = \eta_L F$

5.7.3. Low stakes: duty. If the willingness to pay of the small group is less than the committed bid of the large group then the small group will bid only its committed members and the large group should do the same. The large group is advantaged and wins for certain.

Description of equilibrium: Both groups bid their committed voters with probability one.

5.7.4. Low stakes: chore. If the willingness to pay of the small group is positive but that of the large group is not the large group will not bid nor pay the fixed cost. The advantage lies with the small group and there are two different types of equilibrium. First the small group may also not bid and win for sure. Second the small group may pay the fixed cost and if it does not do so it loses with sufficiently high probability that it does not gain.

Description of equilibrium 1: Neither group pays the fixed cost, the small group wins with probability p_S high enough it does not wish to enter $p_S V \geq V - \eta_S F$.

Description of equilibrium 2: The large group does not pay the fixed cost, the small group does. If the small group fails to pay the fixed cost it wins with probability no greater than p_S determined by the condition it cannot profit by failing to pay the fixed cost $p_S V \leq V - \eta_S F$.

5.7.5. Very low stakes: chore. In the case of chores only it may be that neither group is willing to pay the fixed cost. In this case neither does and the tie-breaking rule is arbitrary.

5.7.6. Low and very low stakes: chore - does the contest take place? In the case of a chore with low or very low stakes there are equilibria in which neither group pays the fixed cost. Rather than applying an arbitrary tie-breaking rule in this case in many applications is may be more interesting to assume that the contest does not take place and both groups get zero. In the case of low stakes this rules out the equilibrium of type 1 leaving on the equilibrium of type 2 in which the small group wins by paying the fixed cost.

5.8. When Do Small Groups Win Elections? The case of School Boards

Small groups do not always do poorly in elections: in school-board elections in the USA a small party of teachers faces a large party of students' parents. The school board controls resources and money that can be allocated either to teachers or to students. If the teachers win the election the money goes to them, if the parents win the election the money goes to the students. Hence we take it as a reasonable approximation that there is a common prize.

Value of the Prize

It is natural to think that the prize might be worth more to teachers since they get the benefits over their career while parents only get benefits while their children are in school. But some rough calculations show that this is probably not a big difference. Assume an election is held every three years. Because this is a relatively short time we ignore discounting. We assume that everyone lives in the district and that nobody leaves the district. We assume that teachers arrive in the district at the time they take the job while parents arrive two years before their first student enter school. From https://nces.ed.gov/pubs2007/ruraled/tables/tablea3_8.asp?referrer=report the average experience of a public school teacher is about 14 years. We assume that every teacher remains on the job the same length of time then retires so that the length of career is 28 years. We assume each parent has three children and four years between children. This means that they will have children in school for 21 years, plus we add two years for arriving in the district early, so they remain in the voting population for 23 years. Hence the average length of a teacher career is about 22% longer than the average length of a parent being in the interested population. However, what is relevant is how this determines the turnover rate during the three years between elections. After three years with a constant departure rate 3/28 of the teachers will have left and 3/23 of the parents. They leave in a continuous stream not all at the end, so the fraction of the prize lost due to departure will be half this amount: 3/56 for teachers and 3/46 for parents. Hence among those party members who are able to vote in the current election teachers will claim 53/56 of the prize and parents 43/46. That means that the ratio of prize value of teachers to parents is $53/43 * 46/56 = 1.012$ meaning that the value of the prize to the teachers is only about 1.2% greater than to the parents.

School board elections are often held at a different time than other elections so that the school board is the only issue on the ballot. In these elections, teacher unions are extremely effective at getting their candidates on the board (86%). While turnout among teachers is extremely high (90%), overall turnout is very low (10-20%), which is consistent with the idea that the many voters who are not parents of students are uninterested in the election. Since school boards generally control budgetary resources and can allocate those between teachers and pupils but they cannot set taxes or determine the overall size of the budget, we can take this as meaning that the prize is indeed of somewhat intermediate size.

Relative Sizes and Turnout

The basic source of information about school board success, turnouts and timing are the empirical studies by Moe (2003) and Moe (2006). We supplement this with some facts about the LAUSD, a large urban school district. From achieve.lausd.net/about we learn that the district has 84,000 employees (about half teachers) and 640,000 students, while the population of the district, which is approximately the same as the city of Los Angeles, is 3.9 million. If we assume three people per family, we have about 1.3 million families, while with three students per family, about 210,000 families of students. This gives a rough estimate of the large party (the families of students) being 2.5 or more times the size of the small party (the employees).

Turnout is reported in http://www.scpr.org/news/2015/05/12/51612/how-low-voter-turnout-impacts-lausd-schools/ as about 10-20% registered voters, while in LA county, for which data is available from http://www.laalmanac.com/election/el18.htm, there are about 4.7 million registered voters out of a population of 10 million, so about half the population are registered voters. This means that roughly 300,000 votes (10-20% of half of 3.9 million) are cast in a school board election. Assuming 90% turnout among employees and spouses means about 150,000 votes (2 times 90% of 84,000), about half of the total. To get an idea of the turnout among parents: if families have four members there are 1 million families in the 3.9 million population, that is 2 million parents, of whom half registered, so we come to roughly half a million parents; so their turnout is $150,000$ in one million or 15%.

Can this election result be explained by monitoring costs? For this to be the case it would have to be that there are few committed voters: this seems plausible since it seems that civic duty does not extend for most people to elections that are viewed as unimportant - and the very low turnout seems consistent with this idea. Hence turnout cost is $C(\varphi) = \theta\varphi + (1/2)(1-2\theta)\varphi^2$. In other words cost is globally concave, favoring the smaller group of teachers, if and only if $\theta > 1/2$, that is, monitoring difficulty is high. Of course since the turnout among teachers is very high, for the teachers monitoring costs do not much matter. In the extreme case of 100% turnout there are no monitoring costs at all. So the key issue is whether it is plausible that the monitoring costs of the parents are sufficiently large as to make overall cost sufficiently concave as to disadvantage them.

What determines monitoring costs?[10] It depends on the ease with which group members observe each other. In practice the people who are likely to be able to observe you are your friends and neighbors. In large general elections with relatively high stakes, those people are likely to be involved in the election. In small special purpose elections involving only a small fraction of the electorate, the other people involved - the other parents - are less likely to be "close" to you, as only a fraction of your friends, co-workers and neighbors are fellow parents. This suggests that indeed it makes sense to think that monitoring costs are relatively high in small special purpose elections like those for school board.

Our theory, incidentally, suggests a hypothesis for which data does not appear to currently exist. All other things equal, if we look at electoral success of teachers unions against the resources controlled by the board (the size of the prize) we might

[10] See section 9.4 for a formal model of how monitoring costs are determined.

expect the curve to be U-shaped: a very weak school board with little power would not be worth controlling, while a very powerful school board with a lot of resources would advantage the larger party. Notice too the implication that the teacher union has an incentive to keep school boards from getting too strong - while it would be better to control a more powerful board, it is less likely to be able to do so.

CHAPTER 6

When Lobbyists Succeed and Pollsters are Wrong

In the title of the book we talk about lobbyists succeeding and why pollsters being wrong. We have now developed the theoretical tools needed to explain.

6.1. When Lobbyists Succeed

When and why do lobbyists succeed? Sometimes they do not. The Disney Corporation is very effective in getting retroactive copyright extensions whenever their Mickey Mouse copyright is due to expire - but large pharmaceutical companies have never managed to get a retroactive patent extension when their blockbuster drug patents are due to expire. The copyright industry does sometimes lose in Congress as it did when it proposed the "Stop Online Piracy Act." If we accept that lobbying is a chore then indeed small groups - "special interests" - have an advantage at lobbying; they derive advantage from the fact that as a group they incur a lower average cost of providing resources for lobbying. On the other hand, they control fewer resources so if the prize is large and both groups are "all in" they will lose. *This is the heart of Theorem* 5.4.1: small groups are advantaged for chores provided the prize is not too large.

Indeed, the cost to the large group of pharmaceutical consumers - and generic manufacturers - of patent extension is very large so it is difficult for the small group of pharmaceutical producers to succeed. On the other hand, the size of the prize involving copyrights over Mickey Mouse is relatively small so that the Disney corporation is quite successful. In general the stakes are low for copyright and the copyright lobby is quite effective. Once in a while the prize gets too large it becomes worthwhile for the large group to suffer the fixed cost and start lobbying. That is a pretty good description of the events surrounding the "Stop Online Piracy Act." Usually copyright laws are fixed in the Congress in the dark of night - and "everyone else" does not find it worthwhile to pay the fixed cost of getting involved. The "Stop Online Piracy Act" was attempted in the same way, but the stakes were higher, organizations such as Wikipedia became involved in coordinating lobbying, and suddenly ordinary people started phoning and emailing their congress members. Rather than passing in the dark of night the "Stop Online Piracy Act" vanished into the dark of night.

Lobbying is advantageous for small groups for another reason. Lobbyists can focus on particular issues and this helps keep the stakes small enough that the small group of special interests has an advantage. By contrast the issues in elections tend to be broad and difficult for special interests to control. Hence elections tend to have higher stakes. The implications for direct democracy should be clear: by calling for a plebiscite over narrow issues special interests may be empowered - the opposite of what is ostensibly intended.

6.2. Why Pollsters are Wrong: The Uncertainty Principle in the Social Sciences

Physicists cannot predict the movement of a particle. Economists cannot predict market crashes. Likewise, political scientists cannot predict the outcome of elections. The failure of physicists has a name - "Heisenberg's uncertainty principle" - and as far as we know nobody criticizes physicists or obsesses over their failure. Economists and political scientists are much criticized for failing to forecast market crashes and elections. This is odd: the uncertainty principle is the foundation of quantum mechanics in which spooky particles seem to anticipate what other particles will do. The failure of economists and political scientists is for the much less spooky reason that people can and do anticipate what other people will do. There is no name for the failure of economists and political scientists: perhaps it will be more acceptable if we make it a principle? The "Lucas critique?" The "von Neumann principle?"

To understand why social scientists are necessarily unable to predict certain things let's start with something simple - the familiar game of rock-paper-scissors. As we know rock breaks scissors, paper wraps the rock and scissors cuts the paper. Suppose Jan and Dean are playing rock-paper-scissors and Nate interviews each of them. Jan tells Nate she is going to play rock and Dean tells Nate he is going to play scissors. Nate publishes his prediction on his website: Jan is going to beat Dean by playing rock to his scissors. They play the game: Jan plays rock and Dean - no fool he - plays paper and beats Jan. Oops...looks like Nate was wrong. As John von Neumann showed in 1928 there is only one solution to this paradox: Jan and Dean cannot know how the other is going to play - they must be uncertain. That uncertainty can be quantified: each must believe the other has one chance in three of playing rock, paper or scissors - or one of them is either stupid or wrong. There is no pure strategy equilibrium. Only if Nate announces that there is a 1/3rd chance of Jan and Dean each playing rock, paper or scissors will Jan and Dean be content to play as he forecasts.

No doubt some investors and voters are stupid and wrong - but most are not. Suppose that clever Nate discovers from his big data analysis that the stock market will crash next week. He announces his discovery to the world. Are you going to wait until next week to sell your stocks? Well nobody else is, so the market is going to crash today. Oops...looks like Nate was wrong again. Just like rock-paper-scissors the only prediction Nate can make that is correct and widely believed is a probabilistic one: For example, he can tell you that every day there is an .01% chance of a stock market crash - but he cannot tell you when the crash will take place. Just as the uncertainty principle underlies quantum mechanics so the fact that *people react to forecasts* is the basis of rational expectations theory in economics. And just as in the simple rock, paper scissors example this theory enables us to quantify our uncertainty.

So elections. As we have argued people vote for lots of reasons: out of civic duty, to register their opinion - and to help their side win. In 2012 voter turnout in swing states was 7.4% higher than in other states. Any analysis of elections must take into account that there are marginal voters who behave strategically - who only vote if they think there is a chance they might contribute to victory. If you are certain your party is going to lose are you more or less inclined to vote? If you are certain it is going to win? Many people - like those in the states that are not swing states - are less inclined to vote when they are confident of the outcome. So when Nate comes along and tells us that the Democrats are definitely going to win, what does the marginal Republican voter Dean do? Skips the vote. But Jan is no dummy, she realizes since Dean is not going to vote, she needn't bother either: her Democrats can win without her. But...Dean should anticipate Jan and vote and so bring his own party to victory. This is exactly the argument we gave proving that the all-pay auction has no pure strategy equilibrium. As we have shown in this chapter there is no solution to to the problem of strategic voter turnout that does not involve uncertainty about the outcome. This is an informal version of our formal proof that all-pay auctions require mixing.

Why are polls wrong? Because people lie to pollsters? Because people change their minds at the last minute? As we argued, by and large this is not the case - even in upset victories polls do a pretty good job of predicting how people are going to vote. What they do not do is do a good job of predicting who is going to vote - they do not predict turnout well. You read this all the time "this year turnout among Hispanic voters was unusually low" and so forth. You get the idea? We may know how many Democrats and Republicans there are and we may know that they are all going to vote for their own candidate: but if we don't know who is going to turn up at the polls we do not know who is going to win the election. And whether voters expect their party to win or lose changes whether they will bother to vote - so that voter turnout is subject to the von Neumann uncertainty principle.

Pollsters argue about their mistakes. Some understand that they do not do a good job of predicting turnout. Some - Sam Wang and his Princeton Election Consortium - made the ludicrous claim - based on "deep math" - that there was a 99% probability that Hillary Clinton would win the 2016 Presidential election. Nate Silver was more conservative giving her only a 73% chance of winning. But as far as we can tell, neither one realizes that the problem with their models is not something they can fix - that the reason that they do not predict the election is not because they do not have enough data, but because they cannot predict the election. Any forecasts of elections that do not take account of the von Neumann uncertainty principle are bound to fail.

That said: the von Neumann uncertainty principle is no more a statement that "life is uncertain" than the Heisenberg uncertainty says that we are unsure where cannon balls are going to land. For example: we know that if the stakes are very high in an election the large party will almost certainly win. We can make specific probability predictions about the chances of one side or the other winning. Moreover, the all-pay characteristic of voting plays a key role. If, for example, there is lobbying with a winner pays auction then we can predict the outcome.

6.3. A Brief History of Mixed Strategies

Choosing randomly? Does that sound realistic? The best selling book ever released by the RAND Corporation is their 1955 table of random numbers. Folklore has it that at least one captain of a nuclear submarine kept it by his bedside to use in plotting evasive maneuvers. As in rock-paper-scissors randomization is crucial when there is a conflict of interest between the players.

One familiar conflict is that of sporting events. The soccer player kicking a penalty goal must keep the goal keeper in the dark about whether he will kick to the right, to the left or to the center of the goal; the tennis player must be unpredictable as to which side of the court she will serve to, the football quarterback must not allow the defense to anticipate run or pass, or whether the play will move to the right or the left, and the baseball catcher must keep the batter uncertain as to how his pitcher will deliver the ball. Indeed, at one time in Japan baseball catchers were equipped with small mechanical randomization devices with which to call the pitch – this was later ruled unsporting and they were banned from play. Professional Bridge players claim that ability in playing mixed strategies makes the difference between a good and an excellent player. Economists have studied sporting events: empirical research (see for example Walker and Wooders (2001)) shows that in real contests - soccer matches, tennis matches - the good players in important matches play randomly and with the right probabilities.

At one time the idea of mixed strategies was not well understood. Sir Arthur Conan Doyle certainly did not understand mixed strategies: in his 1893 short story "The Final Problem" Holmes outwits Moriarity by getting off the train in Canterbury. Moriarity - despite his great intellect - foolishly believes that Holmes will attempt to go to Dover where he will inevitably be captured by Moriarity. There is some excuse for Doyle since the paradox was only solved by John von Neumann in 1928. Never-the-less ancient warriors - even if they did not understand the principle involved - did manage to randomize: the use of omens and oracles to plan military strategy is random indeed. Actually the modern theory of mixed strategies is based on the idea that explicit randomization is not needed at all. Harsanyi (1973) showed that allowing decisions to depend on an unimportant but real random shock to preferences can serve exactly the same purpose as explicit mixing.

What about groups then? Earlier we gave examples where social norms are slow to adjust. In the case of elections randomization seems to suggest that social norms adjust with lightning speed. However, the circumstances are different. Our discussion of social norms that are slow and costly to adjust was for shocks that could not reasonably be anticipated. The need for randomization in elections is hardly unanticipated and a social mechanism should have a contingent plan for dealing with shocks that are likely to happen.

The following interpretation suggests itself. Along the lines of Harsanyi (1973), and as is the case in practice, there is a small common shock to the preferences of party members that they observe, but the other party does not. We might call this "voter enthusiasm." Beforehand party members agree on a social mechanism: for each level of voter of enthusiasm they plan to implement a particular cost threshold for participation and a punishment level for a bad signal. We can think of the latter as a "temporary" social norm. In other words, the party chooses a contingent plan: how the temporary social norm depends upon voter enthusiasm. As Harsanyi showed, if the shock to preferences is relatively small, then the mixed strategy equilibrium is a good approximation to the equilibrium with voter enthusiasm shocks.

That mixing reflects the reality of elections can be seen by examining "GOTV" (Get Out The Vote) efforts. These efforts are an important part of establishing the social norm for the particular election, and indeed, these efforts are variable and strategic. Furthermore parties go to great lengths not to advertise their GOTV effort, and in fact to keep it secret. Accounts in the popular press document both the surprise to the other party of the strength of the GOTV and the secrecy that surrounds it. For example "The power of [Obama's GOTV] stunned Mr. Romney's aides on election night, as they saw voters they never even knew existed turn out..." Nagourney et al. (2012) or "[Romney's] campaign came up with a super-secret, super-duper vote monitoring system [...] to plan voter turnout tactics on Election Day " York (2012).

Note that the secrecy at issue is not over whether or not people voted as for example voting pins: we assume that the act of voting is observable. Rather the secrecy is over the temporary social norm that is enforced on election day. There is no reason to do that unless the GOTV effort is random. That is, the fact that it is secret provides evidence that - consciously or not - political parties engage in randomization when choosing temporary social norms for particular elections.

CHAPTER 7

Lobbying and the Agenda: Subsidies Versus Civil Rights[1]

In Becker (1983)'s classical work on political influence two groups compete over the size of a transfer from one group to the other. How do lobbying groups with control over an agenda determine the size of a transfer payment? From what we have learned so far we expect that a small group will choose an unambitious agenda so that it will be advantaged, while a large group will choose an ambitious agenda which works in its favor. However, we expect also that the nature of the prize will play a role. In the case of a non-rival prize such as civil rights increasing the size of the group increases the value of the rights. As was the case in the simple auction, we shall discover this strongly favors the large group which will indeed choose high stakes. By contrast consider farm subsidies or another prize involving money, goods or services. Here the size of the prize is equal and independent of group size. This means the prize is private - but in the context of lobbying the prize is also *fungible* in the sense that, unlike civil rights, it can be used to pay for the lobbying effort. As we shall discover this strongly favors the small group, that will however choose low stakes.

As we are interested in lobbying, we will consider only the case of a chore and not that of a duty. We examine the linear case in which $\theta = 1/2$. Moreover, we are going to consider the case where one group is an agenda setter that proposes a transfer from the other group to itself. In this analysis we shall make the natural assumption that if neither group makes a bid the status quo is maintained - that is, the non-agenda setter effectively wins.

As we observed in section 5.3 the politician generally prefers the winner pay auction while the groups are indifferent. This is likely why bribery rarely involves all-pay auctions, and from this point on we restrict attention to the winner-pay auctions in which the group with the higher willingness to bid wins and pays the smaller willingness to bid of the opposing group.

As before we continue to assume there are two groups, small and large. We now assume as well that one group, group a, controls the agenda. That is, it can choose the size of the prize.

7.1. Non-Rival Prize

To model a non-rival prize, we assume that each member of the non-agenda setting group $-a$ has ν utility units of a resource we shall call *rights*. Suppose $v \leq \nu$ units of rights are taken away from each member of $-a$. Each member of group a benefits from this loss of rights by group $-a$, that is rights are non-rival. We

[1]This chapter is based on Levine and Modica (2017).

assume that the worth of v to her is ζv where $\zeta < 1$: the benefit of taking someone else's rights is assumed to be less than the cost to the person losing those rights. In particular if a feasible agenda $v \leq \nu$ is selected then - letting η_a and η_{-a} be the group sizes - the value of the prize to $-a$ is $V = \eta_{-a} v$ and the value of the prize to a is $\eta_a \zeta v = (\zeta \eta_a / \eta_{-a}) V \equiv \beta V$. Notice that if the agenda setter is the small group then $\beta < 1$ meaning the transfer is inefficient, while if the agenda is the large group it is possible that $\beta > 1$.[2]

The formal setting then is that the agenda setter chooses a feasible prize $V \leq \eta_{-a}\nu$, then the two groups compete to provide effort costing $C(\varphi_k) = F + \gamma \varphi_k$ in a winner pays auction. We say that the agenda setter has a *winning agenda* if there is a choice of V for which the agenda setting group is advantaged, in which case we can speak of the optimal agenda as the one that maximizes the surplus the agenda setter gets from winning the auction.

The main result is that, as anticipated, if rights are under contention the large group will choose the highest feasible stakes and win:

THEOREM 7.1.1. *Only the large group may have a winning agenda, and has one if and only if $\zeta \nu > F + \gamma(\eta_S/\eta_L)$. If it has a winning agenda it will choose $V = \eta_S \nu$: it asks for and gets the most possible from the small group.*

PROOF. The desire to bid for the agenda setter a solves $\eta_a F + \eta_a \gamma \varphi_k = \beta V$ so $W_a = (1/\gamma)(\beta V - \eta_a F)$; similarly for the non-agenda setter $W_{-a} = (1/\gamma)(V - \eta_{-a} F)$. Both are increasing in V and the desire of the non-agenda setter increases more rapidly if $\beta < 1$. Define the *crossover point* $\hat{V} \equiv F(\eta_{-a} - \eta_a)/(1-\beta)$ as the point where the two desires are equal. We also define the *payoff point* $\tilde{V} \equiv F\eta_a/\beta$ as the point where the desire of the agenda setter is zero. To the right of this point the agenda setter may possibly wish to set an agenda, to the left of this point never. Recall that a winning agenda is a $V \leq \nu \eta_{-a}$ such that willingness to bid $W_a > W_{-a}$.

Case 1: $\beta < 1$.

To the right of the crossover point the non-agenda setter has a higher desire to pay. This means that if the constraints on his ability to pay do not bind he is at least as willing to pay as the agenda-setter. To the left of the crossover point the same is true of the agenda setter.

We first analyze the right of the crossover point, that is, $V > \hat{V}$. Here group a has a winning agenda only if the constraint binds on the non-agenda setter, that is $W_{-a} = \eta_{-a}$. Moreover since the bid of the non-agenda setter cannot increase once the constraint binds the agenda setter should propose the highest possible agenda, that is $V = \nu \eta_{-a}$. For this to be a winning bid it must be $W_a > \eta_{-a}$ which is impossible for $a = S$ (since $W_S \leq \eta_S$) and true for $a = L$ if and only if $(1/\gamma)(\beta \nu \eta_{-a} - F\eta_a) > \eta_{-a}$ which is equivalent to $\beta \nu > \gamma + F\eta_a/\eta_{-a}$. This is the same as the condition in the Theorem: $\zeta \nu > F + \gamma(\eta_S/\eta_L)$.

In case $a = L$ the crossover point $\hat{V} < 0$ so necessarily the optimum satisfies $V = \nu \eta_S > \hat{V}$ hence the large group has a winning agenda if and only if $\zeta \nu > F + \gamma(\eta_S/\eta_L)$ in which case it sets the agenda $V = \nu \eta_S$.

[2] There is a problematic aspect of welfare analysis in this case. It might be that each member of a minority loses two units of utility by being deprived of their rights, while each member of a majority receives one unit of utility by seeing the minority deprived of their rights. If the majority is more than twice the size of the minority then it is apparently efficient to deprive the minority of their rights. From a moral point of view this seems absurd.

In case $a = S$ the crossover point $\hat{V} = F(\eta_L - \eta_S)/(1-\beta)$ is positive, so we must also analyze what happens for $V < \hat{V}$. The small group will not propose any agenda below the payoff point $\tilde{V} = F\eta_S/\beta$. Since $\beta = \zeta\eta_S/\eta_L$ and $\zeta < 1$ it is easily checked that $\hat{V} < \tilde{V}$. Therefore there is no winning agenda for the small group below the crossover point, and we already saw that to the right of the crossover point the small group has no winning agenda.

Case 2: $\beta > 1$.

For this to be true it must be that $a = L$ so that $\hat{V} = F(\eta_S - \eta_L)/(1-\beta) > 0$. Now there is more rather than less desire to pay to the right of the crossover point. Hence to the *left* of $\hat{V}$ group L may have a winning agenda only if S is constrained. Since bigger V is better for L when S is constrained, it follows that if the optimal agenda for L is on the left of $\hat{V}$ it must be at $V = \nu\eta_S$. Similarly to the right of the crossover point group L's desire to pay rises faster than that of group S so it wants as large a transfer as it can, that is again $V = \nu\eta_S$. The maximal agenda is winning if $W_L \geq W_S$, which is to say that $(1/\gamma)(\beta\nu\eta_S - F\eta_L) \geq \eta_S$, that the same condition as in Case 1. □

7.2. Fungible Prize

We turn now to lobbying over a non-rival fungible prize: by this we mean that the proceeds of the prize can also be used to pay the politician. Each group member has a unit endowment that can be used either to lobby or to pay the other group. That means that if the prize is V the agenda setter is limited to $V \leq \eta_{-a}$ and that the individual endowment from which the bribe is payed is $1 + V/\eta_a$. Here the agenda setter gets βV where we focus on the case of inefficient transfers so that $\beta < 1$.[3] We continue to assume that $\theta = 1/2$. Since it takes a dollar to bid a dollar, we take the cost with a unit endowment to be $c_0 + c_1 y^i$ where $c_0 = 1$ is the endowment and $c_1 = 0$.

In the case of fungible prize the result is opposite to the previous one: if at all, the small group wins by choosing "low" stakes:

THEOREM 7.2.1. *Only the small group may have a winning agenda and has one if and only if $\beta > \eta_S/\eta_L$. If it has a winning agenda it chooses $V = \eta_L F$: it asks for and gets just enough to keep the large group from bidding, and receives utility $\eta_L F(\beta - \eta_S/\eta_L)$.*

If the transfer is too inefficient (β small) the status quo is maintained. This is the effect pointed out by Becker (1983): inefficient transfers are less likely to take place. In addition the small group is not too "greedy" in the sense that it asks only for $F\eta_L = \eta_L/2$ while it could ask for as much as $2F\eta_L = \eta_L$. Moreover, the amount that the small group wins $F\eta_L$ is increasing in the fixed cost F.

PROOF. From Lemma 4.4 we get $C(b_k/\eta_k) = 1/2 + (1/2)b_k/\eta_k$. Note that $F = c_0/2 = 1/2$ here. For the non-agenda setter we get $\eta_{-a}C_{-a}(b_{-a}/\eta_{-a}) = \eta_{-a}/2 + b_{-a}/2$ so the desire to bid from $V - \eta_{-a}C_{-a}(b_{-a}/\eta_{-a}) = 0$ is given by $B_{-a} = 2V - \eta_{-a}$. For the agenda setter we have cost $\eta_a C_a(b_a/\eta_a)/(1 + V/\eta_a) = \eta_a/2 + b_a/2$ since we must account for the fact that each participant now provides the endowment $1 + V\eta_a$; therefore the benefit of winning is $\beta V - \eta_a/2 - (1/2)b_a$ whence $B_a = 2\beta V - \eta_a$.

[3]For an efficient transfer government intervention may be unnecessary.

Suppose first that the agenda setter chooses not to constrain the non-agenda setter. In this case the willingness to bid $W_{-a} = B_{-a}$ so if the agenda setter matches the willingness to bid of the non-agenda setter the gain is

$$\beta V - \eta_a/2 - (1/2)B_{-a} = \beta V - \eta_a/2 - (1/2)(2V - \eta_{-a}) = (\beta - 1)V + (\eta_{-a} - \eta_a)/2.$$

This is decreasing in V since $\beta < 1$, so that the optimal choice is to force the non-agenda setter out by choosing $B_{-a} = 0$ that is $V = \eta_{-a}/2$. The gain from doing so is

$$(\beta - 1)\eta_{-a}/2 + (\eta_{-a} - \eta_a)/2 = (\beta\eta_{-a} - \eta_a)/2$$

which is positive if and only if $\beta > \eta_a/\eta_{-a}$; this can be true only if $a = S$, and then becomes the condition in the statement; the gain reads $F(\beta\eta_L - \eta_S)$.

Alternatively the agenda setter might constrain the non-agenda setter. The most the non-agenda setter can bid is $W_{-a} = \eta_{-a}$. This gives the agenda setter the most utility when the prize is as high as possible, that is $V = \eta_{-a}$. In this case, by bidding $b_a = \eta_{-a}$ the agenda setter gets $\beta\eta_{-a} - \eta_a/2 - \eta_{-a}/2 = \beta\eta_{-a} - 1/2 = (2\beta\eta_{-a} - 1)/2$. If $a = L$ then this is negative since $\beta < 1$ and $\eta_{-a} < 1/2$. Otherwise $\eta_{-a} \geq 1/2$ that is $a = S$, in which case it is better not to constrain $-a$ since $2\beta\eta_{-a} - 1 < \beta\eta_{-a} + \eta_{-a} - 1 = \beta\eta_{-a} - \eta_a$. □

7.3. Is Fixed Cost Plausible? The case of farm subsidies

In the fungible case it is the presence of a fixed cost per member that prevents a large group from being effective. But is the level of fixed cost needed to explain the data plausible? Consider the case of farm subsidies.

While there are only about 2 million farms in the US it is not just farmers that benefit from farm subsidies. An upper bound should be the rural population of the US of about 60 million people or roughly 20 million households out of the 120 million U.S. households - which is also about 15%. So we see that a minority of roughly 15% is effective at getting a fungible prize from the remaining 85%. That is, η_L is about 85% and η_S is about 15% of households, so that η_L is indeed much larger than η_S. In Table 5.4.1 the median level of farm subsidies measured in hours per person per year to the large group is 14, or about two working days. According the theory this should be equal to F the fixed cost of joining a lobbying organization - it seems a plausible number.

We should note also that a puzzle remains about the data in Table 5.4.1. In the theory smaller groups extract from each member of the large group a fixed amount F. In the data smaller groups actually extract more from each member of the large group the smaller they are. The theory does point us in a useful direction: the condition for the small group to have a winning agenda is $\beta > \eta_S/\eta_L$. If β is randomly drawn then smaller groups are more likely to succeed. Since farm subsidies are composed of many different pieces of legislation involving different lobbying efforts, the overall level of subsides is a composite of success and failures, so that countries with a smaller farm groups will on average have more successes and hence a higher level of subsidies.

7.4. Subsidies versus Civil Rights: the Facts

Do decisions favoring a group need substantial public support or is a limited public support sufficient? Our agenda setting model suggests that for fungible prizes widespread public support is not so important while for non-rival prizes it is. Two significant non-rival issues have been civil rights for blacks and civil rights for gays. In both cases significant advances have occurred when public support has become widespread. That is, when we talk about the group lobbying for rights we do not mean just those who directly receive the rights but all those who support those rights: while the fraction of blacks or gays may not change much over time those who support them does.

Long term polling by Gallup asks about willingness to vote for a black person for President, which may be taken as an indicator of general attitudes towards civil rights.[4] In 1958 only 38% responded positively, By 1959 this rose to about 50% where it remained until about 1963 when it rose to 60%, dipped briefly in 1967 and then rose steadily to about 95% by the year 2000. Civil rights have been largely reflective of these public attitudes towards blacks. The "separate but equal" doctrine permitting racial discrimination in a variety of domains, but most significantly in education was established in 1896 in Plessy v. Ferguson, and although it was repudiated in law in 1954 in Brown v. Board of Education, desegregation was not immediately implemented: George Wallace's stand in the school house door took place in 1963 - well after turn of public opinion, and the landmark legislation was the 1964 Civil Rights Act. Political action occurred only when the size of the group supporting civil rights became large.

We find a similar story with respect to gay civil rights. The Pew Research center finds that in 2003 only 32% of Americans favored same-sex legal marriage - this increased steadily, reaching parity by 2011.[5] From 1975 to 2000 various states and the Federal government passed a series of laws banning gay marriage. By 2009 only seven states had recognized gay marriage. This rose to thirteen by 2013 and to fifty with the Supreme court decision in 2015. Again the recognition of rights - non-fungible as it is - seems to have followed public opinion and indeed, majority public opinion.

By contrast if we look at an important fungible issue - farm subsidies - we see that support for large farms which receive the bulk of subsidies has only 15% popular support.[6] This is about the same as the calculation we made above of the fraction of households that plausibly benefit from farm subsidies. As we observed a minority of roughly 15% is effective at getting a fungible prize from the remaining 85%. This number 15% is similar to the fraction of the population that is either black or gay - yet those groups have been ineffectual in realizing the non-fungible prize of civil rights until they achieved the support of roughly a majority.

Other examples of fungible and non-fungible lobbying point the same direction. The police and prosecutors have a non-rival wish list of killing all civil liberties to make it easier for them to be nasty to whoever they choose however they choose. They have had fairly low success with this - their greatest success was the Patriot Act when they got a decent part of their wish list by pretending it had something to do with terrorism. On the fungible issue of civil forfeiture (stealing people's property by making accusations without evidence) police and prosecutors have had vastly greater success.[7]

---

[4] www.gallup.com/poll/3400/longterm-gallup-poll-trends-portrait-american-public-opinion.aspx

[5] pewforum.org/2015/07/29/graphics-slideshow-changing-attitudes-on-gay-marriage

[6] www.worldpublicopinion.org/pipa/articles/brunitedstatescanadara/602.php

[7] https://ij.org/report/policing-for-profit/

7.5. Minority and Majority Strategies

How can a minority hope to succeed? The answer depends on the nature of the prize. For fungible prizes lobbying may be an effective alternative to overcoming a voting majority. For rights it is not. However, there are two effective strategies for minorities to establish rights. One is direct action: since more is at stake they may have a greater willingness to engage in protests, for example, so that the resources they are willing to spend on direct action may overcome the resources the majority is willing to spend. While protests occur over many things, the most powerful and prolonged protests have been over issues of civil rights. Second is conversion: the relevant group is not just those that lack rights but those that agree that they should have them. By converting people to the point of view that rights are right, a minority may effectively command a majority. As the public opinion polls indicate - this was an effective strategy for both blacks and gays.

How can a majority hope to succeed? While they are advantaged in voting, the outcome of the election may be subverted by subsequent lobbying. Again, the nature of the prize matters. For fungible prizes it often seems the case that the majority is stuck with successful special interests "milking the cow." For non-fungible prizes - rights - the large party is also advantaged at lobbying so faces less of a problem. None-the-less we have seen that electoral outcomes are necessarily uncertain - small parties will sometimes win. When this happens we can expect active lobbying by the large group over rights. This seems to be the case - when Donald Trump won we saw a big increase in donations to both the American Civil Liberties Union[8] and Planned Parenthood.[9]

How can a politician hope to collect rents? In the non-rival case the politician gets a positive amount. This suggests that it can be lucrative for politicians to sell people down the river to take their rights away - and indeed disdain for minorities is one of the mainstays of "populism."

From a broader perspective, as economists and citizens we are interested in efficiency. It is often suggested that the grip of special interest lobbing over fungible prizes is to increase the cost of lobbying - for example, by putting legal restrictions on campaign contributions, by forcing lobbyists to register, and the like. However in the fungible case the small group gets $F(\beta\eta_L - \eta_S)$ (Theorem 7.2.1). Hence increasing the cost of lobbying increases the amount the small group is able to pick from the pocket of the large group. The point is that the cost of lobbying affects the large group as well as the small group. When the small group has the advantage, lowering the lobbying cost helps the large group overcome this advantage rather than encouraging the small group to lobby more.

The issue of imposing a cost on political participation is a tricky one. The problem is that efficiency demands that the right group win, and this is the group which values the prize the most regardless of whether it is large or small. Efficiency does not say that the majority should always win. We saw this earlier in section 5.3.1 where we discussed efficiency in the context of a non-fungible prize under the neutral assumption that there are no committed members ($\underline{\varphi} = 0$) and no fixed cost ($F = 0$). In that case we showed that lowering the marginal cost of

[8] https://www.huffingtonpost.com/entry/aclu-membership-skyrockets-trump_us_5b3db75de4b07b827cbd69b8

[9] https://www.thecut.com/2016/12/planned-parenthood-donations-are-up-40-fold-since-election.html

participation was inefficient because it advantaged the large group regardless of efficiency considerations. More generally, committed members advantage the large group and fixed cost the small group and this does not take account of which group values the prize the most. Hence the ideal contest is one with few committed members and a low fixed cost. In particular if the variable cost of participation is high (so that the large group is not inclined to take advantage of its size to outbid the small group) then we showed how the political contest can achieve efficient allocations.

Finally, both voting and lobbying waste the resources used to resolve the conflict. It would be better if political contests were resolved by doing something useful such as seeing who could pick up the most garbage rather than who voted the most. Since the all pay auction sometimes results in an inefficient outcome, it would be better still to use a winner pay auction and resolve the political contest by awarding office to whichever group credibly promised to pick up the most garbage.

CHAPTER 8

Crime and Punishment: The Determinants of Cost

For illustrative purposes we have analyzed a simple stark model. We have assumed that the high bid wins for sure; we have assumed a specific monitoring technology and generally worked with linear and/or quadratic functions. We have captured the key idea that monitoring cost increases as monitoring becomes more difficult and as the incentive to deviate gets larger. We have shed light on many of the key issues we raised in the introduction of the book. We now ask how robust these results are. In this chapter we examine in greater detail and generality the underpinning of total cost and monitoring cost. We will find that there are three robust messages from the analysis:

- The most profitable deviation is crucial for determining incentive compatibility and monitoring cost.
- When monitoring costs are taken into account total costs may be concave rather than convex.
- Higher monitoring costs and higher participation costs imply concavity hence favor the smaller group.

In the next chapter we will examine implementation and decentralization. Subsequently we will study more general contests.

The models we have studied have a common structure in which implementing a social norm has a monitoring cost $M = \theta G(\varphi)$ with θ is a measure of monitoring difficulty and G a measure of the gain to deviating. In the simple model without types $\theta = \pi/(\pi_1 - \pi)$; in the case of a public good $G(\varphi) = \varphi$, which is the cost of contributing, and in the case of the cartel $G(\varphi) = \mu(\varphi)(X - \varphi)$. With types we found $\theta = \pi/\pi_1$ and $G(\varphi) = (1 - \varphi)c(\varphi)$, see page 26. In all cases θ is increasing with π and decreasing with π_1. We now consider more general monitoring problems.

8.1. Hidden Actions and Small Deviations

We have so far assumed that the chances of a bad signal depend only on whether the social norm was violated. Broadly speaking this signal technology is a special case of the flexible information systems introduced in Yang (2020). The basic idea is that a variety of information systems are possible, and the one chosen should be sensitive to those deviations that matter. For example, in the design of a bond Yang (2015) shows that the optimal information system should be sensitive near the default boundary. In our setting of social norms in the form of quotas information systems should be sensitive to deviations from the quota. In the case of a discrete choice of whether or not to participate chances of a bad signal necessarily depend only on whether the social norm was violated, but in the case of a continuous effort choice it seems plausible that smaller deviations from the social norm would be harder to detect. We now turn to the study of flexible information systems with this property.

In the case of the cartel the key conclusion was that monitoring costs depend on capacity X: a greater capacity gives a greater incentive to cheat on the cartel and this raises the cost of enforcing incentive compatibility. However: if small deviations are hard to detect and they are appropriately discouraged, perhaps the punishment scheme is already adequate to deter large deviations? In this case X would not matter.

We start with a simple intuition: a large deviation is very easy to detect, hence we can punish a large deviation with probability near one. If an even larger deviation is available it will not result in appreciably higher probability of punishment, and hence the larger deviation demands higher punishment and this should lead to higher cost. Hence intuition suggests that the answer is that X should matter.

To formalize this idea we examine a simple effort provision model based on the cartel example. Individual group members produce output $0 \leq x^i \leq X$ and receive utility $\mu x^i - (\mu \overline{x} + \vartheta(\overline{x}))$ where the externality $\vartheta(\overline{x})$ is convex. Consequently, the cartel may wish to enforce a quota $0 \leq \varphi \leq X$ based on noisy signals of individual behavior. This is easier to analyze than the full cartel model because the marginal individual benefit of increasing output μ is now a constant.

Our basic assumption has been that the probability of a bad signal $\pi_b(|x^i - \varphi|)$ is a non-decreasing function of the absolute extent to which the social norm is violated $|x^i - \varphi|$. Up until now we assumed this was discontinuous: with $\pi_b(0) = \pi$ and $\pi_b(|x^i - \varphi_k|) = \pi_1 > \pi$ for any $|x - \varphi| > 0$. We refer to this as the *simple technology.*

Our first step is to consider the more general technology $\pi_b(x^i - \varphi)$ so that deviations above and below a quota are not necessarily treated in the same way. Indeed, we would not expect that negative values of $x^i - \varphi$ would increase the chance of a bad signal as the simple technology suggests. Instead it makes sense to assume that $\pi_b(x^i - \varphi)$ is non-decreasing. Notice that with the simple technology downward deviations do not matter as there is no point violating the social norm so as to reduce utility without decreasing the chance of punishment. This remains true if we assume that $\pi_b(x^i - \varphi)$ has the property of *left insensitivity*, that is, the chance of punishment remains constant and equal to π for negative values of $x^i - \varphi$.

8.1.1. Sources of Error.

There are a variety of assumption about continuity and left insensitivity of π_b that we can make. To understand what sort of assumptions make sense, it is useful to have concrete examples of the errors made by monitoring systems. Dutta et al. (2021) distinguish three different types of error. The first is a *gross error*, the second is a *measurement error*, and the third arises from *secret sales* which are common, for example, in cartels.

Let us start with a simple example of enforcing a speed limit with a radar system. *Gross error* is an error that is independent of the speed, for example the wrong car is identified by a license plate reader and the individual who receives the fine is not the person that committed the offense. Let us assume that the probability that the wrong car is observed is $0 < \pi/q < 1$ and the probability wrong car is speeding is $0 < q \leq 1$. Then the probability of gross error is π. As this is constant it is both left insensitive and smooth. In a context where the penalty is a fine which benefits other group members, as may be the case in cartels, members have an incentive to make false accusations about violations by other members, and this can be an additional source of gross error.

Measurement error is the usual type of error considered in economic models: if the actual speed is x^i the observed speed is $x^i + \tilde{\eta}$ where the random error $\tilde{\eta}$ is normal with mean 0. The radar system reports the driver if the observed speed exceeds the threshold φ. That is, a bad signal is received if $\tilde{\eta} > -\left(x^i - \varphi\right)$. If there is no gross error the bad signal occurs with probability $1 - H(-(x^i - \varphi))$ where H is the normal cdf. The overall probability of a bad signal is $\pi_b(x^i - \varphi) = \pi + (1 - \pi/q)\left(1 - H(-(x^i - \varphi))\right)$. This is quite different than the simple technology: it is smooth where the simple technology is discontinuous and it is not left insensitive.

By contrast, suppose that measurement error $\tilde{\eta}$ is uniform on $[-\gamma, \gamma]$ for $\gamma > 0$. Since φ is simply a benchmark against which the system report violations, we assume that here the radar system reports the driver if the the observed speed $x^i + \tilde{\eta}$ exceeds a threshold $\varphi + \gamma$. Letting $h = x^i - \varphi$, then $\pi_b(h)$ is continuous: it is constant and equal to π for $h \leq 0$, it is linear for $0 \leq h \leq 2\gamma$ and above that value the observed speed is surely above the threshold so π_b is constant and equal to $\pi + 1 - \pi/q$ for $h > 2\gamma$. In this case π_b is continuous (although not smooth) and is left insensitive.

Finally, we consider *secret sales.* A natural way to enforce a quota is to require transparency: that output or sales be done in such a way that they are easily observed. This is a common rule in cartels.[1] If a member adheres to the quota there is no reason not to comply with the transparency requirement. On the other hand if a member wishes to violate the quota then they will try to conceal their sales in order to avoid being punished. Hence the key monitoring problem is to determine whether or not secret sales took place. This naturally gives rise to left insensitivity: if the quota is adhered to no secret sales are made and negative signals reflect only gross errors, for example, false or mistaken accusations of making secret sales. If the quota is violated then secret sales take place: if a member is engaging in under-the-table transactions there is a chance word will leak out and they will be detected. In the simplest case the chances of getting caught, Q are independent of the number of secret sales. In this case for negative values of $x^i - \varphi$ we have π_b constant and equal to the gross error rate π while for $x^i - \varphi \geq 0$ we have $\pi_b = \pi + (1 - \pi)Q \equiv \pi_1$. This is left insensitive and since downward deviations do not matter with left insensitivity is equivalent to the simple technology. That is, we can think of the simple technology as a model of gross errors plus secret sales.

Generally speaking, however, we would expect that the more secret sales take place, the greater the chance of getting caught. We would also expect that there would be diminishing returns: as secret sales increase the chances of being caught increase at a decreasing rate. Secret sales are given by $h = x^i - \varphi$ if this is positive, so we can model the probability of being caught by a function $H(h)$ which is zero for $h \leq 0$, that may jump up at zero as there is some chance that word leaks out about under-the-table dealings, and is increasing and concave for $h > 0$ to reflect the increased chance of getting caught with diminishing returns. Allowing for gross error, the overall monitoring technology is then

$$\pi_b(h) = \pi + (1 - \pi/q) \cdot \mathtt{1}\{h > 0\} H(h).$$

Notice that this satisfies the property of left insensitivity but may be discontinuous.

---

[1]See, for example, Genesove and Mullin (2001).

Secret Sales. We first analyze the case of secret sales, assuming left insensitivity and that π_b is concave for $x^i - \varphi \geq 0$. In the discontinuous case the analysis is not so different than for the simple technology, so we analyze the case in which π_b is continuous. Since downward deviations do not matter, we may restrict attention to deviations $x^i - \varphi \geq 0$. Hence for any P the objective function $\mu x^i - (\mu \overline{x} + \vartheta(\overline{x})) - P\pi_b(x^i - \varphi)$ is convex so the optimum is either φ or X. Hence the only incentive constraint that matters is $\mu X - P\pi_b(X - \varphi) \leq \mu\varphi - P\pi$ from which we see that the optimal punishment is

$$P = \frac{1}{\pi_b(X-\varphi) - \pi}\mu(X-\varphi)$$

so that

$$M = \pi P = \frac{\pi}{\pi_b(X-\varphi) - \pi}\mu(X-\varphi).$$

Here

$$\theta = \frac{\pi}{\pi_b(X-\varphi) - \pi}$$

is decreasing in $X - \varphi$. Never-the-less M is increasing in $X - \varphi$, see the box below. Moreover, as $X \to \infty$ we must have $\pi_b(X - \varphi)$ converging to an upper limit, which we might as well call π_1 so that for large enough X we are basically in the discontinuous case.

Increasing Monitoring Cost. Differentiate M with respect $X - \varphi$ to to find

$$M'(X-\varphi) = \pi\mu\frac{(\pi_b(X-\varphi) - \pi) - (\pi_b(X-\varphi) - \pi)'(X-\varphi)}{(\pi_b(X-\varphi) - \pi)^2}.$$

Since $\pi_b(X - \varphi) - \pi$ is increasing from the origin and concave by the intermediate value theorem this is positive.

This is quite similar to our earlier analysis in that only the biggest deviation X matters and monitoring cost is strictly increasing in X. Notice that monitoring cost must be weakly increasing in X since it could hardly be the case that giving the individual more choices leads to a decrease. The key point is that it is not flat. This analysis reinforces our comparison of public goods with cartels.

Measurement Error. While secret sales are relevant when considering quotas, there can be measurement error too. This may lead to a failure of left insensitivity. Notice that unless π_b is constant it cannot be either concave or convex since no non-constant function bounded below on the real line is concave and no non-constant function bounded above on the real line is convex. Indeed, the boundaries force in a certain sense convexity to the left and concavity to the right. In this context it is natural to assume that there is a single endogenous inflection point ϕ and that π_b is convex to the left and concave to the right. In this context we write the probability of a bad signal as $\pi_b(x^i - \phi)$.

In this setting it may not be optimal, or even feasible, to choose the quota $\varphi = \phi$. Dutta et al. (2021) show in their Lemma 2 that under fairly weak assumptions it is always optimal to choose φ on the convex portion of π_b, that is, $\varphi \leq \phi$. Previously we assumed that π_b was flat until the quota was reached, then it became concave. In this case indeed the optimal $\varphi = \phi$.

We now study the opposite assumption. Assume that $\pi_b(x^i - \phi)$ reaches 1 and becomes flat at $x^i = \phi$ and is convex for $x^i < \phi$. In this case at $x^i = \phi$

the marginal benefit of deviating is $\mu > 0$ while the marginal cost is $P\pi_b'(0) = 0$ so that in fact $\varphi = \phi$ is not even feasible. The firm's objective function is $\mu x^i - (\mu\overline{x} + \vartheta(\overline{x})) - P\pi_b(x^i - \phi)$. We can now characterize the monitoring costs. Observe that $(1 - \pi_b(h))/\pi_b'(h)$ is a decreasing function for $h \leq 0$.

THEOREM 8.1.1. *For given φ let $M(\varphi)$ be the least monitoring cost of implementing φ. Then there exists an $\hat{h} < 0$ independent of φ such that for $X - \varphi \leq [1 - \pi_b(\hat{h})]/\pi_b'(\hat{h})$ cost is minimized at $\hat{h}$, any other cost minimizing $h > \hat{h}$ and $M = \mu\pi_b(\hat{h})/\pi_b'(\hat{h})$ is independent of X. For $X - \varphi > [1 - \pi_b(\hat{h})]/\pi_b'(\hat{h})$ we have $(1 - \pi_b(h))/\pi_b'(h) = X - \varphi$ so that h is strictly decreasing in X, and $M = \mu\pi_b(h)/\pi_b'(h)$ strictly increasing in X.*

The basic idea - that bigger X leads to greater monitoring costs - remains true, but only once X crosses a threshold. For X close enough to φ increasing X leaves monitoring cost unchanged.

PROOF. There are two incentive constraints, one that it cannot be optimal to deviate locally, that is to $x^i \leq \phi$, and second, since to the right the punishment is constant, that it is not optimal to deviate to X.

We first solve the problem ignoring the second constraint. We fix ϕ and solve for the optimal φ; inverting this solution then tells us for given φ what is the optimal ϕ. Given ϕ the objective function of the individual firm $\mu x^i - P\pi_b(x^i - \phi)$ is concave up to $x^i = \phi$, so if for some $\hat{h} \leq 0$ we set $P = \mu/\pi_b'(\hat{h})$ the function has a local maximum at $\varphi = \phi + \hat{h}$ Observe that for any h if $P = \mu/\pi_b'(h)$ the corresponding monitoring cost is

$$M = \pi_b(h)\frac{\mu}{\pi_b'(h)}.$$

This is continuous on $(0, \overline{h}]$ and from $\lim_{h\downarrow 0} \pi_b'(h) = 0$ and $\pi_b(0) = \pi > 0$ we see that the M-minimizing values of h exist, are bounded away from zero and are independent of μ, ϕ, X. We then take $\hat{h}$ to be the least optimal value of h with the corresponding largest $P = \mu/\pi_b'(\hat{h})$.

We must now consider the no deviation to X constraint. For this $P = \mu/\pi_b'(\hat{h})$ this reads $X - \varphi \leq [\pi_b(X - \varphi + \hat{h}) - \pi_b(\hat{h})]/\pi_b'(\hat{h})$. There are two cases. If $X - \varphi + \hat{h} < 0$ we can use the fundamental theorem of calculus and the convexity of π_b to see that $[\pi_b(X - \varphi + \hat{h}) - \pi_b(\hat{h})]/\pi_b'(\hat{h}) \geq X - \varphi$ so the constraint is satisfied. Hence the relevant constraint is $X - \varphi \leq [1 - \pi_b(\hat{h})]/\pi_b'(\hat{h})$ and when this holds we are done, and it suffices to choose ϕ such that $\phi + \hat{h} = \varphi$.

If the no deviation to X condition fails then the punishment $P = \mu/\pi_b'(\hat{h})$ is too small and we need to lower h to the point where the inequality $X - \varphi \geq [1 - \pi_b(h)]/\pi_b'(h)$ becomes an equality. Then with $P = \mu/\pi_b'(h)$ we get $\mu(X - \varphi) = P(1 - \pi_b(h))$ so that $x^i = \varphi$ is optimal. In this case $X - \varphi = (1 - \pi_b(h))/\pi_b'(h)$; therefore h is unique and strictly decreasing in X. It follows that M cannot be locally constant in X. Since increasing the range of deviations can scarcely lower monitoring cost it is non-decreasing in X and hence it must be strictly increasing.

Finally we must check that when $(1 - \pi_b(h))/\pi_b'(h) = X - \varphi$ it is actually the case that $\phi \leq X$. Observe that by convexity of $\pi_b$for $h \leq 0$ we have $1 - \pi_b(h) = \int_0^{-h} \pi_b'(h)dh \geq -h\pi_b'(h)$ or $(1 - \pi_b(h))/\pi_b'(h) \geq -h$. Then $X - \varphi = (1 - \pi_b(h))/\pi_b'(h)$ implies $X - \varphi \geq -h$ so that indeed $X \geq \phi$. $\square$

8.2. Hidden Types

Recall that in the model with hidden types, individuals draw types indexed by $y^i \in [0,1]$ and face a choice of participating at a cost of $c(y^i)$ increasing in y^i. A social norm is a threshold φ below which participation is required and above which it is excused. We assumed a simple monitoring technology in which participation is observed, but type is not: for a given social norm if $y^i < \varphi$ the probability of a bad signal is π_1 and if $y^i > \varphi$ the probability of a bad signal is $\pi < \pi_1$. Suppose instead we have a continuous probability of the bad signal $1 \geq \pi_b(y^i, \phi) > 0$ where the "target" ϕ lies in a closed interval and $\pi_b(y^i, \phi)$ is decreasing in y^i so that higher cost types are less likely to get a bad signal. Set $\pi = \min_{y^i,\phi} \pi_b(y^i, \phi)$. Since $\pi_b(y^i, \phi)$ is assumed to be positive and continuous we have $\pi > 0$. Monitoring cost can then be characterized:

THEOREM 8.2.1. *Monitoring cost $M(\varphi) = \theta(\varphi)(1-\varphi)c(\varphi)$ where $\pi \leq \theta(\varphi) \leq 1$ is continuous.*

This has two important implications. First, the key property of monitoring cost, that it is 0 at the endpoints $\varphi = \underline{\varphi}$, $\varphi = 1$ and strictly positive in between, does not depend on the details of the monitoring technology. Second, for any given value of π we see that the simple technology where a bad signal from non participation is certain below the social norm and has probability π above the social norm is the most efficient possible in the sense that $\theta(\varphi)$ achieves the lower bound, that is, $\theta(\varphi) = \pi$.

PROOF. For any given target ϕ and punishment P the social norm is determined by $c(\varphi) = \pi_b(\varphi, \phi)P$: for $y^i > \varphi$ participation cost is higher and the chance of punishment lower so it is optimal not to participate, and for $y^i < \varphi$ participation cost is lower and the chance of punishment higher so it is optimal to participate. Hence monitoring cost is given by

$$M(\varphi) = \min_{\phi} \frac{c(\varphi)}{\pi_b(\varphi, \phi)} \int_{\varphi}^{1} \pi_b(y, \phi) dy.$$

Using the bounds $1 \geq \pi_b(y^i, \phi) \geq \pi$ gives the result that

$$\left[\min_{y^i,\phi} \pi_b(y^i, \phi)\right](1-\varphi)c(\varphi) \leq M(\varphi) \leq (1-\varphi)c(\varphi).$$

Together with the fact that $M(\varphi)$ being the minimum of a continuous function is continuous this gives the result. □

8.3. Spillover Costs

We have assumed that the cost of punishment is borne only by the "guilty" party. In practice, however, the cost of punishment may spill over to other group members. The most common forms of punishment - some sort of exclusion, ranging from being denied the opportunity to participate in group events to imprisonment - will generally harm group members as well as the designated target of the punishment. For example, if Tim is punished by being excluded from joining the group at the bar after work then David suffers the loss of Tim's companionship. Or it may be that David feels sorry for Tim. In section 9.3 we consider a more detailed model of how this might work. It is an example of a *punishment cost spillover*. In computing monitoring cost we must account for the additional social cost from the spillover.

It is useful at this point to unify the hidden action and hidden type cases by defining $\Theta = 1/(\pi_1 - \pi)$ in the form case and $\Theta = 1/\pi_1$ in the latter case. Hence in both cases the incentive constraint may be written as $\Theta G \leq P$.

8.3.1. Negative Externality: Rotation, Expertise and Populism. Suppose that a punishment that costs P to the individual had an additional spillover cost of ψP to the group where $\psi \geq 0$, so that

$$M = (1 + \psi)\pi\Theta G.$$

Now $\theta = (1 + \psi)\pi\Theta$, that is, it increased by the the spillover cost.

The intuitive formula for monitoring costs is not without implications. We can think of ψ as a measure of social closeness between punishers and producers, while $\pi\Theta > 1$ measures the noisiness of the signal. These two variables are related, because by manipulating social interactions a group may be able to vary the social distance between the punisher and producer, but this leads in a natural way to a tradeoff: using punishers with greater social distance makes it less likely they will interact socially with the producer (lower ψ), but may also make it more difficult to accurately observe the production decision (higher $\pi\Theta$). Here we explore the implications.

One practical method of increasing the social distance between punisher and producer is the use of supervision rather than peer evaluation. In the literature on personnel management a great deal of attention is on which system provides the best incentives. Generally speaking we expect that peers interact with each other and supervisors interact with each other, but interactions between the two groups is less common - in other words supervisors have greater social distance than peers. Indeed, in some instances peers and supervisors are actively discouraged from interacting: for example in the military officers clubs used to be common to encourage officers to socialize with one another but not with enlisted ranks. We expect then that supervisor evaluation will deliver lower ψ - albeit at the cost of higher Θ. Indeed there is data - see for example Kraut (1975)[2] that indicates that peer evaluation is substantially more accurate than supervisor evaluation.

A second method of varying the social distance between punisher and producer is the system of rotation. In the police, for example, police officers may be periodically moved between precincts to deliberately break social ties. Rotation increases social distance because it makes police officers know their colleagues less well. Hence we expect that it will lower ψ - again at the cost of higher $\pi\Theta$. As in the case of supervisor evaluation a common complaint is that the effectiveness of evaluation is reduced as the evaluators have less interaction with and knowledge of the producers.

To study more clearly the trade-off between $\pi\Theta$ and ψ let us assume a trade-off in the form of a smooth *accuracy function*

$$\log \pi\Theta = Af(\log(1 + \psi))$$

[2]Most studies in this literature look only at the correlation between peer and supervisor rating or the within group correlation of rankings (“reliability”). Kraut (1975), by contrast, looks at peer and supervisor evaluations made at the end of a four week training course and shows that peer evaluation is a far better predictor of subsequent promotions.

where f is decreasing and convex. Here we interpret the parameter A as a deterioration in the monitoring technology, and we take ψ as a choice variable in the minimization of the monitoring cost. To focus thoughts and as an illustrative example we consider the police as a prototypical working class occupation and surgeons as a prototypical professional occupation. In both cases there is a substantial public goods output in the form of good reputation: corruption or excessive use of force by the police gives all police a bad reputation; lack of effort by doctors results in poorer patient outcomes, reduced demand for the services of doctors and less income for all doctors. In both cases group members have incentives to self-organize to reduce bad behavior.[3] We observe that in the police - as in working class occupations more generally - social distance to peers is low but supervisor evaluation is common and rotation is sometimes used as well. By contrast in professions such as the medical profession evaluation and punishment is done almost entirely by peers. Can this be explained by differences in the monitoring technology A between the two types of occupations?

What are the economic fundamentals - that is, what do we expect A to be in the two cases? For any given level of expertise it is more difficult to observe surgical output than police output: that argues that $A_S > A_P$ (S for surgeons, P for police). Compare, for example, improper behavior by police versus malpractice by doctors: from survey and other data the fraction of bad signals in response to bad behavior is greater for the police (about 3.2%, see Langton and Durose (2013)) than for doctors (less than 1%)[4] indicating better signal quality for police than surgeons. It is also the case that changes in social distance has a greater impact on the accuracy with which surgical output is observed than for police output. That is, increasing the level of expertise of the evaluator will make a great deal of difference for the accuracy with which surgical output is observed but not so for police output. This is a natural consequence of the fact that surgeons require a high level of specialized knowledge - more than a decade of specialized training[5] - while police officers require less than a year.[6] We interpret that to mean that outsiders are unlikely to have the specialized knowledge needed to evaluate surgical output while it is not so difficult for an outsider to evaluate police output. This also argues that $A_S > A_P$. Let $m = (1+\psi)\pi\Theta$ denote marginal monitoring cost.

THEOREM 8.3.1. *Suppose that f is an accuracy function (decreasing convex) and that $A_S > A_P > 0$. Then the minimizers of M are such that $m_S > m_P$ and $\hat{\psi}_S > \hat{\psi}_P$.*

PROOF. Write $\log M = \log m + \log G = \log(1+\psi) + Af(\log(1+\psi)) + \log G$. Hence minimizing M is equivalent to minimizing $\log m$. Since increases in A increase the objective function they increase its minimum value. As m is increasing in $\log m$ this gives the first result. Moreover this is a concave optimization problem since f is convex, so the optimum is determined by the first order condition $f'(\log(1 +$

---

[3] A common form of ostracism in the medical profession is to refuse to refer patients to other doctors: see Kinchen et al. (2004) and Sarsons (2017) who document that perceived medical skill is the most important factor in surgeon referrals and that bad surgical events lead to reduced referrals.

[4] In Group (2000) about 3% of cases where malpractice is documented in medical records lead to claims, while the actual incidence of malpractice is estimated to be 4 times higher.

[5] https://study.com/articles/Surgeon_Career_Summary_and_Required_Education.html

[6] http://work.chron.com/long-train-cop-21366.html

$\psi)) = -1/A$. Applying the implicit function theorem gives $d(\log(1+\psi))/dA = (1/A^2)/f''(\log(1+\psi)) > 0$. $\square$

The last part of the result tells us we should see greater social closeness in the monitoring of surgeons than of police. To understand more clearly the first part, that there is a higher marginal monitoring cost for surgeons, observe that this implies less output of any public good. We imagine that a social norm anticipates that there can be many different public goods that may be produced - this includes such things as timeliness and politeness. As a result for any particular problem surgeons will provide less public good. Roughly speaking, the first part of the theorem says that surgeons are chummy with their punishers and the second part says that they "get away with" more stuff than police.

The theorem seems to reflect reality. One form of public good, for example, is being on time: nobody can have failed to notice that doctors are never on time for appointments, while working class who are late for work are generally punished. These facts are relevant to the political analysis of populism. One of the root causes of populism is working class resentment of professionals.[7] One source of this resentment is the (correct) perception that professionals are laxly monitored: they are chummy with their evaluators and get away with more stuff. This is often attributed by the working class to the political power of elites. This analysis shows that instead it may be due to the different nature of signal accuracy.

Notice that while the monitoring of doctors may be optimal from the point of view of doctors, it isn't necessarily so from a social point of view: there is no reason to believe, for example, that the cost to surgeons of unnecessary surgery is as great as it is for the patients. In this case greater monitoring of doctors would be socially desirable. Our analysis does point to an appropriate remedy: not populism, whatever that means, but rather collective punishment for the professional class. What might be politically and practically feasible is hard to say - but, for example, a tax on all surgeons based on the number of fatal surgical accidents would encourage surgeons to tighten their self-regulation.

8.3.2. Limited Punishment. So far punishments are unlimited and their social cost is a linear function $(1+\psi)P$ of the size of punishment. In practice punishments come in many different sizes and shapes, and particular punishments are limited in the amount of utility cost they can inflict: slapping on the wrist is rather different than boiling in hot oil. Here we consider the implications of many different limited punishments.

We consider a set of punishment pairs $(p, s) \in \mathcal{P} \subseteq \Re^+ \times \Re^+$ where p represents a punishment and s a negative externality so that social cost is given by $p + s$. We make the minimal assumption that $\mathcal{P}$ is closed and that is contains $(0,0)$ so that no punishment is possible and has no social cost. In addition to specific punishments it is possible also to randomize over punishments. Hence the feasible space of punishment/externality pairs is the convex hull $\mathcal{H}$ of $\mathcal{P}$. Cost minimization requires that only punishments on the lower boundary of this set be used. Since $\mathcal{P}$ is closed this is given by a continuous convex function $\Psi(p)$ on an interval $[0, \overline{P}]$.

[7] See, for example, Williams (2016).

Limited or Unlimited Punishments? For technical reasons it is sometimes convenient to assume that unlimited punishments are available and $\overline{P} = \infty$. This is perhaps not literally true, but that does not matter a great deal. Very great punishments are available but have such great social costs they will never be used. Hence the exact size of $\overline{P}$ is not important. That is, in principle we could try to punish the failure of an individual to vote in an election by destroying the world with nuclear weapons - but of course it would never be a good idea to do so.

THEOREM 8.3.2. *The social cost function satisfies* $\Psi(0) = 0$ *and* $\Psi(p)$ *is strictly increasing for* $\Psi(p) > 0$.

PROOF. Since $(0, 0) \in \mathcal{P}$ it follows that $\Psi(0) = 0$. Suppose that $p' > p$. Convexity of Ψ implies that $(p/p')\Psi(p') + (1 - p/p')\Psi(0) \geq \Psi(p)$, so $\Psi(p') \geq (p'/p)\Psi(p) > \Psi(p)$. □

What does this tell us about monitoring costs? For incentive compatibility we require a punishment $P = \Theta G$. The social cost of this is $P + \Psi(P) = \Theta G + \Psi(\Theta G)$, and the probability the punishment is triggered on the equilibrium path is π. Hence for monitoring cost we have

$$M = \pi\Theta G + \pi\Psi(\Theta G).$$

The difference is that where previously we had a linear function of G now we have a convex function of G.

To understand the implications of this better, consider a simple example where there is a single punishment $\overline{P} > 0$ with $s = 0$ (so that there is no additional social cost beyond that of the punishment itself). This is a special case of convexity: M is flat up to $\overline{P}$ then vertical. By randomizing any punishment $0 \leq P \leq \overline{P}$ is feasible. Consider the simple public goods provision model and the possibility and cost of enforcing a quota φ. Suppose that if there is a bad signal there is only a chance λ that it is seen. In other words, assume that $\pi_1 = \lambda\overline{\pi}_1$ and $\pi = \lambda\overline{\pi}$. The incentive constraint is $P = \Theta G = (1/(\lambda(\overline{\pi}_1 - \overline{\pi})))\, G$. The corresponding monitoring cost is $M = \lambda\overline{\pi}\Theta G = (1/(\overline{\pi}_1 - \overline{\pi}))\, G$. The monitoring cost, in other words, is independent of λ. If there is a lesser chance of seeing the bad signal this is compensated for with a larger punishment issued less frequently. Hence - as long as $P \leq \overline{P}$ - the cost (and hence desirability) of enforcing φ does not change. However, if λ is sufficiently small that $P = (1/(\lambda(\overline{\pi}_1 - \overline{\pi})))\, G > \overline{P}$,that is, if $\lambda < \left(1/(\overline{P}(\overline{\pi}_1 - \overline{\pi}))\right) G$ then it is no longer feasible to enforce the quota φ at all, so in a sense the social norm abruptly breaks down.

Implications for Laboratory Experiments. It may well be that we have witnessed the effect of changing λ in the laboratory. It is well recognized that much of the "behavioral" phenomena observed in the laboratory is the implementation of social norms from outside the laboratory. If participants suspect that word of their behavior in the laboratory may leak out to peers outside the laboratory then they will wish to adhere to social norms while inside the laboratory. Here we may interpret λ as the chance that behavior in the laboratory leaks. This has been systematically varied in the laboratory - for example by using single or double-blind treatments and other promises of anonymity. What is interesting is that we typically see little response in behavior in response to changes in these promises - which is exactly what the above result predicts. Some extreme cases where great effort is made to reduce λ we do see less pro-social laboratory behavior which is consistent with the idea that once enforcement is no longer feasible bad behavior should be excused. See in particular Tisserand et al. (2015) for a discussion of the impact of double-blind treatments in dictator games.

8.3.3. Transfer Payments and Rewards. In addition to a negative externality from punishment, there can be a positive externality: it can be that the punishment to those with bad signals benefits those with good signals. For example, in a business firm the salary of those with bad signals can be reduced and the proceeds used to increases the salary of those with good signals.

A simple model is this. For any given punishment P issued an amount ψP is lost to the group where $0 \leq \psi \leq 1$ and the remainder is divided among those with good signals. When $\psi = 1$ the entire punishment is lost, which is our "standard" model. When $\psi = 0$ punishments have no net cost to the group and we would expect that there is no monitoring cost.

On the equilibrium path a fraction $1 - \pi$ of the population get a good signal and π a bad signal. Hence a good signal receives a reward of $(1 - \psi)P\pi/(1 - \pi)$. It follows that the net punishment or the punishment plus loss of reward, is

$$\tilde{P} = P + \frac{(1-\psi)\pi}{1-\pi}P = \frac{1-\pi+(1-\psi)\pi}{1-\pi}P.$$

It is this that figures into the incentive constraint, that is, incentive compatibility requires that the gain from deviating be no greater than the difference in signal probabilities times the net punishment: $G \leq (\pi_1 - \pi)\tilde{P}$.

To compute the monitoring cost, observe that a fraction of the population π is punished with P and each punishment costs the group ψP the amount that is lost during the transfer. Hence $M = \psi\pi P$. Solving the incentive constraint for equality and computing P as a function of $\tilde{P}$ then gives the monitoring cost

$$M = \psi\pi\frac{1-\pi}{(1-\pi\psi)(\pi_1-\pi)}G.$$

When $\psi = 1$ this is as before. As ψ decreases there are two effects. First, less of the punishment represents a loss so in particular if $\psi = 0$ there is no monitoring cost as we expect. Second, lower ψ improves the effectiveness of punishment by increasing the reward to good signals, hence less punishment is needed. This further lowers monitoring cost, but does so by interacting with π: the bigger is π, that is, the greater the monitoring difficulty, the greater the reduction in monitoring cost from lower ψ. Put differently: in general increasing π increases monitoring cost, but the effect is less strong the smaller is ψ.

The case $\psi = 0$ of pure transfer payments is a standard model in mechanism design where it is called budget balance. A fundamental result is that if the information structure is adequate to reveal in statistical terms what individuals did - called identification conditions - then the first best can be obtained: see, for example, d'Aspremont et al. (2003) and Fudenberg et al. (1995). In our context the first best means the absence of monitoring costs. As we have independent individual signals for each individual the identification condition is easily established and well known to be satisfied. The same idea underlies the folk theorem with public information: see Fudenberg and Maskin (1994). In that setting the key is to show that when the discount factor is high then it is "as if" cost free transfers are possible. However, as Myerson and Satterthwaite (1983) showed not all information technologies allow the first best to be obtained with budget balance so it can be optimal to "burn money," corresponding to socially costly punishments.

The bottom line is that transfers, if they are available, are good. In the cartel setting we often see transfers: in the form of fines or the buy-backs documented by Harrington and Skrzypacz (2011). Both fines and exclusion seem common in practice. Estimates of the fraction of cartels using exclusion[8] range from 5 to 27% and those using fines[9] from 4 to 64%. These however involve very different samples: Hyytinen et al. (2019) report that in their sample 27% use exclusion and 15% fines.

One reason to use exclusion is that transfers are not a panacea. First, transfers must ultimately be backed by some other more costly form of punishment - the punishment for refusing a transfer cannot simply be a larger transfer that will also be refused. For example, in Harrington and Skrzypacz (2011) transfers are backed by collective punishment for refusal to pay.[10] Second, transfer payments are generally not 1-1 - that is there are costs and inefficiencies of collecting fines or buybacks - as with taxes - and the value to the recipient will generally be less than the cost to the payer. Finally, transfers introduce malincentives for monitoring. Those who receive the transfers have an incentive to make false accusations.[11] Indeed, if any cartel member can secretly plant false evidence - switch a signal to a bad signal - then all will choose to do so and the signal will be useless. All of these considerations lead to the conclusion that while we may get a substantial reduction in ψ by using transfers, it is unlikely to be zero.

There are two key conclusions: first, if transfers are available they are likely to be used, and second, if transfers are available cartelization is more likely. In particular in the case of large cartels, a large gain from deviating may be offset by a low value of ψ. A case in point is that of the Consorzio Grana Padano, that owns the trademark for Grana Padano cheese in Italy. It is a reasonably large group consisting of about 200 producers and collects fees from farmers and monitors

[8] Hay and Kelley (1974), Frass and Greer (1977), Hyytinen et al. (2019).

[9] Posner (1970), Suslow (2005), Fink et al. (2017), Hyytinen et al. (2019).

[10] In their setting a firm refuses to pay by lying about output.

[11] We see this, for example, when local governments lower speed limits and reduce the length of yellow lights in order to increase revenue from traffic fines. Indeed the problem of false accusations is an ancient one - one element of the code of Hammurabi is punishment for false accusation - see Fudenberg and Levine (2006).

quality of cheese production before authorizing the use of the trademark.[12] The Consorzio was fined by the Italian Competition Authority (decision 4352) essentially for imposing fines on members producing too much.[13] The Consorzio has very good monitoring technology (they have inspectors on the floor of the producers) and they believed that their system of fines was legal. Hence in this case θ was quite small both because π is small and because ψ is small. The cartel activities were also very visible and as a result anti-trust action was effective. The general implication for anti-trust authorities is that since fines are easy to observe, illegalizing them should be an effective means of reducing cartelization.

8.3.4. Adulation with Hidden Types. The possibility of rewards has interesting implications in the case of hidden types. Consider a minor variation on the previous model where rewards are in the form of transfer payments: suppose that the group in addition to punishments has a budget $R \geq 0$ from which rewards can be issued. Just as people wish to avoid condemnation by their peers (punishment) so they also wish for the admiration of their peers (reward). Of course if we give everyone a gold star then gold stars cease to be a reward so it is natural to think in terms of a fixed budget for adulation or reward.

To get the most "bang for the buck" it is clear that the reward should be divided among those who participate. Hence with the social norm φ and costs $c(y^i)$ we see for $y^i \leq \varphi$ a participant gets $R/\varphi - c(y^i)$ for participating and $-\pi_1 P$ for not participating. The incentive constraint for the marginal group member is $R/\varphi - c(\varphi) \geq -\pi_1 P$. Now, however, even if $c(y^i) > 0$ so that there are no committed types, we see that for $\varphi c(\varphi) \leq R$ there is no need for punishment. Letting $\underline{\varphi} c(\underline{\varphi}) = R$ be the unique solution there is no cost of inducing participation up to $\underline{\varphi}$ - adulation alone is enough to do the trick, and the situation is the same as with committed types. For larger $\varphi > \underline{\varphi}$ we require the punishment $P = (c(\varphi) - R/\varphi)/\pi_1$ so that the impact of adulation on reducing the need for punishment is diminished at higher participation levels.

The point here is that adulation works well when limited participation is required, but is less useful when mass participation is required. So it is in practice as we admire the few who fight on behalf of the many, but of course when we all participate and get gold stars, gold stars are not worth so much.

8.3.5. Internalization of Social Norms. There is a simple variation on the idea of adulation: this is that individuals may have a preference for adhering to the social norm. This may be due to moral considerations as in Feddersen and Sandroni (2006) on which we base a simple model. We assume that there is a utility benefit B for adhering to a social norm. We refer to this as *internalization* of the social norm. The incentive constraint for the marginal member is then $-c(\varphi)+B \geq -\pi_1 P$. Hence $c(\underline{\varphi}) = B$ and for $\varphi > \underline{\varphi}$ we have $P = (c(\varphi) - B)/\pi_1$. Not surprisingly this is the same as a negative cost intercept c_0 for individual cost - which can be given exactly the same interpretation.

This simple model does not contain much new. As we shall indicate in the next section when punishments are costly there is potentially a more important role for internalization. Moreover, in social design problems where truth-telling may be

[12] See https://www.granapadano.it/public/file/201904USA-34282.pdf

[13] See Consorzio Grana Padano, 24 June 2004, n. I569, Bulletin 26/2004, and Siragusa et al. (2009).

possible only through indifference between reports internalization can give strict incentives for truth-telling - and can play an additional role if information is costly to acquire.

There is a fruitful research agenda still to be fulfilled on internalization of social norms. In this context it should be observed that groups invest heavily in convincing members to internalize social norms: propaganda, myth, and education are all used in this respect. There is a presumption in the extant literature that the purpose of communicating information is to communicate facts - for example, about an unknown state of the world. In practice communication frequently plays a role in communicating social norms and in convincing people to internalize them. We tell stories not necessarily because they are true but often because they reinforce social norms. Communication of this sort is often disparagingly called "virtue signalling" but those who throw the greatest accusations are often those most guilty. Regardless, the investigation of internalization as an endogenous cost that can be invested in is a promising area for research.

8.4. Continuous Signal

So far we have assumed that the signal is either that the social norm was violated or not, that is to punish or not to punish. In practice there can be stronger and weaker signals, so it might be a good idea to give more severe punishments for stronger signals of a violation. Here we examine what happens when there is a continuous signal.

We examine a simple case. We assume that either the norm can be followed or it can be violated resulting in a utility gain of G. Punishment, up to a maximum level of $\overline{P} < \infty$ is based on a continuous signal z which we may without loss of generality take to lie in the unit interval. If the norm is followed we assume this has a continuous density $f(z)$, which again without loss of generality, by choosing suitable units for z, we may take to be uniform. If the norm is violated we take the density of z also to be continuous and given by $f_b(z)$. By ordering the signals appropriately, we may assume that larger values of z are more likely when the norm is violated, this is called the *monotone likelihood ratio* assumption and here simply means that $f_b(z)$ is increasing in z. We restrict attention to the case where it is strictly increasing.

The problem is to choose a measurable punishment function $0 \leq P(z) \leq \overline{P}$ that is incentive compatible, so that $-\int_0^1 P(z)dz \geq G - \int_0^1 P(z)f_b(z)dz$ to minimize monitoring cost $M(G) = \int_0^1 P(z)dz$.

THEOREM 8.4.1. *The optimal strategy is to choose a cutpoint z^* give the maximum punishment $\overline{P}$ for $z > z^*$ and no punishment for $z < z^*$. Monitoring cost $M(G)$ is continuous, non-decreasing and convex, and $M(0) = 0$. Finally, as $\overline{P} \to \infty$ we have $M(G) \to \theta G$ where $\theta = f(1)/(f_b(1) - f(1))$.*

There are several points to be made here. First, it is indeed optimal either to punish a fixed amount or not punish at all. However the probability of punishment when adhering to the social norm and violating it π and π_1 are endogenous and depend on $\overline{P}$. Never-the-less the basic properties of monitoring cost - which is what matters for economic analysis - of increasing marginal cost with respect to the gain to deviating is preserved. Finally, while it has become common these days to think of a limit result as $\overline{P} \to \infty$ as meaning something like "$\overline{P}$ is nearly infinite" which is meaningless, what is actually means is more or less the opposite. What it means is that for large enough $\overline{P}$ to a good approximation $M(G)$ is θG and that increasing $\overline{P}$ further has practically no benefit in reducing monitoring cost.

PROOF. First, we observe that the space of measurable functions bounded by $0, \overline{P}$ on $[0,1]$ is compact in the weak topology and the objective being given by an integral is clearly continuous, so the problem has a solution. Suppose that $P(z)$ is a solution, and define $P_1 = \int_0^1 P(z) f_b(z) dz$. The solution must clearly minimize $\int_0^1 P(z)dz$ subject to $\int_0^1 P(z) f_b(z) dz \geq P_1$ since a smaller solution would be incentive compatible with lower cost. Consider the Lagrangean for the linear programming problem with Lagrange multiplier ν. The derivative of the Lagrangean with respect to $P(z)$ is $1 - \nu f_b(z)$ so that the optimum is to punish zero for $f_b(z) < 1/\nu$ and to punish $\overline{P}$ for $f_b(z) > 1/\nu$. Hence $z^*$is the unique solution of $f_b(z) = 1/\nu$. This give the cutpoint property.[14]

That monitoring cost is non-decreasing follows directly from the fact that if the incentive constraint is satisfied for G then it is satisfied for all $G' < G$ so an optimal scheme for G is incentive compatible for G' and yields the same monitoring cost. For convexity, suppose that $G' = \lambda G_0 + (1-\lambda)G_1$ and let P_0, P_1be optimal for G_0 and G_1respectively. Define $P' = \lambda P_0 + (1-\lambda)P_1$. Then $-\int_0^1 P_0(z)dz \geq G_0 - \int_0^1 P_0(z) f_b(z)dz$, $-\int_0^1 P_1(z)dz \geq G_1 - \int_0^1 P_1(z) f_b(z)dz$ implies $-\int_0^1 P'(z)dz \geq G' - \int_0^1 P'(z) f_b(z)dz$ so P' is incentive compatible for G'. Moreover, $M(G') \leq \int_0^1 P'(z)dz = \lambda M(G_0) + (1-\lambda)M(G_1)$.

That $M(0) = 0$ is obvious, and any non-decreasing convex function on the entire real line must be continuous there.

Define

$$\theta(z^*) = \frac{(1-z^*)}{\int_{z^*}^1 f_b(z)dz - (1-z^*)} = \frac{1}{\int_{z^*}^1 f_b(z)dz/(1-z^*) - 1}$$

and observe that $M(G) = \theta(z^*)G$. At the optimum we clearly require $f_b(z^*) > 1$, so $\int_{z^*}^1 f_b(z)dz/(1-z^*)$ is increasing in z^*. Hence as $\overline{P} \to \infty$ we should take $z^* \to 1$, implying that $\theta(z^*) \to 1/(f_b(1) - 1)$. □

8.5. Who Will Guard the Guardians?

We have allowed the possibility that punishment has a social cost: it may well have a private cost as well. Punishments must executed and that can be expensive to to the punisher as well as to the group. If David excludes Tim from parties at his house that is a social cost to other party goers who have no control over the invitations, but also a private cost to David who does. When punishments have a private cost incentives must be provided to carry them out. There are two ways in which such incentives can be provided. One discussed in the previous section 8.3.5 is that social norms may be internalized so that individuals suffer a psychic cost for violating them. This may work up to an extent, but few systems rely primarily on psychic costs and rewards. To provide real incentives the guardians must guard each other - members must be punished for failing to carry out punishments. This requires multiple rounds of monitoring and punishment. As the experiment in Fu et al. (2017) shows these additional rounds are relevant in practice. Even more

[14] This argument is from Fudenberg and Levine (2007).

notable is Skarbek (2014)'s work on prison gangs. He analyzes the written constitution of a particular gang. The constitution contains the self-referential idea of multiple rounds of punishment: it is the duty of each gang member to report and punish any member who fails in their duty to report and punish others. Indeed we are all familiar with the idea that to be "good" people we must take an active role in punishing the "bad." To tolerate heresy is to be heretical.

To understand how this works we examine a simple theory based on the foundational work of Kandori (1992) and Levine and Modica (2016). In the initial period $t = 1$ an initial punishment is $P_1 = \Theta G$ is required. Any punishment of P has a social cost of $(1+\psi)P$ but also a private cost of ΨP. By refusing to carry out the punishment group members may avoid this private cost. Hence subsequent rounds $t = 2, 3, \ldots$ of monitoring and punishment are required to give incentives to carry out punishments. In each round signals are generated about whether punishments were properly carried out: a bad signal is generated with probability Π_1 if the punishment is not properly executed and with probability $\Pi < \Pi_1$ if it is. Note that we do not assume the same monitoring technology as that for observing fulfillment of the quota φ since observing whether a punishment is properly executed is a different activity.

Incentive compatibility requires that to execute a punishment of P_{t-1} the punishers must be threatened with a punishment of at least $P_t = (\Psi/(\Pi_1 - \Pi))\, P_{t-1}$. Applying this recursively this gives

$$P_t = (\Psi/(\Pi_1 - \Pi))^{t-1}\, \Theta G.$$

Define $\Gamma \equiv \Psi/(\Pi_1 - \Pi)$. If $\Gamma \geq 1$ there is a Hatfield-McCoy feud situation in which punishments escalate over time. In this case φ cannot be enforced. By contrast, if $\Gamma < 1$ punishments diminish over time.

The social cost of punishment is $(1+\psi)P_t$ and on the equilibrium path this is triggered with initial probability π and subsequent probability Π. As the time periods over which punishment rounds take place are typically short we will ignore discounting and simply add the costs over time to get the monitoring cost. Presuming that $\Gamma < 1$ so that punishments are diminishing over time we see that

$$M(\varphi) = (1+\psi)\left(\pi P_1 + \sum_{t=2}^{\infty} \Pi P_t\right) = \Theta(1+\psi)\left(\pi + \Pi \sum_{t=2}^{\infty} \Gamma^{t-1}\right) G$$
$$= \Theta(1+\psi)\left(\pi + \Pi \frac{\Gamma}{1-\Gamma}\right) G.$$

To summarize:

.

THEOREM 8.5.1. *The quota can be implemented if and only* $\Gamma = \Psi/(\Pi_1 - \Pi) < 1$ *in which case to minimize the monitoring cost the group optimally chooses punishment probabilities*

$$P_t = \Gamma^{t-1}\Theta G.$$

The corresponding monitoring cost of punishment is

$$M(\varphi) = \Theta(1+\psi)\left(\pi + \Pi \frac{\Gamma}{1-\Gamma}\right) G.$$

There are several new points here:

- The punishment process itself must satisfy constraints that private cost Ψ be small relative to the increased probability of detection when cheating $\Pi_1 - \Pi$.
- It remains true that monitoring cost has the form $M = \theta G$. However perfect monitoring on the equilibrium path $\pi = 0$ no longer implies zero monitoring costs $\theta = 0$. If the cost of punishment is positive, $\Psi, \Pi > 0$ then certainly $\theta > 0$, and indeed if they are too large the quota cannot be implemented. The point is, even with a perfect signal in the initial stage, it is still necessary to monitor the punishers, and if signals about punishers are noisy costly punishments must be executed against them.

Endless Punishment? In implementing multiple rounds of punishments it does not make sense to have minuscule punishments that drag on for decades. Fortunately this is not necessary. Because punishment costs must decline exponentially, punishment rounds can be ended stochastically. For example, there might be a $1 - \Gamma$ probability that at each time $t > 1$ punishment rounds end entirely and forever and a Γ probability that another round will take place and if there is a bad signal a punishment of ΘG is issued. In this case punishment rounds do not go on forever but end in finite time with probability one.

8.6. The Cost Distribution with Hidden Types

With hidden types the concavity or convexity of costs depends upon two elements. One is the difficulty of monitoring. This we discussed extensively in section 4.3. When it is low, that is most weight is on direct cost, convexity is favored and when it is high, that is most weight is on monitoring cost, concavity is favored. The other element is the underlying distribution of costs, which we have so far assumed to be uniform. Here we examine the role of the cost distribution more closely: the key point is that having a high probability of low cost is similar to a duty and favors convexity of cost while high probability of high cost is similar to a chore and favors concavity.

We have modeled the direct cost by taking types y^i to be uniformly distributed on $[0, 1]$ and costs $c(y^i)$ an increasing function of type. While useful, this approach obscures the underlying distribution of costs. For this reason the alternative *distributional approach* is often used.[15] This assumes that rather than drawing a type each individual draws a cost c^i from a density function $g(c^i)$. These two approaches are equivalent and easily related as we now show.

In our approach the direct cost of a participation rate φ is given by $D(\varphi) = \int_0^{\varphi} c(y^i)dy^i$. To compute the direct cost for the distributional approach let $G(c) = \int_0^c g(\xi)d\xi$ be the cdf corresponding to $g(c)$. Then the fraction of members with cost less than or equal to ξ is given by $\varphi = G(\xi)$, and we can invert this to find $\xi = G^{-1}(\varphi)$. Hence the direct cost is given by $D(\varphi) = \int_0^{G^{-1}(\varphi)} cg(c)dc$, and cost as a function of type is

$$c(y^i) = D'(y^i) = G^{-1}(y^i).$$

This gives as well $M(\varphi) = \theta(1 - \varphi)c(\varphi) = (1 - \varphi)G^{-1}(\varphi)$. In other words these two approaches are equivalent with $c(y^i) = G^{-1}(y^i)$.

[15]For example, Feddersen and Sandroni (2006) or Coate and Conlin (2004a), among many others.

The distributional approach gives additional insight into the source of both costs and convexity as it is easy to understand what $g(c)$ means. We illustrate this with an example. For some $\alpha > 0$ assume the density function to be the power function $g(c) = \alpha c^{\alpha-1}$ on $[0,1]$. Hence low α means that cost is likely to be close to zero - low, while a high α means that cost is likely to be close to one - high.

To see how this plays out, the cdf is $G(c) = \int_0^c \alpha\xi^{\alpha-1}d\xi = c^\alpha$. and $G^{-1}(\varphi) = \varphi^{1/\alpha}$. Hence the direct cost is

$$D(\varphi) = \int_0^{\varphi^{1/\alpha}} c\alpha c^{\alpha-1}dc = \int_0^{\varphi^{1/\alpha}} \alpha c^\alpha dc = \frac{\alpha}{\alpha+1}\varphi^{(\alpha+1)/\alpha}$$

so $c(y^i) = \varphi^{1/\alpha}$, $M(\varphi) = \theta(1-\varphi)c(\varphi) = \theta(1-\varphi)\varphi^{1/\alpha}$ and

$$C(\varphi) = \frac{\alpha}{\alpha+1}\varphi^{(\alpha+1)/\alpha} + \theta(1-\varphi)\varphi^{1/\alpha} = \left(\frac{\alpha}{\alpha+1} - \theta\right)\varphi^{(\alpha+1)/\alpha} + \theta\varphi^{1/\alpha}.$$

We already examined the linear case $\alpha = 1$ where $c(y^i) = y^i$ corresponding to a uniform distribution on $[0,1]$. In that case we noticed that for the benchmark case $\theta = \alpha/(1+\alpha)$ the cost function itself was linear. If we make the same assumption for general α we get as a benchmark the *homogeneous cost* function

$$C(\varphi) = \frac{\alpha}{1+\alpha}\varphi^{1/\alpha}$$

which we see to be concave if $\alpha < 1$ and convex if $\alpha > 1$. What does this correspond to? Increasing α in the benchmark case has two effects. One is to raise monitoring difficulty θ. The other is to shift the distribution $g(c) = \alpha c^{\alpha-1}$ towards higher costs. Hence homogeneous costs enable us to increase small group advantage (or reduce disadvantage) by increasing α.

To see the link with previous findings recall from Theorem 4.2.1 that in the linear case with committed types or fixed costs we had global convexity if $F = 0$ (so $\underline{\varphi} \geq 0$) and $\theta \leq 1/2$ and global concavity if $F \geq 0$ (so $\underline{\varphi} = 0$) and $\theta \geq 1/2$. Here if $\alpha < 1$ costs are likely to be low: this is akin to a duty. That is, this is similar to $\underline{\varphi} \geq 0$ and $\theta = \alpha/(1+\alpha) \leq 1/2$. As we expect this gives global convexity. By contrast when $\alpha > 1$ and costs are likely to be high this is akin to a chore. That is, this is similar to $F \geq 0$ and $\theta = \alpha/(1+\alpha) \leq 1/2$. As we expect this gives global concavity.

CHAPTER 9

Implementation

The strength of mechanism design theory is that it abstracts from the details of how the mechanism is implemented. In general there are many implementations: the key fact is that they all lead to the same economic result. Never-the-less implementation is important, and questions about decentralized implementation especially so. It is not plausible, for example, that in a lobbying group or political party with many millions of members bad behavior by a single individual is somehow known by every member of the group. Monitoring and punishment are abstract in mechanism design theory and it is important to think by whom and how they are carried out. In this chapter we look at a series of examples illustrating how social mechanisms can be implemented in a decentralized way.

Beyond implementation, our foundational assumption is that groups are able to reach agreements that benefit the group as a whole. As we indicated, we wish it were not so: the world might well be a better place if special interest lobbying groups were less good at operating in their own interest. Never-the-less it is legitimate to examine how and why we think agreements can be reached.

9.1. Reaching an Agreement

It is crucial to emphasize that the world of social mechanisms we contemplate is a world of communication. We do not imagine that individuals sit in isolation in their homes, work out the best mechanism for the group, and then out of moral considerations all "do their bit" in the implementation.[1] We imagine that there is discussion, debate, argument, that consensus are reached and communicated. In large groups we expect that smaller groups and leaders play an important role in this process of communication.

Does in fact make sense to assume that group members can coordinate their effort to pick the most preferred alternative? Coordination games have been extensively studied in the laboratory and the famous experiment of Van Huyck et al. (1990) shows that despite common interests if there are many participants coordination may break down. However: the laboratory environment they study is very far removed from that of the groups we are studying. In the laboratory the participants are isolated and cannot communicate with one another. As indicated, the environment we are interested in is one where there is ample opportunity for communication and discussion - it is in this environment where "collective decision making" makes sense. And indeed there is support for this idea in the laboratory as well - the experiments of Cooper et al. (1992), for example, highlight the

[1]Something of the sort seems to be contemplated in the ethical voting literature such as Feddersen and Sandroni (2006).

role of communication in enabling coordination. Even very limited communication channels - allowing the exchange of messages - enables participants to achieve the efficient equilibrium. Even in the difficult environment of the Van Huyck et al. (1990) many players minimum game in which the least efficient equilibrium is highly robust we know from the work of Blume and Ortmann (2007) that when communication is allowed coordination on the efficient equilibrium occurs most of the time. Charness (2006) reach similar conclusions when group members can make non-binding promises: "We find that promises (or statements of intent) sent from agents to principals enhance trust, cooperation, and efficiency." Devetag and Ortmann (2007) find that along with communication shared experience also helps in achieving coordination, while Brandts et al. (2015) find that having leaders also enhances coordination and that having leaders who are viewed as "legitimate" helps the most.

One robust finding in the experimental literature is that when interests are too diverse coordination often breaks down.[2] Here we study groups in which members are *ex ante* identical so this is not an issue. Never-the-less it must be recognized that real groups may have substantial conflict which may make it difficult to reach an agreement. There is another issue as well: even if interests are common the mechanism may create incentives for subgroups to secretly collude. We argued in Section 5.5 that this was why in the face of fixed costs it is not easy to delegate action to a smaller group. It is also an issuing in monitoring: for example if Jan is assigned to monitor Dean and Dean to monitor Jan it may be relatively easy for the two of them to collude with each other not to file bad reports. For this reason we are careful to avoid this type of implementation.

9.2. Single Monitor Models

We consider a group with a large number N of members. Each member i chooses an amount of public good $x^i \in [0, X]$ to produce at unit marginal cost. As we are going to deal only with incentive compatibility and monitoring cost the objective function is irrelevant. We do assume for simplicity that individual group members ignore the fact that their own contribution to the public good returns some (very small if N is large) amount of the public good to themselves. We continue to assume that the group establishes a social norm in the form of a target level of output φ and that there is a noisy signal $z^i \in \{0, 1\}$ about whether member i respected the social norm where 0 means "good, respected the social norm" and 1 means "bad, failed to respect the social norm." If the social norm was respected ($x^i = \varphi$) the bad signal occurs with probability $\pi \geq 0$; if the social norm was violated ($x^i \neq \varphi$) the probability of the bad signal is at least as high $\pi_1 \geq \pi$. We now examine different models of how the signal is received and acted on.

A basic implementation is one of peer monitoring over a simple circular network. Here each group member i serves as a *monitor* for the member to his left $i-1$ and we identify group member 0 with group member N to make a circle. The signal about member i is received by the monitor $i+1$ who reports it to the group. Notice that if $N > 2$ then no monitor is monitored by the person they monitor.

More generally, let $I = \{1, 2, \ldots, N\}$ be the set of members. We can consider a matching of members to monitors by a map $m : I \to I$ with the convention that if a member is assigned to audit himself $m(j) = j$ then no audit takes place.

[2] See, for example, Crawford (1998).

A generalized matching process is then a random choice of m. Note that this formulation allows the possibility that a monitor is assigned to monitor several members. Alternatively we may assume that the map is 1-1 so that each monitor audits exactly one member.

We consider groups where there are separate subgroups of producers and monitors and the monitors are assigned randomly to producers.

9.3. Ostracism and Monitor Incentives[3]

Punishments to producers may be costly to monitors as well. In this case monitors must be provided with incentives to tell the truth. There are several ways in which this may occur. One is that monitors may have a mild preference for telling the truth. There is laboratory evidence that this is the case: see, for example, the meta-study Abeler et al. (2019). Alternatively as in section 8.5 there can be repeated rounds of monitoring in which the monitors themselves are monitored. Here we examine a third possibility. We model more closely the process of ostracism as a punishment and show how this also can be used to provide monitor incentives along with producer incentives.

For simplicity we now restrict public good provision to be binary, that is, $x^i \in \{0, X\}$ and assume that $\pi_1 = 1 - \pi$. We assume that monitors and producers are drawn from separate populations and that there is one monitor per producer. After production signals are observed by monitors who produce public reports (which may or may not be truthful).

After the reports are produced a social interaction takes place. The population is rematched into social subgroups of size $L \geq 4$ for a *social interaction,* with the probability that a producer is in the same subgroup as their monitor (and vice versa) equal to $h \geq 0$. Exactly one of the L members of each subgroup - the presenter - is chosen with equal probability to learn an interesting story and may volunteer to share it with the subgroup. The identity of the presenter is known and any report by or about the presenter is also known to the audience. The audience itself is anonymous, so, for example, if the presenter is a monitor the monitor does not know if the producer they reported on is in the audience.

If the presenter chooses to share, the remaining $L-1$ members - the anonymous audience - read the report by or about the presenter. In particular if the presenter is a producer the group observes her monitor's report even if the latter is in another group. Following this the group votes on whether the presentation will be allowed or whether the presenter should be ostracized. There is a number $1 < K < L-1$ such that the presenter is ostracized if and only if K or more members of the audience vote against the presentation. Prior to the vote the audience has access to a public randomizing device.

If the presentation takes place it provides a value of L to the presenter and to each member of the audience - so that the per capita value of a presentation is L in the social group. As there is a chance $1/L$ of being chosen to present, the *ex ante* expected (opportunity) cost prior to the speaker being chosen of being ostracized when chosen to present is then $L \cdot (1/L) = 1$. The ostracism decision is assumed to satisfy *audience anonymity* in that it is independent of the composition of the audience.

[3]This section is based on unpublished work with Rohan Dutta.

This framework allows for two types of mechanisms. One social mechanism - the *default mechanism* - is for no output to be produced (that is $x^i = 0$ all i), all stories to be volunteered and to ostracize nobody. The social utility from this is the (per capita) utility L from only the social interaction. An alternative is to try to provide incentives for producing output so that $\varphi = X$, the monitor tells the truth about the signal she observed and all stories are volunteered. The only tool for providing incentives is ostracism, that is sacrificing some part of the utility from the social interaction to encourage production. Notice that ostracism is costly for the monitor: if $h > 0$ the monitor shares part of the cost of ostracizing the producer and so has incentive to let the producer off the hook. Hence ostracism must be used also to provide incentives for the monitor.

9.3.1. Incentive Constraints. A key feature of the model is that any ostracism rule in which all members of the audience concur with each other is an equilibrium of that subgame: since $K < L - 1$ no individual audience member can force the presentation to take place when everyone else votes against and since $K > 1$ no individual can block a presentation agreed to by everyone else. In other words, there are Nash equilibria of the voting subgame in which ostracism takes place and in which ostracism does not take place. In particular it is subgame perfect to ostracize the speaker even though it is not Pareto efficient to do so. By making use of the public randomizing device the audience can randomize between the two types of equilibria. Hence potential mechanisms correspond to a probability P of ostracizing the producer for a bad report and a probability Q of ostracizing the monitor for a good report.

It is useful to understand the model through the incentive constraints. The incentive constraint for a monitor is simple. If it is to be optimal to tell the truth the monitor must be indifferent between reporting 0 and X. If a good report ($x = X$) is filed the monitor is excluded with probability Q. If a bad report ($x = 0$) is filed there is probability P that the producer is ostracized. Hence the incentive constraint for a monitor is

$$Q = hP.$$

Concretely, if the monitor is selected to make a presentation (probability $1/L$) and has made a good report then the audience randomizes with probability Q all voting for ostracism and with probability $1 - Q$ voting to let the presentation take place. If the monitor has filed a bad report there is a probability h the producer is in the social group, a probability $1/L$ that the producer is chosen to make a presentation and probability P that the producer is ostracized. In both cases the cost of the ostracism when it takes place to the monitor is L.

When a producer produces there is a direct cost X but also an indirect cost due to monitoring. The monitor (who reports truthfully owing to her incentive constraint) receives a bad signal with probability π resulting in a probability P that the producer is ostracized and a good signal with probability $1 - \pi$ resulting in a probability Q that the monitor is ostracized. Hence the overall expected cost to the producer of contributing is $X + (\pi P + (1 - \pi)Qh)$.

When a producer does not produce there is no longer a direct cost, but there still remain the indirect costs. This is $\pi' P + (1 - \pi')Qh = (1 - \pi)P + \pi Qh$.

The incentive constraint for the producer is that the cost of contributing must be no greater than the cost of not contributing: this can be written as

$$P - hQ \geq \frac{X}{1-2\pi}.$$

9.3.2. Implementation with Ostracism.

THEOREM 9.3.1. *If and only if the* implementation condition

$$\frac{X}{(1-2\pi)(1-h^2)} \leq 1$$

is satisfied can production be implemented. The ostracism probabilities are

$$P = \frac{X}{(1-2\pi)(1-h^2)}, \quad Q = hP.$$

The per capita cost of implementation is $X/2$ plus the monitoring cost

$$M = \frac{1}{2}\left[\frac{\pi + (1-\pi)h}{(1-2\pi)(1-h^2)}\right] X$$

so it is equal to

$$\frac{1}{2}\left[\frac{\pi + (1-\pi)h}{(1-2\pi)(1-h^2)} + 1\right] X.$$

We prove this result at the end of the section: first we comment on what it means and some comparative statics.

- The implementation condition is crucial: if it is satisfied then any sufficiently valuable public good will be produced, and if it is not no public good will be produced no matter how valuable.
- The implementation condition clearly implies that $X < 1$: that is, the cost of producing the good cannot be greater than the cost of being ostracized.
- Fixing the other parameters if h is sufficiently close to one the implementation condition fails. The reason for this is a feedback effect: a bigger punishment for the producer implies a bigger punishment for the monitor. The feedback effect is that the latter reduces the incentive for the producer to produce: by not producing she can reduce the probability the monitor is punished for sending a good report. A high degree of social interaction (h) makes this feedback effect very strong and consequently implementation becomes impossible.
- The only case in which implementation cost is zero is if both $\pi = 0$ so that the producer is not punished due to erroneous signals and $h = 0$ so that the monitor does not bear any of the cost of punishing the producer. Notice that if $\pi = 0$ and $h > 0$ we must punish the monitor for good reports even though that is the only kind submitted and they are known to be true: the reason is that the monitor must be willing to report the producer if the producer deviates from equilibrium.
- The only way to get the monitor to tell the truth is to make her indifferent between the two reports. This means that if there is even a small cost of monitoring it will never be paid: this can only be overcome if there is a mild preference for telling the truth.
- Malicious gossip is valued in the sense that a monitor is less likely to be ostracized for filing a bad report.

- The cost of implementation is proportional to X, the incentive to cheat on the social norm. This is our standard finding.

PROOF. Recall the incentive constraint for a monitor

$$Q = hP$$

and the producer

$$P - hQ \geq \frac{X}{1-2\pi}.$$

Since punishment is costly the producer constraint must hold with equality.

If we plug in the incentive constraint for the monitor $Q = hP$ into the incentive constraint for the producer we find the probability with which the producer must be excluded

$$P = \frac{1}{1-h^2}\frac{X}{1-2\pi}.$$

From this we see that if the punishment is feasible $P \leq 1$ if and only if the condition given in the theorem holds. Since $Q = hP$ and $h \leq 1$ it follows that if P is feasible, so is Q.

Finally we compute the minimum implementation cost. This is $X/2$ (since half the population produces) plus the monitoring cost M which is the ostracism probability times the cost of ostracism which is 1. So M is just the probability of ostracism, which since the presenter is a producer or a monitor with equal probability is given by

$$M = (1/2)\left[\pi P + (1-\pi)Q\right] = (1/2)\left[\pi + (1-\pi)h\right]P = \frac{1}{2}\frac{1}{1-h^2}\frac{\pi+(1-\pi)h}{1-2\pi}X.$$

From this the expression in the text follows. □

9.4. Multiple Monitors: Formal Versus Informal Monitoring[7]

If there are multiple auditors for each producer the incentive problem for auditors is greatly mitigated: if their information is perfectly correlated it is possible to compare the reports of the auditors and punish them for disagreeing. Here we examine another aspect of multiple auditors: the possibility that if their information is imperfectly correlated it may be possible to use multiple reports (many witnesses to a crime) to improve the quality of information.

To model multiple auditors we use a network approach. We assume that the network over which monitoring takes place is formed for reasons exogenous to monitoring and public goods production: farmers have social and trade relations with other farmers because they farm and because they live in farm communities and not in order to lobby for farm subsidies. So it is with bankers and many other groups.[8] However once the social network is formed - for whatever reason - it can be used for monitoring and this in turn can be used to enforce social norms for collective activity.

In our formulation we are given a network - for example of farmers - and the network members must decide how much to invest in social interactions with other network members - how much time to spend at the pub with neighboring farmers, say. Following this the provision of public goods must be determined - how much effort should the network spend in lobbying for farm subsidies, say?

---

[7]This section is based on unpublished work with Andrea Galeotti.

[8]By contrast secret police generally exist only for the purpose of monitoring.

Following the literature in sociology we highlight two aspects of informal interaction that seem to be key to understanding social norms. First, when a relationship is strong a lot of information is conveyed and monitoring is effective. Second, high density allows for information aggregation and better monitoring. Based on these two simple forces, we ask how density in a group affects social norms. In this setting a decision must be made on how much information to aggregate - that is how to audit. An interesting aspect of the network problem is that this generally depends on the local characteristics of the person being audited. Our main finding is that the relationship is not monotone. We say that monitoring is *formal* if the monitor must consult other people prior to reporting, and it is *informal* if she acts independently on his own signal without gathering information from others. When network ties are either very strong or very weak only informal monitoring is used: it is only in the intermediate case that we will find formal monitoring.

9.4.1. The Model. We examine a multi-stage game played by N individuals in a social network. In the first stage, each member of the network invests in the strength of social interactions with the other network members. We assume that members have preferences for variety and that investing in social interactions is costly but provides a direct benefit in the form of privately consumed output. In addition social ties can serve to transmit information in subsequent stages of the game. In the second stage the society collectively determines a social norm for public good provision. During the third stage a public good is produced from costly individual contributions. In the fourth stage of the game information about contributions is transmitted along social ties and it is used to punish free-riders and so provide incentives for public good contributions. Finally, in the last stage information transmitted along social ties is used to provide incentives for punishing deviators in the previous stage. We now provide the formal details.

The Social Network. Following Jackson (2010) the social network is an exogenous undirected graph denoted by $\mathbf{g}$ with nodes $i = \{1, ..., N\}$ referring to different members of the network. For a member i, the set N_i are i's social contacts - the set of members that have a link with i, and $k = |N_i|$ is the common degree of all network members. For any network member i there is a monitor denoted by $a(i)$ who is a social contact of i with the greatest possible number of mutual social contacts. Specifically $a(i)$ satisfies:

$$a(i) = \arg\max_{j \in N_i} |N_i \cap N_j|$$

Each mutual contact is assumed to know who the other mutual contacts are. We assume that all network members have the same number of mutual contacts with their monitors, which is equal to $\chi = |N_i \cap N_{a(i)}|$. Furthermore, if $a(i)$ is not unique, one of them is selected randomly. With some abuse of terminology we call χ the network *density*.

We will wish to relate k and χ to the number N of nodes or network members. It is useful to think of a physical region where people settle and form connections with other network members who live nearby. We define an exogenous threshold $\underline{N}$ such that if $N < \underline{N}$ we say that the region is *sparsely populated*. In this case, spreading people evenly through the region will result in few or no network connections. Since we imagine that people do not want to live that far apart, we assume in this case that $k = \underline{k} > 1$. If instead the region is *densely populated* so that $N \geq \underline{N}$, then people are forced to live near each other. In this case we assume that $k = \mu(N-1)$, where

$\mu \in (\underline{k}/(N-1), ..., 1)$ so that as the population increases the number of "nearby" people increases proportionality and so does the number of network connections. Finally, for convenience of notation define an integer $0 < \ell < \underline{k}$ such that $\chi = k - \ell$.

EXAMPLE 9.4.1. Suppose that the network lies on a circle of circumference R. Network contacts are those people within a distance d on the circle. Hence the fraction of the population who are contacts of a member on each side is d/R. It follows that $k = (2d/R)N$. The members with the greatest number of mutual contacts are the two individuals who are adjacent on the circle, and $\chi = k - 1$ (that is, $\ell = 1$).

Stage 1: Private Good Provision. Given the network $\mathbf{g}$, member i privately and unobservably chooses how much effort y_{ij} to invest in interacting with the jth social contact where $j \in N_i$. Investing in social interactions is costly but provides a direct benefit in the form of privately consumed output. We assume net utility of the private good ω_i follows a Dixit-Stiglitz technology

$$\Omega_i = \sum_{j \in N_i} \omega(y_{ij}) - \kappa(\sum_{j \in N_i} y_{ij}).$$

The utility function ω is smooth and has strictly diminishing returns with $\omega'(0)$ finite. We assume that $\kappa(\cdot)$ has constant marginal cost $\vartheta \geq 0$ up to a maximum available level of investment Y, that is, $\kappa(\sum_{j \in N_i} y_{ij}) = \vartheta \sum_{j \in N_i} y_{ij}$ for $\sum_{j \in N_i} y_{ij} \leq Y$ and, otherwise, $\kappa(\sum_{j \in N_i} y_{ij}) = +\infty$. Individual effort y_{ij} is known only to the individual exerting that effort - in particular it is not possible to punish a member in a subsequent stage for their choice of y_{ij}.

Stage 2: The Social Norm. The group collectively determines a social norm. This consists of a public good effort target φ, a target number of social contacts $\hat{\eta}$ a monitor is expected to consult with and a punishment P for failing to meet the effort target. For simplicity we assume that the behavior of monitors is perfectly observed so that there is no social cost of punishing them.

Stage 3: Public Good Provision. Each member i privately chooses a public good effort $x^i \in [0, X]$ representing the cost of contributing to the public good.

Stage 4: Monitoring. The contribution norm is enforced by monitoring behavior and punishing those who violate the social norm. Each social contact of a member i receives noisy private signals about the member's action: these signals depend entirely on whether the individual target was met $x^i = \varphi$ or not $x^i \neq \varphi$. Our basic assumption is that the number of signals that each social contact j receives is proportional to the strength of the relationship so that each social contact receives y_{ji} signals. These signals are observed without cost to the member. The monitor chooses a number $\eta_{a(i)}$ of mutual social contacts and receives their signals for a cost of b per signal.

Let $\overline{y}_i = \sum_{j \in N_i \cap N_{a(i)} \setminus a(i)} y_{ji}/(\chi - 1)$ be the average strength of ties of common social contacts excluding the monitor. We assume that it is the strength of the contacts' social ties to the *auditee* that determines signal strength and not the strength of the auditee's social ties to the contacts. This implies that network members have no incentive to manipulate the strength of their social ties in hopes of influencing the strength of signals that will be sent about them in the monitoring subgame.

We assume that the monitor receives a summary of $s = y_{a(i)i} + \eta_{a(i)}\overline{y}_i$ signals about the auditee. That is, s represents how many signals the monitor receives.

If output $x^i = \varphi$ meets the target the probability these signals are "bad" resulting in punishment P, is $q(s)$ and if output $x^i \neq \varphi$ the probability of a bad signal is $p(s) \geq q(s)$. We assume that $q(s), p(s)$ are smooth, that $q'(s) < 0, p'(s) > 0$, so that with more signals the chances of being punished when innocent declines and of being punished when guilty goes up, that $p(0) = q(0)$ so that no signal is uninformative and that as $s \to \infty$ we have $q(s) \to 0, p(s) \to 1$ and the signal becomes fully informative. Furthermore, define

$$\pi(s) = \frac{q(s)}{p(s) - q(s)}$$

and assume that $\pi''(s) > 0$; from our assumptions on $p(s), q(s)$ it follows also that $\pi'(s) < 0$. We assume moreover that $\pi'(0) > -\infty$. Finally letting $h(x) = \pi'(s)s$ we assume that $h(\infty) = 0$.

Example 9.4.2. [Normal] Suppose that s signals are received, each independent and normally distributed with variance σ^2, with mean $1 > 0$ if output meets the target, and mean -1 if not. Punishment is issued if the sum of the signals is negative. Hence $q(s) = \Phi(-\sqrt{s}/\sigma)$, $p(s) = \Phi(\sqrt{s}\sigma)$, where Φ denotes the standard normal cdf. It is easily seen that these functions satisfy our assumptions, and we have

$$\pi'(s) = -\frac{(1/(2\sigma))s^{-1/2}\phi(\sqrt{s}/\sigma)}{\left(\Phi(\sqrt{s}/\sigma) - \Phi(-\sqrt{s}/\sigma)\right)^2}$$

which is clearly increasing so that $\pi"(s) > 0$. We see as well that $h(\infty) = 0$.

An alternative simple signal technology is this

Example 9.4.3. [Exponential] Now suppose that s represents the amount of time spent looking for a signal that the social norm has been complied with. If output does not meet the norm, such a signal is never received, so $p(s) = 1$. If output does meet the norm the signal arrives with an exponential delay, so that the probability of getting a negative signal is $q(s) = e^{-s/\sigma}$. We may compute that $\pi(s)$ is not only convex, it is in fact log convex, that is

$$[\log \pi(s)]" = \frac{e^{-s/\sigma}}{\sigma^2(1 - e^{-s/\sigma})^2} > 0.$$

9.4.2. The Strength of Social Ties. We have assumed the individual effort is known only to that individual and that it has an effect on signals received, but not signals sent, so that from an individual point of view the only significance of individual effort is in producing the private output. Recall that this is

$$\Omega_i = \sum_{j \in N_i} \omega(y_{ij}) - \kappa(\sum_{j \in N_i} y_{ij}).$$

Because we have assumed that ω has strictly diminishing returns, the optimum is to allocate the same amount of effort $y_{ij} = y_i$ to every relationship, so that the objective function can be written as

$$\Omega = ku(y_i) - \kappa(ky_i).$$

Let $\overline{y}$ denote the unique solution of $\omega'(\overline{y}) = \vartheta$, then $y_i = \hat{y} = \min\{\overline{y}, Y/k\}$. We refer to $\hat{y}$ as the *strength* of a relationship. When $\hat{y} = \overline{y}$ we say that members have *strong* ties; when $\hat{y} = Y/k$ we say that the member has *weak* ties. Letting $\overline{k} = Y/\overline{y}$, we have that the member has strong ties if and only if $k \leq \overline{k}$. Given that all members choose the same strength of social ties on all links, we may write the signal received when $\hat{\eta}$ social contacts are contacted as $s = (1+\hat{\eta})\hat{y}$.

The key substantive point here is that network ties are chosen solely to maximize the benefits of the network for individual private good production: the fact that they may serve also to enable the production of public good is purely incidental.

9.4.3. When Is There Formal Monitoring? Since punishment for failing to meet the public good contribution target is the same regardless of how low the effort is, each member must prefer producing φ to producing 0. If the target is met, with probability $q(s)$ the member is punished anyway and receives utility $-\varphi - q(s)P$. If nothing is produced the utility is $-p(s)P$ hence the incentive constraint is $p(s)P \geq q(s)P + \varphi$. The equilibrium cost of punishment when everyone complies with the social norm is $q(s)P$ and this is minimized by choosing P as small as possible subject to incentive compatibility, that is $P = \varphi/(p(s) - q(s))$. Hence we can write the total cost of monitoring as

$$M(\hat{\eta}) = \pi(\hat{\eta}+1)\hat{y}\varphi + b\hat{\eta}$$

A necessary condition for the optimal mechanism is that $\hat{\eta}$ is chosen to minimize the total monitoring cost. In this setting there are two different types of institutions that may achieve this. If $\hat{\eta} > 0$ we say that there is *formal* monitoring - auditing is costly and the monitor must consult other people. If $\hat{\eta} = 0$ then the monitor acts independently on his own signal without gathering information from others, auditing is costless, and the total cost of monitoring is $\pi(\hat{y})\varphi$. In this case we say that there is *informal* monitoring. With this terminology we want to capture how "formally organized" is the monitoring system: an informal monitoring is executed with the involvement of only the monitor, whereas a formal monitoring requires the participation of most of the social contacts.

THEOREM 9.4.4. *If $-\pi'(\hat{y})\hat{y} \leq b/\varphi$ optimal monitoring is informal. Otherwise, denoting by $\overline{\eta}$ the unique solution to $\pi'((\eta+1)\hat{y})\hat{y}\varphi + b = 0$, optimal monitoring is formal and given by $\hat{\eta} = \min\{\overline{\eta}, \chi\}$. If $0 < \hat{\eta} < \chi$ then*

$$\frac{d\hat{s}}{d(\hat{y}\varphi)} = -\frac{\pi'(\hat{s})}{\pi''(\hat{s})\hat{y}\varphi} > 0.$$

otherwise $d\hat{s}/d\varphi = 0$ and $d\hat{s}/d\hat{y} > 0$. As $\hat{y}\varphi \to 0$ and $\hat{y} \to 0$ we have $\hat{s} \to 0$.

PROOF. Monitoring cost is given by $M = \pi((\eta+1)\hat{y})\varphi + b\eta$ and η must be chosen to minimize this cost. The derivatives with respect to η are $M' = \pi'((\eta+1)\hat{y})\hat{y}\varphi + b = 0$ and $M'' = \pi''((\eta+1)\hat{y})\hat{y}^2\varphi$. The latter is by assumption strictly positive so there is a unique solution given by the first order condition: either the boundary case where $-\pi'(\hat{y})\hat{y} \leq b/\varphi)$ or the interior case where $\pi'((\overline{\eta}+1)\hat{y})\hat{y}\varphi + b = 0$ if $\overline{\eta} \leq \chi$ so that this is feasible, and χ otherwise. The derivative of the optimal signal $\hat{s} = (\hat{\eta}+1)\hat{y}$ follows from the implicit function theorem.

Finally from $M' = \pi'(\hat{s})\hat{y}\varphi + b = 0$ we see as $\hat{y}\varphi \to 0$ we must have $\pi'(\hat{s}) \to -\infty$ which is possible only if $\hat{s} \to 0$, or if $m' = \pi'(\hat{s})\hat{y}\varphi + b = 0 > 0$ then $\hat{s} = \hat{y} \to 0$. □

COROLLARY 9.4.5. *There exists a $0 < \underline{y} \leq \overline{y}$ such that for $\hat{y} < \underline{y}$ and $\hat{y} > \overline{y}$ monitoring is informal. For fixed $\hat{y}$ if b/φ is sufficiently small there is formal monitoring.*

PROOF. The first follows from the fact that $\pi'(0) > -\infty$. The second follows directly from the assumption that the continuous function $h(s) = \pi'(s)s$ satisfies $h(\infty) = 0$ and Theorem 9.4.4. The third directly from Theorem 9.4.4 $\square$

Roughly speaking, if ties are very strong there is no reason to consult with others, while if they are very weak it is not worth paying the cost for the negligible additional information.

CHAPTER 10

Games Between Groups[1]

Groups choose mechanisms to enforce social norms, and they often do so in competition with one another. One form of competition is an auction, but it is certainly not the only type of competition we see between groups: in the next chapter we will examine contests more general than auctions. Here we consider more broadly the idea of games between groups. For notational and conceptual simplicity we restrict attention to finite games. Because groups, unlike individuals, have incentive constraints, games between groups are different than games between individuals. Earlier work by Myerson (1982) and Dutta et al. (2018) show that if we define an equilibrium as the best incentive compatible mechanism for each group given the mechanisms of the other groups then they may fail to exist. They propose a modified definition of equilibrium: quasi-equilibrium in Myerson (1982) and collusion constrained equilibrium in Dutta et al. (2018). We briefly review a counter-example and then show that if adequate punishments are available the problem does not arise, equilibria in the ordinary sense exist, and they can be analyzed in a straightforward way. Once again the tool of monitoring cost places a key role in the analysis.

10.1. The Model

We study multiple groups $k = 1, \ldots, K$. Each group chooses a social norm $\varphi^k \in A^k$ a finite set. Individuals i in group k have deviations $\phi^i \in \Phi^k$ a finite set with $0 \in \Phi^k$ having the interpretation "adhere to the social norm." Utility of individual i in group k is given by $u^k(\phi^i, \varphi^k, \varphi^{-k})$. Crucially, groups have monitoring technologies for detecting whether or not $\phi^i = 0$. Specifically there is a probability of a bad signal $\pi_1^k(\varphi^k)$ if $\phi^i \neq 0$ and $\pi^k(\varphi^k) < \pi_1^k(\varphi^k)$ if $\phi^i = 0$. Finally, punishments P^k up to $\overline{P}^k$ are available with social cost equal to punishment value.

To define the game between groups we specify a *strategy* σ^i for group k as a non-negative function $P^k(\varphi^k)$ with a probability density f^k over social norms A^k. A strategy profile σ is *incentive compatible* if for each group $u^k(\phi^i, \varphi^k, f^{-k}) - P^k(\varphi^k)\pi_1^k(\varphi^k) \leq u^k(0, \varphi^k, f^{-k}) - P^k(\varphi^k)\pi^k(\varphi^k)$ for every deviation $\phi^i \in \Phi^k$ and for each φ^k with $f^k(\varphi^k) > 0$. An *equilibrium* consists of a strategy profile σ that is incentive compatible and such that for each group k and for each $\tilde{\varphi}^k, \tilde{P}^k$ that is incentive compatible with respect to f^{-k} we have for $f^k(\varphi^k) > 0$

$$u^k(0, \tilde{\varphi}^k, f^{-k}) - \pi(\tilde{\varphi}^k)\tilde{P}^k \leq u^k(0, \varphi^k, f^{-k}) - \pi(\varphi^k)P^k(\varphi^k).$$

[1] This chapter is based on Dutta et al. (2018).

10.2. Equilibrium May Not Exist: an Example

We start with an example adapted from Dutta et al. (2018) in which $\overline{P}^k = 0$ so monitoring and punishment are irrelevant. There are two groups $K = 2$ and each has two social norms C, D. Each individual in group 1 has one non-trivial deviation $\phi = 1$ and group 2 has none. Utility when there is no deviation is given by

	C	D
C	$6,5$	$10,0$
D	$2,0$	$2,5$

In group 1 the utility from deviating to $\phi = 1$ is

	C	D
C	8	8
D	0	0

Let μ^k denote the probability with which group k plays C. What is incentive compatible for group $k = 1$? The gain to deviating is given by

$$\begin{array}{ll} \text{when the group is playing } C: & 2\mu^2 - 2(1-\mu^2) \\ \text{when the group is playing } D: & \qquad -2 \end{array}$$

from which we see that D is always incentive compatible and that C is incentive compatible if and only if $\mu^2 \leq 1/2$. Since C dominates D it should be used whenever incentive compatible, that is when $\mu^2 \leq 1/2$. Otherwise D must be used.

For group 2 every social norm is incentive compatible so it is optimal to choose C if $\mu^1 \geq 1/2$ and D if $\mu^1 \leq 1/2$.

We now have the following problem. If $\mu^2 \leq 1/2$ then optimality for group 1 implies that $\mu^1 = 1$. This implies that $\mu^2 = 1$, a contradiction. On the other hand if $\mu^2 > 1/2$ then optimality for group 1 implies that $\mu^1 = 0$. This implies that $\mu^2 = 0$, another contradiction. There is no equilibrium.

There are two "solutions" to this problem. One is to follow the lead of Myerson (1982) and Dutta et al. (2018) and modify the definition of equilibrium to allow the possibility of "shadow mixing" where groups can mix onto strategies to which they are not indifferent. In this example we would allow group 1 to respond to $\mu^2 = 1/2$ by mixing onto the inferior plan D because C is not incentive compatible for small changes in μ^2. Dutta et al. (2018) show that such "shadow mixing" equilibria well capture equilibria of more realistic settings in which perceptions of opponents strategies are not so sharp. They also show that if the punishment technology is "adequate" then the problem does not arise. Since that is the relevant case for social mechanisms we outline that approach here.

10.3. Monitoring Cost and the Reduced Game

We define the *gain function* in group k as

$$g^k(\varphi^k, f^{-k}) \equiv \max_{\phi^i} u^k(\phi^i, \varphi^k, f^{-k}) - u^k(0, \varphi^k, f^{-k}).$$

DEFINITION 10.3.1. We say that the *punishment technology is adequate* if

$$\max_{\varphi^k, f^{-k}} \frac{g^k(\varphi^k, f^{-k})}{\pi_1(\varphi_k) - \pi(\varphi_k)} \leq \overline{P}.$$

This requires that it be possible to issue sufficiently large punishments: it means that, at some cost, any social norm can be made incentive compatible. It is this assumption that fails in the counter-example.

When the monitoring technology is adequate we can define the monitoring cost

$$M^k(\varphi^k, f^{-k}) \equiv \frac{\pi(\varphi^k)}{\pi_1(\varphi_k) - \pi(\varphi_k)} g^k(\varphi^k, f^{-k}).$$

We can then define the *reduced game* between the groups in which the strategies are simply mixtures f^k over social norms and the payoffs to a pure strategy φ^k by k and a mixture f^{-k} by $-k$ is given by $u^k(0, \varphi^k, f^{-k}) - M^k(\varphi^k, f^{-k})$. For such a game we can speak of Nash equilibrium in the usual sense.

THEOREM 10.3.2. *If the punishment technology is adequate a profile f is a Nash equilibrium of the reduced game if and only if there are functions $P^k(\varphi^k)$ that together with f are an equilibrium of the original game. An equilibrium exists.*

PROOF. All the incentive constraints satisfied if and only if

$$P^k \geq \frac{g^k(\varphi^k, f^{-k})}{\pi_1(\varphi_k) - \pi(\varphi_k)}.$$

Hence, since the punishment technology is adequate, in any equilibrium we may choose

$$P^k(\varphi^k) = \frac{g^k(\varphi^k, f^{-k})}{\pi_1(\varphi_k) - \pi(\varphi_k)}$$

and the corresponding utility is given by $u^k(0, \varphi^k, f^{-k}) - M^k(\varphi^k, f^{-k})$. This shows the equivalence between Nash equilibria of the reduced game and equilibria of the original game.

For existence we look to the reduced game. This does not follow directly from existence of Nash equilibrium in finite games since $M^k(\varphi^k, f^{-k})$ need not be linear in f^{-k}. However it is continuous in f^{-k} and that is all that is required for the standard Kakutani fixed point argument to go through. $\square$

10.4. Monitoring in the Example

We re-examine the counterexample in section 10.2 assuming now that $\overline{P}^k = \infty$. Recall that for group 1 if $\mu^2 \leq 1/2$ then C is incentive compatible without punishment so is chosen, while group 2 has no incentive problem, so chooses C if $\mu^1 > 1/2$ and D if $\mu^1 < 1/2$ and is otherwise indifferent.

The extant incentive problem is when $\mu^2 > 1/2$ in which case the gain to deviating is given by $g^1(\mu^2) = 4\mu^2 - 2$. Hence when $\mu^2 > 1/2$ a punishment $P^1(C) = \left(4\mu^2 - 2\right) / (\pi_1 - \pi)$ should be used with corresponding social cost $M^1 = \pi\left(4\mu^2 - 2\right) / (\pi_1 - \pi)$. When playing C members of group 1 get $6\mu^2 + 10(1-\mu^2) - \pi P^1(C)$, and with D they get 2 (no punishment there). Hence C will be chosen when $6\mu^2 + 10(1-\mu^2) - \pi P^1(C) > 2$ that is $8 - 4\mu^2 > \pi\left(4\mu^2 - 2\right) / (\pi_1 - \pi)$; when the reverse is true D will be chosen and when equality holds group 1 will be indifferent. As there cannot be a pure strategy equilibrium we must have $\mu^1 = 1/2$ and $8 - 4\mu^2 = \pi\left(4\mu^2 - 2\right) / (\pi_1 - \pi)$ or

$$\mu^2 = \frac{8\pi_1 - 6\pi}{4\pi_1} > \frac{1}{2}.$$

Observe that as $\pi \to \pi^1$ we have $\mu^2 \downarrow 1/2$ which is the shadow mixing equilibrium we identified in the counter-example (but note that here $P^1 \to \infty$). Notice also that the utility of group 1 in equilibrium is that from choosing D (as they are indifferent) so equal to 2. Hence while the strategies with increased monitoring difficulty converge to the shadow mixing strategies the utilities do not as in the shadow mixing equilibrium group 1 gets 5 (group 1 gets higher utility because it incurs no punishment costs).

10.5. Generalizations and Specializations

The basic principle exposed here is that with a suitable assumption about monitoring and punishment a game between groups choosing mechanisms can be reduced to a game in which the strategies are social norms and the utility function incorporates monitoring costs: $u^k(0, \varphi^k, f^{-k}) - M^k(\varphi^k, f^{-k})$. This is useful not only for proving the existence of equilibrium but in the analysis of equilibrium. For conceptual and notational clarity we proved this assuming that the underlying game is finite, that the monitoring technology of a group is independent of the actions of other groups, that the chance of detecting a deviation is independent of the deviation, that there is no choice of monitoring technology, and that the social cost of punishment is the punishment itself. The principle extends to more general settings: continuum of actions, more general monitoring technologies, and convex social cost of punishment. Indeed we have already shown that it holds for a variety of contests. The notation and definition of what it means to have adequate punishment is correspondingly more complicated, but not difficult to verify in specific cases. They key idea is in equilibrium group k knows the mixture over social norms of the other groups f^{-k} and views it as outside of its control. Conditional on this fixed f^{-k} the monitoring and punishment can be chosen optimally with respect to any particular choice of social norm φ^k by group k. Provided that every social norm φ^k can be made incentive compatible for some (possibly large) cost for each f^{-k} we can accordingly reduce the game.

In the opposite direction, we have allowed the gain to deviating $u^k(\phi^i, \varphi^k, \varphi^{-k}) - u^k(0, \varphi^k, \varphi^{-k})$ within group k to depend on the actions of other groups φ^{-k}. In the setting of all-pay contests (but not winner-pay contests) the gain to deviating does not in fact depend on the actions of other groups. If $u^k(\phi^i, \varphi^k, \varphi^{-k}) - u^k(0, \varphi^k, \varphi^{-k}) = G^k(\phi^i, \varphi^k)$ then it is easy to see that $g^k(\varphi^k, f^k)$ is independent of f^k and therefore so is the monitoring cost $M^k(\varphi^k, f^{-k}) = M^k(\varphi^k)$. This additional structure can simplify the analysis of equilibrium when it is relevant.

CHAPTER 11

Technical: All Pay Contests[1]

So far we have focused on auctions as a conflict resolution mechanism. Our theory of groups solving a mechanism design problem differs from the general literature on auctions only in that the cost of bids must take account of monitoring cost. Despite the tripartite auction theorem there are practical differences between auctions where bids are submitted and auctions where the parties "duke it out" whether in the voting booth or the streets. So far we have assumed that the highest bidder wins: the group that provides the most effort wins the prize. In an all-pay setting this is not always realistic - not in the case of voting and even less so in the case of political conflicts such as street demonstrations or civil war. When both sides put forth effort, quite often there is an element of uncertainty about who will win. In the case of voting for example, note that:

- There is uncertainty about the outcome because participation by individuals is not certain. If a finite number of individuals independently draw participation costs the total effort of each group is random: it is the sum of independent random decisions on whether or not to vote and so follows a binomial distribution as in Palfrey and Rosenthal (1985) or Levine and Palfrey (2007). Moreover, individual draws of participation costs may be correlated: for example, bad weather may raise participation costs for all members in regions where a party is heavily concentrated.
- The size of the two parties may be uncertain - nobody knows for sure how many people support their candidate or cause. This is the approach taken, for example, in Shachar and Nalebuff (1999) and Coate and Conlin (2004b).
- There can be random errors in the counting of votes, in the way that votes are validated, or courts may intervene in the vote counting - as happened, for example, in the 2000 U.S. Presidential election between Bush and Gore.

How robust are all-pay auction results to the addition of noise? We examine this question with a great deal of agnosticism about the cost of bidding as we know this depends in a complicated way on the distribution of individual participation costs and monitoring. Our conclusion is that auction results are quite robust: indeed, for homogeneous cost, a type of noise known as the Tullock function, and non-binding resource constraints we show that even for implausibly large levels of noise equilibrium utility in the contest is identical to that in the all-pay auction.

11.1. The Model

Two contestants $k \in \{1, -1\}$ compete for a prize worth $V_k > 0$ to contestant k. Each contestant chooses an effort level $b_k \geq 0$. The probability of contestant k

[1] This chapter is based on Levine and Mattozzi (2019).

winning the prize is given by a *contest success function* $0 \le p(b_k, b_{-k}) \le 1$ that is symmetric in the sense that it depends on the efforts of the two contestants and not on their names.

The contest success function is assumed to be continuous for $b_k \ne b_{-k}$, non-decreasing in b_k, and it must satisfy the adding-up condition $p(b_k, b_{-k}) + p(b_{-k}, b_k) = 1$. We allow for a discontinuous jump in the winning probability when we move away from $b_k = b_{-k}$, but require that when there is a tie the probability of winning is 1/2. Two standard contest resolution functions have this type of discontinuity: we are familiar with the all-pay auction in which the highest effort wins for sure so the probability of winning jumps from zero to one when there is a tie.

The Tullock Function. Another convenient function that satisfies these properties is that introduced by Tullock (1967) and analyzed in the case of voting by Herrera et al. (2015):

$$p(b_k, b_{-k}) = \frac{b_k^\beta}{b_k^\beta + b_{-k}^\beta}.$$

with $\beta > 0$. Here $1/\beta$ may be interpreted as "noise" or the level of exogenous uncertainty. As $1/\beta \to 0$ the Tullock function approaches the ordinary all-pay auction in which the highest bidder has probability 1 of winning. To see this suppose $b_k > b_{-k}$. Writing $p(b_k, b_{-k}) = 1/\left[1 + (b_{-k}/b_k)^\beta\right]$ we see that $p(b_k, b_{-k}) \to 1$ as $\beta \to \infty$. By contrast $p(b_k, b_{-k}) \to 1/2$ as $\beta \to 0$. Notice that $p(b_k, b_{-k})$is discontinuous when there is a tie at zero.

The raw cost of effort b_k is $V_k \varrho_k(b_k)$ and it is incurred regardless of the outcome of the contest. The function $\varrho_k(\cdot)$, which we refer to simply as the cost of effort, measures cost relative to the value of the prize $V_k > 0$. We assume that $\varrho_k(\cdot)$ is continuous, non-decreasing, it satisfies $\varrho_k(0) = 0$, and for some W_k called the *willingness to bid* $\varrho_k(W_k) = 1$ and if $b_k > W_k$ then $\varrho_k(b_k) > 1$. To avoid degeneracy we assume that for contestant -1 the cost function $\varrho_{-1}(\cdot)$ is strictly increasing at the origin.

The raw objective function of contestant k is given by $V_k p(b_k, b_{-k}) - V_k \varrho(b_k)$, or equivalently $p(b_k, b_{-k}) - \varrho(b_k)$ which we refer to as the utility function. Since choosing effort higher than the willingness to bid is strictly dominated by choosing zero effort, we may restrict the choice of effort to $[0, \overline{W}]$, where $\overline{W} > \max\{W_k, W_{-k}\}$. Hence, a strategy for contestant k is a cdf G_k on $[0, \overline{W}]$. Define $p(G_k, G_{-k}) \equiv \int_0^{\overline{W}} \int_0^{\overline{W}} p(b_k, b_{-k}) dG_k(b_k) dG_{-k}(b_{-k})$ and $\varrho_k(G_k) \equiv \int_0^{\overline{W}} \varrho_k(b_k) dG_k(b_k)$. A *Nash equilibrium* is a pair of strategies (G_k, G_{-k}) such that for each contestant k and all strategies $\tilde{G}_k$ we have

$$p(G_k, G_{-k}) - \varrho_k(G_k) \ge p(\tilde{G}_k, G_{-k}) - \varrho_k(\tilde{G}_k).$$

Since this is an expected utility model this definition is equivalent to restricting deviations to pure strategies b_k.

11.2. War, Peace and Rents

There are a variety of properties of contests that are important in political economy: here we focus on success, rent dissipation, effort provision, and the use of pure strategies.

A basic question in any contest is: who wins? As we know from the all-pay auction the answer is uncertain, so we may well ask: who is most likely to win.

Formally we say that k has *outcome success* if $p(G_k, G_{-k}) > 1/2$ or equivalently $p(G_k, G_{-k}) > p(G_{-k}, G_k)$. This, however, fails to take into account the cost of the resources used in achieving success, so we say that k has *greater success* if $p(G_k, G_{-k}) - \varrho_k(G_k) > p(G_{-k}, G_k) - \varrho_{-k}(G_{-k})$, that is, k gets a greater fraction of achievable utility than $-k$. Since the contest success function is symmetric it is natural to think that the contestant with the lower cost will have greater success. This is our first question.

A second question is: how well do the contestant do? In equilibrium each contestant gets a rent $p(G_k, G_{-k}) - \varrho_k(G_k)$. This is non-negative since it is possible to bid zero and incur no cost. The combined rents of the two contestants are $1 - \varrho_k(G_k) - \varrho_{-k}(G_{-k})$. How much of the value of the prize, normalized here to one, is wasted or dissipated in the process of competing for the prize? If the combined rents are zero we refer to *complete rent dissipation*. Recall that in our study of the all-pay auction the disadvantaged contestant gets no rent. If the contest is symmetric then neither contestant gets any rent and we have complete rent dissipation. Is this typical of contests where contestants have identical costs?

Related to rent dissipation is the question: how much effort do contestants provide? For example, if neither provides any effort there would be no rent dissipation. We refer to this as a *peaceful equilibrium*. Hirshleifer (1989) points out that it is likely to be the case in practice that effort makes the greatest difference when the contest is close. He argues that in this case one contestant should be expected to provide zero effort. He suggests this means that there should be positive rents. Is this correct?

Recall that in the all-pay auction the disadvantaged contestant gets no rent. Is this generally the case? It is in fact possible because bidding zero when the opponent bids positive gives no chance of success. Not all contests have this property. A simple contest that does not is the *generalized all pay* auction where $b_k > b_{-k}$ results in $p(b_k, b_{-k}) = 1 - 2q$ where $0 \leq q < 1/2$. Here higher effort guarantees a greater chance of winning, but the loser also has a chance of winning: for example this could model an electoral process where there is a chance of corruption overturning the electoral result. In this case it is clear that no contestant can get less than q since this is guaranteed by bidding zero. It is clear as well that in such a case there cannot be complete rent dissipation regardless of symmetry.

Finally: what do equilibrium strategies look like? Do players play pure strategies or do they have to mix as they do in the all-pay auction? If they mix, what do their mixed strategies look like? Are they given by continuous densities? Do they have finite support? Or are they given by some sort of unfathomable cdf like the Cantor function, which is continuous and manages to creep upwards from zero to one while having a derivative almost everywhere equal to zero.

11.3. Existence of Pure and Mixed Equilibrium

Theorems about things that do not exist are meaningless. Hence, as in Chapter 10, a key first step is to prove that an equilibrium exists. In game theory this is usually done by waving hands and saying "Kakutani (1941)." Here, however the strategy space is infinite dimensional. If the objective function were continuous we could wave our hands and say "Glicksberg (1952)." However, because we wish to include contests such as the all-pay auction which is discontinuous when there is a tie we cannot do that either. Nor yet can we wave our hands and say "Dasgupta and Maskin (1986)" because their result on existence of equilibrium in discontinuous games does not apply to this model. With a little work we can say "Prokopovych and Yannelis (2014)," and in addition existence follows from a deeper result on upper hemi-continuity of the equilibrium correspondence which we discuss below. The crucial fact is that an equilibrium, possibly in mixed strategies, does exist. It is also important to know whether the equilibrium is in pure strategies and the next result gives some basic information about equilibria.

THEOREM 11.3.1. *Nash equilibria exist. In every Nash equilibrium the probability of a tie at a point of discontinuity is zero. If both contestants have the same costs there is a symmetric Nash equilibrium. However, if $p(b, b)$ is a point of discontinuity for all $0 \le b \le \overline{W}$ the symmetric equilibrium is not in pure strategies.*

This partially generalizes the result from all-pay auctions that mixed strategies must be used: it shows that if the probability of winning jumps when there is a tie then when the two contestants have the same costs mixed strategies must be used.

PROOF. Existence is shown in Corollary 11.8.4 below.

We show that if $p(b, b)$ is a point of discontinuity then both contestants cannot have an atom at b so the probability of (b, b) is zero. Notice that this immediately implies that if $p(b, b)$ is a point of discontinuity for all $0 \le b \le \overline{W}$ a symmetric equilibrium cannot be in pure strategies.

To show that both contestants cannot have an atom at b we show that if G_{-k} has an atom at $b < \overline{W}$ and $p(b, b)$ is a point of discontinuity then $b_k = b$ is not a best response by k to G_k. If $b = \overline{W}$ this is obvious since that effort level is strictly dominated by 0.

Define $p^+(b) = \lim_{\epsilon \to 0^+} p(b + \epsilon, b)$. First we show that if $p(b, b)$ is a point of discontinuity of b then $p^+(b) > 1/2$. Discontinuity implies that there is a sequence $b^n \to (b, b)$ with $\lim p(b^n) \neq p(b)$. From symmetry we may assume $\lim p(b^n) > 1/2$. Fix $b + \epsilon$ where $\epsilon > 0$. For n sufficiently large $b_1^n < b + \epsilon$. Hence $p(b + \epsilon, b_{-1}^n) \ge p(b_1^n, b_{-1}^n)$. Since $p(b + \epsilon, b)$ is a point of continuity of $p(b_1, b_{-1})$ we have $p(b + \epsilon, b) = \lim p(b + \epsilon, b_{-1}^n) \ge \lim p(b_1^n, b_{-1}^n)$. Hence $p^+(b) = \lim_{\epsilon \to 0^+} p(b + \epsilon, b) \ge \lim p(b_1^n, b_{-1}^n) > 1/2$.

The remainder of the proof is to show that when $p^+(b) > 1/2$ it would be better to choose a little bit more effort than b so as to break the tie and get a jump in the probability of winning at trivial additional cost. Specifically, suppose that $-k$ has an atom $g_{-k}(b)$ at b. If k provides effort $b + \epsilon$ instead of b then k gains at least

$$g_{-k}(b)(p^+(b) - 1/2) + \varrho(b) - \varrho(b + \epsilon).$$

In the limit as $\epsilon \to 0$ this is strictly positive proving the result.[2] □

[2] It may be a bit puzzling when $b = W_k$ to think of contestant k deviating to $b + \epsilon$. Clearly this cannot be optimal. However, the argument shows that although such a deviation to a strictly dominated strategy is suboptimal it does better than b, which is just another way of saying b was not a terribly good idea in the first place.

11.4. The Zero Sum Property and Consequences[3]

When a contest is resolved by physical conflict such as warfare or protests the effort provided by a contestant generally creates a negative externality for the opponent: battle damage. Specifically, in such a contest the objective function would be given by $p(b_k, b_{-k}) - \varrho_k(b_k) - d_k(b_{-k})$ where $d_k(b_{-k})$ represents harm inflicted on k by the effort of the opponent. It could also represent costs of $-k$ that are incurred by k. Another example is Feddersen and Sandroni (2006) who assume for ethical reasons that contestant cares equally about the cost of the other contestant as about their own. Subsequent models of ethical voters (also known as rule utilitarian) such as Coate and Conlin (2004b) have dropped that assumption, however.

The key point is that subtracting a function that depends only on the other contestant's choice modifies the payoffs of the game but the modified game is *best response equivalent* in the sense that the best response functions for both contestants are exactly the same. In particular the modified game has exactly the same Nash equilibria. If we agree that it is still sensible to measure achievement of goals by $p(b_k, b_{-k}) - \varrho_k(b_k)$, that is net of battle damage about which k can do nothing, then $d(b_{-k})$ does not matter. It does matter, however, for assessment of the efficiency of different contest success functions: low collateral damage is obviously socially desirable. For example, in choosing between an election and a military conflict both designed to exactly mimic the contest success, the former is preferred because it avoids collateral damage.

One particular case of importance is to take $d(b_{-k}) = (1/2) - \varrho_{-k}(b_{-k})$. In this case the objective function is given by $u_k^0(b_k, b_{-k}) = p(b_k, b_{-k}) - \varrho_k(b_k) + \varrho_{-k}(b_{-k}) - 1/2$ and adding the objective functions of the two contestants we see that this is a zero sum game. The key point is that Nash equilibria of zero sum games have strong properties.

THEOREM 11.4.1. *$\hat{G}_k \hat{G}_{-k}$ is a Nash equilibrium if and only if $u_k^0(\hat{G}_k, \hat{G}_{-k}) = \sup_{G_k} \inf_{G_{-k}} u_k^0(G_k, G_{-k}) = \inf_{G_{-k}} \sup_{G_k} u_k^0(G_k, G_{-k})$.*

PROOF. We cannot use standard versions of the minimax theorem because the objective function is not upper semi-continuous. However, we can use the fact from Theorem 11.3.1 to prove it directly. Specifically, by the definition of Nash equilibrium $u_k^0(\hat{G}_k, \hat{G}_{-k}) = \sup_{G_k} u_k^0(G_k, \hat{G}_{-k})$ Hence

$$u_k^0(\hat{G}_k, \hat{G}_{-k}) \geq \sup_{G_k} \inf_{G_{-k}} u_k^0(G_k, G_{-k}), \inf_{G_{-k}} \sup_{G_k} u_k^0(G_k, G_{-k}).$$

Applying this to the opponent since $u_{-k}^0(\hat{G}_{-k}, \hat{G}_k) = -u_k^0(\hat{G}_k, \hat{G}_{-k})$

$$u_k^0(\hat{G}_k, \hat{G}_{-k}) \leq \sup_{G_k} \inf_{G_{-k}} u_k^0(G_k, G_{-k}), \inf_{G_{-k}} \sup_{G_k} u_k^0(G_k, G_{-k})$$

which is possible only if the minimax result stated in the theorem is true. Note that this result does not depend upon the previous theorem. □

The zero-sum property has an important consequence for equilibrium first shown in Ewerhart and Valkanova (2020):

[3]We are grateful to Christian Ewerhart for pointing out to us the equivalence with zero-sum games. This result can be found in Ewerhart and Valkanova (2020) and the mix and match property of minimax equilibria can be found in his earlier work in Ewerhart (2017) and Ewerhart and Sun (2018).

THEOREM 11.4.2. *A strict Nash equilibrium is the unique Nash equilibrium.*

PROOF. A strict Nash equilibrium is necessarily in pure strategies. Let b_k, b_{-k} be a strict Nash equilibrium. Suppose without loss of generality that there is a second Nash equilibrium G_k, G_{-k} in which $G_{-k} \neq b_{-k}$. By Theorem 11.4.1 $u_k^0(b_k, b_{-k}) = u_k^0(G_k, G_{-k})$. Optimality for $-k$ at b_k, b_{-k} implies $u_k^0(b_k, b_{-k}) \leq u_k^0(b_k, G_{-k})$, while optimality for k at G_k, G_{-k} implies $u_k^0(b_k, G_{-k}) \leq u_k^0(G_k, G_{-k})$. Hence $u_k^0(b_k, b_{-k}) = u_k^0(b_k, G_{-k})$ contradicting the assumption that $-k$ is playing a strict best response at b_k, b_{-k}. $\square$

11.5. Peaceful Equilibria

In the all-pay auction if one contestant provides no effort the other will not since a small effort increases the chances of winning from $1/2$ to 1 while continuity of the cost function implies this leads to only a minuscule increase in cost. In other words, in the all-pay auction some effort must be provided in equilibrium and there must be some rent dissipation. With general contest success functions this need not be the case. Consider the following example:

EXAMPLE 11.5.1. Suppose that suppose that $\varrho_1(b_1) = b_1, \varrho_{-1}(b_{-1}) = 2b_{-1}$ but that $p(b_k, b_{-k}) \equiv 1/2$ so that effort does not matter. Then the unique equilibrium is for each to provide zero effort so both get $1/2$. Notice that despite fact that 1 has lower cost than -1 this does not lead to greater success or outcome success.

We define *peaceful* equilibria as those in which both contestants choose to incur zero cost of effort and have a probability of winning of $1/2$ and, recalling our utility normalization, utility equal to $1/2$. In particular there is no rent dissipation and neither contestant has greater success or greater outcome success regardless of any cost advantages. To have a *contested* equilibrium[5] in which this is not the case we must rule out situations such as Example 11.5.1 in which the cost function rises too fast relative to the steepness of the contest success function. We begin with the relevant definitions.

We start with the possibility that contestant k finds it a strict best response to reply to zero effort with zero effort, that is, $p(0,0) - \varrho_k(0) > p(b_k, 0) - \varrho_k(b_k)$ for all $b_k > 0$. This we can rewrite as $\varrho_k(b_k) > p(b_k, 0) - p(0,0)$. This separates the cost from the contest success function, and the right hand side is the same for both contestants. If $p(b_k, 0) - p(0,0)$ is continuous at 0 all sufficiently large $\varrho_k(b_k)$ will satisfy this condition so we call it *high cost*. Notice that when p is discontinuous at $(0,0)$ as it is in the all-pay auction or the Tullock case, high cost is ruled out because $\varrho_k(b_k)$ is continuous and $\varrho_k(0) = 0$.

By contrast, we say that contestant k has *low cost* if zero effort is not a best response to zero effort, that is, for some b_k we have $\varrho_k(b_k) < p(b_k, 0) - p(0,0)$, and in particular high cost and low cost are mutually exclusive.

THEOREM 11.5.2. *If both have high cost then the unique equilibrium is peaceful and neither provides effort. If* 1 *has low cost all equilibria are contested.*

PROOF. If both have high cost then each finds it strictly optimal to provide zero effort when the other is doing so. Since this is a strict Nash equilibrium by Theorem 11.4.2 it is the unique equilibrium.

[5]Note that if contestant 1 has a headstart, that is, a flat cost function at 0 then there can be a contested equilibrium in which 1 provides effort but the total cost of effort by both contestants is still zero.

Finally, at a peaceful equilibrium since $\varrho_{-1}(b_{-1})$ is assumed to be strictly increasing at the origin, as we noted above, it must be that -1 provides zero effort. The condition for 1 having low cost may be written as $p(b_1, 0) - \varrho_1(b_1) > p(0,0) - \varrho_1(0)$ implying that 1 gets strictly more than $1/2$ in equilibrium. This requires that the chance of 1 winning is greater than $1/2$ contradicting the definition of a peaceful equilibrium. $\square$

Note that while this result establishes necessary conditions for a contested equilibrium and also sufficient conditions, there is a gap between the two conditions

11.6. Cost and Success

Does lower cost lead to greater success? Intuition suggests it should, but intuition about equilibria is often wrong. In examining cost it will be convenient to assume that it is contestant 1 who has the lower cost. We start with the simplest notion of lower cost which we call *a pure cost advantage*: if $b > 0$ then $\varrho_1(b) < \varrho_{-1}(b)$. We consider first as a prototypical example the generalized all pay auction where recall that $b_k > b_{-k}$ results in $p(b_k, b_{-k}) = 1 - 2q$ where $0 \leq q < 1/2$. Here higher effort guarantees a greater chance of winning, but the loser also has a chance of winning. The following result adapts a well known result for the standard all pay auction for which $q = 0$.[6]

THEOREM 11.6.1. *In the generalized all-pay auction if* 1 *has a pure cost advantage then in any equilibrium* 1 *has greater success.*

PROOF. Let (G_1, G_{-1}) be an equilibrium of the game. Define $\overline{b}_{-1} \equiv \max \operatorname{supp} G_{-1}$. Consider the strategy for 1 of providing effort $b_\epsilon \equiv \overline{b}_{-1} + \epsilon < \overline{W}$. In the all-pay auction this guarantees a win, so

$$p(G_1, G_{-1}) - \varrho_1(G_1) \geq 1 - 2q - \varrho_1(b_\epsilon).$$

By the continuity of ϱ_1 this implies

$$p(G_1, G_{-1}) - \varrho_1(G_1) \geq 1 - 2q - \varrho_1(\overline{b}_{-1}).$$

Because 1 has a pure cost advantage, the right hand side of the inequality is strictly larger than $1 - 2q - \varrho_{-1}(\overline{b}_{-1})$.

Because $\overline{b}_{-1} \in \operatorname{supp} G_{-1}$ there is a sequence $b^n \to \overline{b}_{-1}$ with

$$p(b^n, G_1) - \varrho_{-1}(b^n) = p(G_{-1}, G_1) - \varrho_{-1}(G_{-1}).$$

By the continuity of ϱ_{-1} this implies

$$q - \varrho_{-1}(\overline{b}_{-1}) \geq p(G_{-1}, G_1) - \varrho_{-1}(G_{-1}).$$

Hence it is indeed the case that 1 has greater success. $\square$

Our goal is to understand how this result extends to more general contest success functions. We know this cannot be the case with peaceful equilibria, so we focus on contested equilibria. We first show that even in this case pure cost advantage is not in general sufficient for the cost advantaged contestant to have greater success.

---

[6]See, for example, Siegel (2014).

EXAMPLE 11.6.2. Here we construct a contested pure strategy equilibrium in which 1 has a pure cost advantage but -1 has greater success. Take $p(b_k, b_{-k}) = (1/2)+(1/2)(b_k - b_{-k})$ truncated by 0 below and 1 above. The cost function for 1 is $\varrho_1(b_1) = (4/7)(b_1 - 1)$ for $b_1 \geq 1$ and 0 otherwise. For -1 it is $\varrho_{-1}(b_{-1}) = (3/7)b_{-1}$ for $0 \leq b_{-1} \leq 2$ and $6/7+(4/7)(b_{-1}-2)$ otherwise. At $b = 0$ we have $\varrho_1 = \varrho_{-1} = 0$. At $b = 1$ we have $\varrho_1 = 0$, and $\varrho_{-1} = 3/7$. At $b = 2$ we have $\varrho_1 = 4/7$, and $\varrho_{-1} = 6/7$. Above 2 the cost difference remain equal to $2/7$ in favor of -1. So 1 has a pure cost advantage. We claim that $(b_1, b_{-1}) = (1, 2)$ is a pure strategy equilibrium. Here 1 loses for certain and has no cost so gets 0 while -1 wins for sure and has a cost of $6/7$ so gets $1/7$. Hence certainly -1 is more successful. To see this is an equilibrium observe that 1 is indifferent to reducing effort below 1: there is no cost and no chance of winning there. Increasing effort above 1 increases the chances of winning at the rate of $1/2$ while it increases costs at the rate of $4/7$ so in fact $b_1 = 1$ is optimal for contestant 1. For -1 reducing effort below 2 reduces the chances of winning at the rate of $1/2$ but decreases costs only at the rate of $3/7$. Increasing effort above 2 has no effect on the chances of winning but simply increases costs. Hence $b_{-1} = 2$ is optimal for contestant -1.

This leads us to consider two strengthened notions of cost advantage

(1) 1 has a *marginal cost advantage* if for $b_2 > b_1$we have $\varrho_1(b_2) - \varrho_1(b_1) < \varrho_{-1}(b_2) - \varrho_{-1}(b_1)$

(2) 1 has a *homogeneous cost advantage* if $\varrho_1(b) = \nu\varrho_{-1}(b)$ for some $0 < \nu < 1$

Given these notions, we have that homogeneous cost advantage implies marginal cost advantage, and marginal cost advantage implies pure cost advantage. An important special case of homogeneous cost advantage occurs when both contestants have the same absolute cost: for all b we have $V_1\varrho_1(b) = V_{-1}\varrho_{-1}(b)$. In this case 1 has a homogeneous cost advantage if and only if the prize is valued more highly: $V_1 > V_{-1}$.

The notions of pure, marginal, and homogeneous cost advantage are defined independent of the contest success function. An alternative approach is to relate the size of the cost advantage to measures of the steepness of the contest success function. A simple but quite strong form of cost advantage is the following: we say that 1 has a *strong cost advantage over* -1 if for some $b_1 > W_{-1}$, where W_{-1} is the willingness to bid defined earlier, we have $p(W_{-1}, W_{-1}) = \frac{1}{2} < p(b_1, W_{-1}) - \varrho(b_1)$. This condition implies that, no matter how small player 1's cost at W_{-1} (that is, even if it is zero) there is an $b_1 > W_{-1}$ that yields higher payoff when played against W_{-1} than playing W_{-1} does. To understand this condition better fix W_{-1}, -1's willingness to bid. If contest success has a strict increase above this point, a sufficiently low cost for 1 will always lead to a strong cost advantage. On the other hand, strong cost advantage in the all pay auction requires that $\varrho_1(W_{-1}) < 1/2$, while greater success requires only that $\varrho_1(W_{-1}) < 1$.

For this reason we introduce a weaker condition applied over a broader range of effort levels. We say that 1 has a *uniform cost advantage* over -1 if for any $0 \leq b_{-1} \leq W_{-1}$ there is an $b_1 > b_{-1}$ with $\varrho_1(b_1) < \varrho_{-1}(b_{-1}) - (p(b_{-1}, 0) - p(b_1, b_{-1}))$, that is player 1 earns strictly more playing b_1 against b_{-1} than player -1 earns playing b_1 against 0. Notice that this condition is satisfied in the all-pay auction provided that 1 has a cost advantage. It is also satisfied in a difference model in which $p(b_1, b_{-1}) = p(b_1 - b_{-1}, 0)$ if $\varrho_1(2b_1) < \varrho_{-1}(b_1)$. One particularly important case of a uniform cost advantage arises when there is a common underlying strictly increasing cost function $\varrho_2(b)$ but contestant 1 has a sufficient effort advantage of $\bar{b}_1 > 0$, meaning that the probability that 1 wins with underlying effort $\tilde{b}_1$ is given by $p(\tilde{b}_1 + \bar{b}_1, b_{-1})$. This is known in the literature as a "head start advantage." This can be made equivalent to the original model by defining $\varrho_1(b_1) = \tilde{\varrho}_2(b_1 - \bar{b}_1)$ for $b_1 \geq \bar{b}_1$ and 0 otherwise. Notice that in this case the cost advantage cannot be homogeneous.

Global Noise. There are two distinct types of noise: there can be local noise, meaning that a close contest may have substantial uncertainty. Our results below on upper hemi-continuity of the equilibrium correspondence show that this type of noise has little effect on contest success. There can also be global noise, meaning that there is a substantial chance of winning with minimal effort - for example, if the overwhelmingly superior fleet of Genghis Khan were to be sunk by a divine wind on its way to Japan, or if Republican election officials were to declare that all votes cast by Democrat voters are invalid.

Global noise can easily be added to any contest success function by following the method we used to generalize the all-pay auction. We assume that with probability $2q$ the contest is decided by a coin flip and with probability $1 - 2q$ the contest is decided by the original function $p(b_k, b_{-k})$. This results in a new contest success function $\tilde{p}(b_k, b_{-k}) = q + (1 - 2q)p(b_k, b_{-k})$ and as indicated the generalized all-pay auction described above is an example of such $q-$*global noise*.

It is quite easy to work out the implications of a q-global noise. The modified objective function is $q + (1 - 2q)p(b_k, b_{-k}) - \varrho_k(b_k)$ which is equivalent to $p(b_k, b_{-k}) - (1/(1-2q))\varrho_k(b_k)$ so that strategically the q-global noise is equivalent to increasing both costs by the common factor $1/(1 - 2q)$, that is, using the cost $\tilde{\varrho}_k(b_k) = (1/(1 - 2q))\varrho_k(b_k)$. Naturally this reduces the incentive to compete. For example, if cost is strictly increasing and $p(b_k, b_{-k})$ is continuous then as q approaches $1/2$ the high cost condition $\tilde{\varrho}_k(b_k) > p(b_k, b_{-k}) - p(0, b_{-k})$ for all $b_k > 0$ will be satisfied and the equilibrium will be peaceful.
More broadly, the conditions of cost advantage, marginal cost advantage, and uniform cost advantage are invariant to the addition of q-global noise. By contrast, the notions of strong and uniform cost advantage, depending as they do on the magnitude of cost are not. Similarly willingness to bid is decreased by a q-modification, and indeed the willingness to bid of the two parties can be reversed by q-global noise.

Finally, once we have retrieved the utilities $\hat{u}_k$ from an equilibrium in the strategically equivalent model, we may map them back to original units as $q+(1-2q)\hat{u}_k$ so that in particular each contestant gets at least q.

Finally, we introduce the concept of *preemptive equilibrium* and say that G_1, G_{-1} is a preemptive equilibrium if either one distribution first order stochastically dominates the other or the two are equal. Any pure strategy equilibrium is preemptive. Equipped with these new definitions we can map the relationship between cost advantage and success:

THEOREM 11.6.3. *In a contested equilibrium 1 has greater success if any of the following conditions are satisfied*

(0) 1 *has a pure cost advantage and* -1 *does not have outcome success,*

(i) 1 *has a marginal cost advantage and the equilibrium is preemptive,*

(ii) 1 *has a homogeneous cost advantage,*
(iii) 1 *has a strong cost advantage,*
(iv) 1 *has a uniform cost advantage.*

Notice that in Example 11.6.2 while 1 had a pure cost advantage in the range $[1,2]$, 1 also had higher marginal cost than -1. This possibility is ruled out by marginal cost advantage. With this assumption 1 has greater success in all preemptive equilibria. For pure strategies this trivially "works" since all pure strategy equilibria are preemptive. Unfortunately pure strategy equilibria do not always exist and we do not have general results about when equilibria are preemptive. If we further strengthen the cost advantage assumption to homogeneous cost advantage then we get a general result for all equilibria pure or mixed.

PROOF. Suppose that (G_1, G_{-1}) is an equilibrium. From optimality of G_k and symmetry we have

$$p(G_k, G_{-k}) - \varrho_k(G_k) \geq p(G_{-k}, G_{-k}) - \varrho_k(G_{-k}) = 1/2 - \varrho_k(G_{-k}). \tag{11.6.1}$$

By rearranging the terms we also have

$$p(G_k, G_{-k}) - 1/2 \geq \varrho_k(G_k) - \varrho_k(G_{-k}). \tag{11.6.2}$$

First, we show (0). Suppose that 1 has a pure cost advantage but does not have greater success. Then

$$p(G_{-1}, G_1) - \varrho_{-1}(G_{-1}) \geq p(G_1, G_{-1}) - \varrho_1(G_1) \geq 1/2 - \varrho_1(G_{-1}). \tag{11.6.3}$$

Where the first inequality follows from the fact that 1 does not have greater success, and the second from equation 11.6.1. Suppose first-1 is not providing effort. Then $\varrho_{-1}(G_{-1}) = \varrho_1(G_{-1}) = 0$ so 11.6.3 implies $p(G_{-1}, G_1) \geq 1/2$. Moreover p nondecreasing implies $p(G_{-1}, G_1) = p(0, G_1) \leq 1/2$ so $p(G_{-1}, G_1) = 1/2$. Since this is not a peaceful equilibrium it must be that $\varrho_1(G_1) > 0$ so $p(G_1, G_{-1}) - \varrho_1(G_1) = 1/2 - \varrho_1(G_1) < 1/2$ while choosing $b_1 = 0$ gives a utility of $1/2$ contradicting the fact that 1 is playing optimally. Suppose second that -1 is providing effort. By the pure cost advantage equation

$$1/2 - \varrho_1(G_{-1}) > 1/2 - \varrho_{-1}(G_{-1})$$

From equation 11.6.3 this gives $p(G_{-1}, G_1) > 1/2$. Consequently -1 has outcome success. This proves (0).

To show (i), notice that from equation 11.6.2 with $k = 1$ we have

$$p(G_1, G_{-1}) - 1/2 \geq \varrho_1(G_1) - \varrho_1(G_{-1}).$$

From symmetry this gives

$$-p(G_{-1}, G_1) + 1/2 \geq \varrho_1(G_1) - \varrho_1(G_{-1})$$

or

$$p(G_{-1}, G_1) - 1/2 \leq \varrho_1(G_{-1}) - \varrho_1(G_1). \tag{11.6.4}$$

From equation 11.6.2 with $k = -1$ we have

$$p(G_{-1}, G_1) - 1/2 \geq \varrho_{-1}(G_{-1}) - \varrho_{-1}(G_1)$$

Hence

$$\varrho_1(G_{-1}) - \varrho_1(G_1) \geq \varrho_{-1}(G_{-1}) - \varrho_{-1}(G_1). \tag{11.6.5}$$

Suppose that 1 has a marginal cost advantage. If G_1 first order stochastically dominates G_{-1} or the two are equal then -1 does not have a outcome advantage so 1 has greater success by (0). Suppose instead that G_{-1} first order stochastically dominates G_1. For $b_2 > b_1$ the condition for marginal cost advantage can be written as $\varrho_{-1}(b_2) - \varrho_1(b_2) > \varrho_{-1}(b_1) - \varrho_1(b_1)$. It follows that $\varrho_{-1}(G_{-1}) - \varrho_1(G_{-1}) > \varrho_{-1}(G_1) - \varrho_1(G_1)$. This contradicts equation 11.6.5. This shows (i).

Next, we show (ii). Suppose that 1 has a homogeneous cost advantage. From equation 11.6.5

$$\varrho_1(G_{-1}) - \varrho_1(G_1) \geq \varrho_{-1}(G_{-1}) - \varrho_{-1}(G_1) = (1/\nu)\left(\varrho_1(G_{-1}) - \varrho_1(G_1)\right).$$

Since $\nu < 1$ it follows that $\varrho_1(G_{-1}) - \varrho_1(G_1) \leq 0$. From equation 11.6.4

$$p(G_{-1}, G_1) - 1/2 \leq \varrho_1(G_{-1}) - \varrho_1(G_1) \leq 0$$

so -1 does not have an outcome success. There are two possibilities. First, if 1 does not have an outcome success either, then, it must be that $p(G_{-1}, G_1) = 1/2$ so that also $p(G_1, G_{-1}) = 1/2$. By (0) we may assume that -1 does not provide zero effort with probability one so by cost advantage

$$p(G_1, G_{-1}) - \varrho_1(G_{-1}) > p(G_1, G_{-1}) - \varrho_{-1}(G_{-1}) = p(G_{-1}, G_1) - \varrho_{-1}(G_{-1})$$

and indeed 1 instead has greater success. The second possibility is that 1 does have outcome success. In this case by (0) 1 also has greater success. This proves (ii).

We now show (iii). If 1 has a strong cost advantage then there is a $\hat{b}_1$ with $\varrho_1(\hat{b}_1) < p(\hat{b}_1, W_{-1}) - p(W_{-1}, W_{-1}) = p(\hat{b}_1, W_{-1}) - 1/2$. Hence $p(\hat{b}_1, W_{-1}) - \varrho_1(\hat{b}_1) > 1/2$. Observe that $G_{-1} \leq W_{-1}$ so $p(\hat{b}_1, W_{-1}) \leq p(\hat{b}_1, G_{-1})$. Finally, from optimality

$$p(G_1, G_{-1}) - \varrho_1(G_1) \geq p(\hat{b}_1, G_{-1}) - \varrho_1(\hat{b}_1) \geq p(\hat{b}_1, W_{-1}) - \varrho_1(\hat{b}_1) > 1/2$$

which as both contestants cannot have a utility greater than 1/2 implies greater success. This proves (iii)

Finally we prove (iv). Let $\hat{b}_{-1}$ be the top of the support of the equilibrium G_{-1}. Let $b_{-1}^n \leq \hat{b}_{-1}$ with $b_{-1}^n \to \hat{b}_{-1}$ and $p(b_{-1}^n, G_1) - \varrho_{-1}(b_{-1}^n) = p(G_{-1}, G_1) - \varrho_{-1}(G_{-1})$. Since at points of discontinuity of p the jump is up this implies

$$p(G_{-1}, G_1) - \varrho_{-1}(G_{-1}) \leq p(\hat{b}_{-1}, 0) - \varrho_{-1}(\hat{b}_{-1}).$$

From the definition of a uniform cost advantage there is a $\hat{b}_1$ such that

$$p(G_{-1}, G_1) - \varrho_{-1}(G_{-1}) < p(\hat{b}_1, \hat{b}_{-1}) - \varrho_1(\hat{b}_1).$$

Moreover because $\hat{b}_{-1}$ is the top of the support of G_{-1} we get

$$p(G_{-1}, G_1) - \varrho_{-1}(G_{-1}) < p(\hat{b}_1, G_{-1}) - \varrho_1(\hat{b}_1)$$

By optimality of G_1 this gives

$$p(G_{-1}, G_1) - \varrho_{-1}(G_{-1}) < p(G_1, G_{-1}) - \varrho_1(G_1)$$

that is to say, greater success. □

The following special case of parts (i) and (ii) of Theorem 11.6.3 is useful in a variety of applications.

Corollary 11.6.4. *In a contested equilibrium* 1 *has greater success if either of the following two conditions is satisfied:*

(i) Cost is linear for both contestants and 1 *has a pure cost advantage.*[7]

[7]This assumption is very popular in the literature.

(ii) 1 has a marginal cost advantage and one contestant provides no effort.

We also have an important result about symmetric contests.

COROLLARY 11.6.5. *In a symmetric contest, that is, $\varrho_1(\cdot) = \varrho_{-1}(\cdot)$ in any Nash equilibrium each has an equal chance of winning, and all Nash equilibria give both the same cost and normalized utility.*

PROOF. Suppose $\varrho_1(G_1) > \varrho_1(G_{-1})$. From Theorem 11.4.1 contestant 1's deviating to another equilibrium strategy cannot change -1's zero sum payoff

$$p(G_{-1}, G_1) - \varrho_{-1}(G_{-1}) + \varrho_1(G_1) - 1/2 = p(G_{-1}, G_{-1}) - \varrho_1(G_{-1}) + \varrho_{-1}(G_{-1}) - 1/2$$

implying that

$$p(G_{-1}, G_1) - \varrho_{-1}(G_{-1}) < p(G_{-1}, G_{-1}) - \varrho_1(G_{-1}) = p(G_1, G_{-1}) - \varrho_1(G_1)$$

contradicting the fact from Theorem 11.6.3 (ii) in any equilibrium both contestants must get the same payoff as each other.

Suppose then that G_1, G_{-1} is an equilibrium in which $p(G_1, G_{-1}) > 1/2$. By symmetry G_{-1}, G_1 is also an equilibrium, so by the mix and match property of minimax strategies so is G_{-1}, G_{-1}. From Theorem 11.4.1

$$p(G_1, G_{-1}) - \varrho_1(G_1) + \varrho_{-1}(G_{-1}) - 1/2 = p(G_{-1}, G_{-1}) - \varrho_1(G_{-1}) + \varrho_{-1}(G_{-1}) - 1/2.$$

This implies that $\varrho_1(G_1) > \varrho_1(G_{-1})$ which we just showed is impossible.

If two different equilibria $G_1, \tilde{G}_{-1}$ and $G_{-1}, \tilde{G}_1$ has $\varrho_1(G_1) > \varrho_1(G_{-1})$ this is also impossible since G_1, G_{-1} is also an equilibrium. $\square$

11.7. Convexity and An All-Pay Auction Result

As we earlier noted, Hirshleifer (1989) points out that it is likely to be the case in practice that effort makes the greatest difference when the contest is close. If this is the case, we would expect that the contest success function $p(b_k, b_{-k})$ should be convex in b_k for $b_k < b_{-k}$. He argues that in this case one contestant should be expected to provide zero effort. Alternatively it could lead to mixed strategy equilibria. We now examine the situation more closely: it leads to a remarkable result of Alcalde and Dahm (2007) that shows that many contests with substantial noise have equilibria payoff equivalent to the all-pay auction.

Consider a real valued function $h(b)$ on $[0, \infty)$. If the function is continuously differentiable then strict convexity implies that if $h(b) \geq h(0)$ for $b > 0$ then $h'(b) > 0$. For want of a better name generalize this idea by calling $h(b)$ *generalized convex up to* $\bar{b}$ if for $b \in (0, \bar{b}]$ and $h(b) \geq h(0)$

$$h^+(b) \equiv \lim \sup_{\epsilon \to 0^+} \frac{h(b+\epsilon) - h(b)}{\epsilon} > 0$$

where we allow the possibility that $h^+(b) = \pm\infty$ so that this is well defined. Generalized convex functions cannot achieve a maximum in $(0, \bar{b}]$ since $h(b) \geq h(0)$ implies $h^+(b) > 0$.

A contest is *generalized convex* if for each contestant k and all $b_{-k} > 0$ the objective $p(b_k, b_{-k}) - \varrho_k(b_k)$ is generalized convex as a function of b_k up to b_{-k}. If cost is strictly positive for $b_1 > 0$ the all-pay auction is generalized convex: the condition $p(b_k, b_{-k}) - \varrho_k(b_k) \geq p(0, b_{-k})$ is never satisfied for $0 < b_k < b_{-k}$, while at $b_k = b_{-k}$ the right derivative is positive infinity.

Hirshleifer (1989)'s argument suggests that contest success functions should be generalized convex. This condition is satisfied by many standard contest success functions. Ewerhart (2017) studies continuously differentiable contest success functions and shows that if the elasticity of the odds ratio with respect to own effort is globally larger than 2 then generalized convexity holds. He shows that if $\beta > 2$ this elasticity condition is satisfied by the Tullock function. He shows in that case that it is also satisfied by the *serial* contest success function $p(b_k, b_{-k}) = (1/2)(b_k/b_{-k})^\beta$ for $b_k < b_{-k}$ studied by Alcalde and Dahm (2007).[8]

Generalized convexity not only applies to discontinuous contest success functions, it is weaker than the elasticity condition even for continuously differentiable functions. For example, while the serial contest success fails the elasticity condition for $\beta \leq 2$ it is continuously differentiable and for $b_k \leq b_{-k}$ and $\beta > 1$ strictly convex in b_k so it is generalized convex even for $1 < \beta \leq 2$.

If the contest success function is generalized convex and the cost functions are not "too convex," and certainly if they are weakly concave, then the contest will be generalized convex.[9]

Let us say that a contest is *insensitive* if for each contestant k and $b_{-k} > 0$ we have $p(0, b_{-k}) = q < 1/2$. This is a strong condition but is satisfied in cases such as the Tullock and serial cases where $q = 0$, and more generally in any ratio form contest success function with the condition that a zero ratio yields a probability of success strictly lower than $1/2$. Notice also that if we add q-global noise to any insensitive contest the property of insensivity will be retained.

If G_k, G_{-k} are an equilibrium, we write $\hat{u}_k = p(G_k, G_{-k}) - \varrho_k(G_k)$ for the corresponding normalized utility. The results of Hirshleifer (1989), Alcalde and Dahm (2007) and Ewerhart (2017) are generalized in Levine and Mattozzi (2019) and we add to that a partial converse:

THEOREM 11.7.1. *(i) If a contest is generalized convex then in any equilibrium there is at least one contestant who provides effort with positive probability in every interval containing zero.*

(ii) If in addition the contest is insensitive then in any equilibrium G_k, G_{-k} neither contestant uses a pure strategy and there is a less successful contestant $-k$ who receives $\hat{u}_{-k} = q$ and and more successful contestant k who receives $\hat{u}_k = (1 - G_{-k}(0))q + G_{-k}(0)(1 - q)$.

(iii) If in addition 1 has a homogeneous cost advantage $\varrho_1(b_1) = \nu\varrho_{-1}(b_1)$ then in any equilibrium 1 is more successful, has greater outcome success, and $G_1(0) = 0$ and $G_{-1}(0) = 1 - \nu$, and wins with probability $(1 - \nu)(1 - q) + \nu/2$.

Part (iii) says is that if we retain the generalized convexity and insensitivity property of the generalized all-pay auction but assume a homogeneous cost advantage the contest is payoff equivalent to the generalized all-pay auction with the same costs and q. Note that there can be no peaceful equilibrium when the contest is insensitive: if the both are providing zero effort a small effort raises the probability of winning from $1/2$ to $1 - q$.

The proof of Theorem 11.7.1 uses the following technical lemma from Levine and Mattozzi (2019) proved as Lemma 11.12.11 in the Appendix to this Chapter:

[8] Feddersen and Sandroni (2006) study $\beta = 1$ with quadratic cost.

[9] For example cost functions are linear in Ewerhart (2017).

LEMMA 11.7.2. *Suppose that either G_{-j} does not have an atom at b or p is continuous at (b,b). Then $p(b_j, F_{-j})$ as a function of b_j is right continuous at $b_j = b$.*

PROOF. Suppose the contest is generalized convex. Define $\underline{b}_j$ to be the lowest point of support in the equilibrium G_j. If $\underline{b}_j > 0$ then for $-j$ the objective $p(b_{-j}, G_j) - \varrho_{-j}(b_{-j})$ is generalized convex up to $\underline{b}_j$ meaning that it is strictly suboptimal for $-j$ to provide effort in $(0, \underline{b}_j]$. This implies that either one of the $\underline{b}_j$'s is zero or both are equal. If both are equal, Theorem 11.3.1 and Lemma 11.7.2 imply that for one j the function $p(b_j, G_{-j})$ is continuous in b_j at $\underline{b}_j$ hence so is the objective function. Since $\underline{b}_j$ is strictly suboptimal and $p(b_j, G_{-j})$ is continuous there, it follows that there is an $\epsilon > 0$ such that b_j is strictly suboptimal in $[\underline{b}_j, \underline{b}_j + \epsilon]$ contradicting the definition of $\underline{b}_j$.

Suppose next that $\underline{b}_{-j} = 0$, that the contest is insensitive and that $\underline{b}_j > 0$. We will show this is impossible.

Since $p(0, b_{-j})$ is constant for $b_{-j} > 0$ and $-j$ does not provide effort in $(0, \underline{b}_j]$ define the function $v_j(b_j) = p(b_j, G_{-j}) - \varrho_j(b_{-j})$ for $b_j > 0$ and $v_j(0) = \lim_{b_j \to 0} p(b_j, G_{-j}) - \varrho_j(b_j)$. This is generalized convex up to $\underline{b}_j$.

If p is discontinuous at $(\underline{b}_j, \underline{b}_j)$ and $-j$ has an atom there then j does not by Theorem 11.3.1. It follows from Lemma 11.7.2 that there is an $\epsilon > 0$ such that b_j is strictly suboptimal in $[\underline{b}_j, \underline{b}_j + \epsilon]$. Hence v_j is in fact generalized convex up to $\underline{b}_j + \epsilon$, so for $\hat{b}_j \in [\underline{b}_j, \underline{b}_j + \epsilon]$ we have $v_j(\hat{b}_j) < \lim_{b_j \to 0} p(b_j, G_{-j}) - \varrho_j(b_j)$. Hence $\hat{b}_j$ is not optimal. This contradicts the definition of $\underline{b}_j$.

If either p is continuous at $(\underline{b}_j, \underline{b}_j)$ or $-j$ has no atom there, the generalized convexity of v_j up to $\underline{b}_j$ implies that $v_j(\underline{b}_j) < \lim_{b_j \to 0} p(b_j, G_{-j}) - \varrho_j(b_j)$. It follows from Lemma 11.7.2 that there is an ϵ' so that for $\hat{b}_j \in [\underline{b}_j, \underline{b}_j + \epsilon']$ we have $v_j(\hat{b}_j) < \lim_{b_j \to 0} p(b_j, G_{-j}) - \varrho_j(b_j)$. Hence $\hat{b}_j$ is not optimal, again contradicting the definition of $\underline{b}_j$. As all cases have been covered, we conclude that $\underline{b}_j = 0$ for both contestants.

We next derive the equilibrium normalized utility under the insensitivity assumption. Fix j and choose a positive sequence $b_j^n, b_{-j}^n \to 0$ such that $p_j(b_j^n, b_{-j}^n) \to q$. Since $\underline{b}_j = 0$ the support of G_j must contain points arbitrarily near 0. Hence for both contestants we can choose a sequence $\tilde{b}_j^n \le b_j^n$ in the support of G_j and this implies that $p_j(\tilde{b}_j^n, G_{-j}) - \varrho_j(\tilde{b}_j^n) = \hat{u}_j$. Since cost is continuous

$$\hat{u}_j = \liminf p_j(\tilde{b}_j^n, G_{-j}) = \liminf \int_{0<b_{-j}} \min\{q, p_j(\tilde{b}_j^n, b_{-j})\} dG_{-j}(b_{-j}) + G_{-j}(0)(1-q)$$

$$+ \int_{0<b_{-j}<b_{-j}^n} \left[p_j(\tilde{b}_j^n, b_{-j}) - \min\{q, p_j(\tilde{b}_j^n, b_{-j})\} \right] dG_{-j}(b_{-j})$$

$$+ \int_{b_{-j} \ge b_{-j}^n} \left[p_j(\tilde{b}_j^n, b_{-j}) - \min\{q, p_j(\tilde{b}_j^n, b_{-j})\} \right] dG_{-j}(b_{-j})$$

$$\le (1 - G_{-j}(0))q + G_{-j}(0)(1-q).$$

The second line vanishes in the limit since the range of integration goes to zero, the third line because it is bounded above by $|p_j(b_j^n, b_{-j}^n) - q|$ which goes to zero by construction. Since nearly $(1-G_{-j}(0))q+G_{-j}(0)(1-q)$ is obtained by providing near zero effort, it follows that in fact

$$\hat{u}_j = (1-G_{-j}(0))q+G_{-j}(0)(1-q). \tag{11.7.1}$$

Since insensitivity implies discontinuity at zero, by Theorem 11.3.1 both contestants do not have an atom at zero. If j has no atom then $-j$ gets q. If $-j$ provides zero effort with probability one then j has no best response so this is not an equilibrium. Since $\underline{b}_j = 0$ it must be that j is mixing as well. This proves (ii).

To prove (iii) , observe that if cost is homogeneous it follows from Theorem 11.6.3 (ii) that 1 must be more successful. Hence -1 gets q and if 1 had an atom at zero -1 could get nearly $(1-G_1(0))q+G_1(0)(1-q)$ by providing near zero effort. That is to say, 1 cannot have an atom at zero. The final part of the argument is derived from Ewerhart (2017) and Alcalde and Dahm (2007). Consider the contest in which 1 has cost $(\nu/(1-G_{-1}(0)))\varrho_{-1}(b_1)$. We then modify -1's strategy to get rid of the atom taking the strategies to be G_1 and $(G_{-1}-G_{-1}(0))/(1-G_{-1}(0))$ and observing that these are an equilibrium of this modified game. Hence both contestants get q as neither has an atom at zero. By Theorem 11.6.3 this implies $\nu/(1-G_{-1}(0)) = 1$. This in turn implies the modified game is symmetric so by Corollary 11.6.5 each wins with equal probability. It follows that 1 wins with probability $(1-\nu)(1-q)+\nu/2$ in the original game as asserted. Since an average of $1-q$ and $1/2$ is bigger than $1/2$ this means outcome success. $\square$

For the case of the Tullock function we can give a simple condition on the cost function that insures that the contest is generalized convex for sufficiently large β. Denote the nth derivative of $\varrho_k(b_k)$ by $\varrho_k^{(n)}(b_k)$.

THEOREM 11.7.3. *For any fixed smooth convex cost functions with $\varrho_k^{(n)}(0) \neq 0$ for some $n \geq 1$ and the Tullock contest success function there exists a $\overline{\beta}$ such that for $\beta > \overline{\beta}$ the contest is generalized convex.*

Note that in case $\varrho_k(b_k)$ is real analytic the assumption $\varrho_k^{(n)}(0) \neq 0$ for some $n \geq 1$ is always satisfied: if $\varrho_k^{(n)}(0) = 0$ for all n then $\varrho_k(b_k)$l is identically zero, which is inconsistent with our assumption that it must be somewhere strictly increasing. There are, however, increasing smooth functions such as $\exp(-1/b_k)$ for which derivatives of all order vanish.

PROOF. The objective function is

$$u_k(b_k, b_{-k}) = \frac{b_k^\beta}{b_k^\beta + b_{-k}^\beta} - \varrho_k(b_k).$$

The condition for generalized convexity is that is for $0 < b_k \leq b_{-k} \leq \overline{W}$ and $u_k(b_k, b_{-k}) \geq u_k(0, b_{-k}) = 0$ we have the derivative $D_k u_k(b_k, b_{-k}) > 0$ which is to say

$$\beta \frac{b_k^{\beta-1} b_{-k}^\beta}{\left(b_k^\beta + b_{-k}^\beta\right)^2} - \varrho_k'(b_k) > 0$$

or equivalently

$$\beta \frac{b_k^{\beta} b_{-k}^\beta}{\left(b_k^\beta + b_{-k}^\beta\right)^2} - \varrho_k'(b_k) b_k > 0 \tag{11.7.2}$$

Since we may assume $u_k(b_k, b_{-k}) \geq 0$ and $b_{-k}^\beta / \left(b_k^\beta + b_{-k}^\beta\right) \geq 1/2$ we have

$$\beta \frac{b_k^\beta b_{-k}^\beta}{\left(b_k^\beta + b_{-k}^\beta\right)^2} - \varrho_k'(b_k) b_k \geq \frac{\beta}{2} \varrho_k(b_k) - \varrho_k'(b_k) b_k \equiv d_k(b_k).$$

It suffices to show that $d_k(b_k) > 0$ for sufficiently large β.

There are two cases depending on whether b_k is small or large. The small case is first. Write

$$d_k(b_k) = \left(\frac{\beta}{2} \frac{\varrho_k(b_k)}{\varrho_k'(b_k) b_k} - 1 \right) \varrho_k'(b_k) b_k.$$

As $b_k \to 0$ apply l'Hospital's rule to

$$\frac{\varrho_k(b_k)}{\varrho_k'(b_k) b_k} \to \frac{\varrho_k'(b_k)}{\varrho_k'(b_k) + \varrho_k''(b_k) b_k}. \tag{11.7.3}$$

If $\varrho_k'(0) = 0$ iterate until we find the first n with $\varrho_k^{(n)}(0) \neq 0$. Since $\varrho_k(b_k)$ is non-decreasing it cannot be that $\varrho_k^{(n)}(0) < 0$, so $\varrho_k^{(n)}(0) > 0$. This gives

$$\frac{\varrho_k(b_k)}{\varrho_k'(b_k) b_k} \to \frac{\varrho_k^{(n)}(b_k)}{n \varrho_k^{(n)}(b_k) + \varrho_k^{(n+1)}(b_k) b_k} \to 1/n.$$

Hence we may choose $\underline{b}_k > 0$ such that for $0 < b_k \leq \underline{b}_k$ we have

$$\frac{\varrho_k(b_k)}{\varrho_k'(b_k) b_k} > 1/2.$$

Hence for $0 < b_k \leq \underline{b}_k$ and $\beta > 4n$ we have $d_k(b_k)$ positive.

The large case is $b_k > \underline{b}_k$. Define $\overline{\varrho}$ and $\underline{\varrho}$ to be the max and min of $\varrho'(b_k)$ on $[\underline{b}_k, \overline{W}]$ and observe that both are finite and non-zero. Here we use

$$d_k(b_k) = \frac{\beta}{2} \varrho_k(b_k) - \varrho_k'(b_k) b_k \geq \frac{\beta}{2} \underline{\varrho} \underline{b}_k - \overline{\varrho} \overline{W}.$$

Hence for $\beta > 2\overline{\varrho}\overline{W}/(\underline{\varrho}\underline{b}_k)$ we have $d_k(b_k)$ positive.

The theorem now follows by taking $\overline{\beta} = \max\{4n, 2\overline{\varrho}\overline{W}/(\underline{\varrho}\underline{b}_k)\}$. □

11.8. Robustness and the Equilibrium Correspondence

In order to discuss robustness we must perturb the model. It is easiest to do the by introducing sequences of contests $p_n(b_1, b_{-1}), \varrho_{1n}(b_1), \varrho_{-1n}(b_{-1})$. We first give a slightly more formal definition of a contest. A *contest* on $\overline{W}$ is a contest success function $p(b_k, b_{-k}) \geq 0$ for $0 \leq b_1, b_{-1} \leq \overline{W}$, which is non-decreasing in the first argument, continuous except possibly at $b_k = b_{-k}$, and satisfying the adding-up condition $p(b_k, b_{-k}) + p(b_{-k}, b_k) = 1$ together with a pair of cost functions $\varrho_k(b_k) \geq 0$ non-decreasing and continuous with $\varrho_k(0) = 0$, $\varrho_k(\overline{W}) > 1$, and ϱ_{-1} strictly increasing at 0. For a contest on $\overline{W}$ we take the strategy space to be of cumulative distribution functions on $[0, \overline{W}]$. In the Appendix to this Chapter we show

THEOREM 11.8.1. *Suppose $p_n, p_0, \varrho_{kn}, \varrho_{k0}$ are a sequence of contests in $\overline{W}$ with $p_n(b_1, b_{-1}) \to p_0(b_1, b_{-1})$, $\varrho_{kn}(b_k) \to \varrho_{k0}(b_k)$ for each $0 \leq b_1, b_{-1} \leq \overline{W}$ and that G_{1n}, G_{-1n} are equilibria for p_n, ϱ_{kn} converging weakly to G_{10}, G_{-10}. Then $p_n(G_{kn}, G_{-kn}) \to p_0(G_{k0}, G_{-k0})$, $\varrho_{kn}(G_{kn}) \to \varrho_{k0}(G_{k0})$ and G_{10}, G_{-10} is an equilibrium for $p_0(b_1, b_{-1}), \varrho_{k0}(b_k)$.*

We should emphasize that this result requires only the pointwise convergence of p_n, ϱ_{kn}. Pointwise convergence is easy to check, but has strong consequences for non-decreasing functions on rectangles: if the limiting function is continuous the convergence is uniform. Even if the limiting function is discontinuous on the diagonal - as we allow for contest success function - convergence is uniform on the set of effort pairs that is bounded away from the diagonal.

As an example the Tullock contest success function $b_k^\beta/(b_k^\beta + b_{-k}^\beta)$ converges pointwises to the all-pay auction as $\beta \to \infty$, so in any sequence of equilibria the payoff of -1 converges to zero and that of 1 to $1 - \varrho_1(W_{-1})$. Say for $b > 0$ that a conflict resolution function is *perturbed* Tullock if $p(b_k, b_{-k}) = (b + b_k)^\beta / \left((b_k + b)^\beta + (b_{-k} + b)^\beta\right)$ where recall that $\beta > 0$.[10] Alternatively, a conflict resolution function is *perturbed serial* if $p(b_k, b_{-k}) = (1/2)((b_k + b)/(b_{-k} + b))^\beta$ for $b_k < b_{-k}$. Notice that both of these functions are continuous but fail the insensitivity condition of Theorem 11.7.1, never-the-less that theorem together with Theorem 11.8.1 imply the following:

Corollary 11.8.2. *Suppose that the conflict resolution function is perturbed Tullock with $\beta > 2$ or perturbed serial with with $\beta > 1$ and that 1 has a homogeneous cost advantage $\varrho_1(b_1) = \nu\varrho_{-1}(b_1)$. Then in the limit as $b \to 0$ in any sequence of equilibria the utility of 1 converges to $1 - \nu$ and that of -1 to zero.*

11.8.1. Well-behaved Contests. We say that a contest is *well-behaved* if $p(b_k, b_{-k}) > 0$, p is strictly increasing in the first argument, ϱ_k is strictly increasing, and both have an extension to an open neighborhood of $[0, \overline{W}] \times [0, \overline{W}]$ that is real analytic. Some contest success functions studied in the literature have real analytic extensions. This is true of the perturbed Tullock function. The *logit function*

$$p(b_k, b_{-k}) = \frac{\exp(\beta b_k)}{\exp(\beta b_k) + \exp(\beta b_{-k})}$$

introduced by Hirshleifer (1989) is another example. Notice that like the Tullock function as $\beta \to \infty$ the logit function converges pointwise to the all-pay auction. Another example can be found in Shachar and Nalebuff (1999) who take

$$p_k(b_k, b_{-k}) = H\left(\frac{1}{2} + \frac{\exp(b_k) - \exp(b_{-k})}{\exp(b_k) + \exp(b_{-k})}\right)$$

where H is a cdf with support in $[0, 1]$. If the cdf H is symmetric around $1/2$ then $p_k(b_k, b_{-k})$ is a contest success function, and if in addition H admits a real analytic extension to $(-\epsilon, 1 + \epsilon)$ then so does $p_k(b_k, b_{-k})$.

Other contest success functions studied in the literature are not well-behaved either being discontinuous as is the case with the all-pay auction and Tullock function, or having discontinuities in the derivatives as is the case with the *quasi-linear* function $P(b_k - b_{-k})$ which is linear when it is not 0 or 1. Never-the-less Levine and Mattozzi (2019) show that all contests can be approximated by well behaved contests:

Theorem 11.8.3. *If p, ϱ_k is a contest on $\overline{W}$ then there is a sequence of well-behaved contests p_n, ϱ_{kn} on $\overline{W}$ with $p_n(b_k, b_{-k}) \to p(b_k, b_{-k})$, $\varrho_{kn}(b_k) \to \varrho_k(b_k)$ for every $(b_1, b_{-1}) \in [0, \overline{W}] \times [0, \overline{W}]$.*

Corollary 11.8.4. *A Nash equilibrium exists.*

[10] As for example in Amegashie (2006).

PROOF. Since real analytic functions are continuous the Glicksberg fixed point theorem implies the existence of Nash equilibrium in the real analytic case. Theorem 11.8.3 and Theorem 11.8.1 imply existence for the general case. □

We show in the Appendix that well-behaved contests have a relatively simple equilibrium structure:

THEOREM 11.8.5. *Suppose that $\varrho_1(b_1) = 0$ for $0 \leq b_1 \leq \underline{W}_1$ and if $\underline{W}_1 > 0$ we require that $p(b_k, b_{-k})$ is strictly increasing in the first argument (so in particular in any equilibrium $\lim_{W \to \underline{W}_1^-} G_1(W) = 0$). If $p(b_k, b_{-k}), \varrho_k(b_k)$ have real analytic extensions to an open neighborhood of $[\underline{W}_1, \overline{W}] \times [0, \overline{W}]$ then every equilibrium has finite support.*

We note that the finiteness property holds also for some contests that are not well-behaved. Che and Gale (2000) show that with quasi-linear contest success and linear costs there is an equilibrium with finite support and they explicitly compute it. Ashworth and Mesquita (2009) extend that analysis to the case where one contestant has a head start advantage. Ewerhart (2015) who developed the technique we use in the appendix analyzed the symmetric Tullock contest for large β. That function is not well-behaved since it is discontinuous at zero and without the extension of analyticity below zero the finiteness result fails: with linear costs Ewerhart (2015) shows that the support is countable with a single accumulation point at zero and explicitly computes the equilibrium.

11.8.2. Robust Properties. We are interested in understanding properties of contests that are robust. By a *property* we mean a statement $\Pi(p, \varrho, G)$ such as: there is complete rent dissipation, contestant 1 has greater success, or one contestant has zero utility. We say that a property is true in a contest if it is true for all equilibria of the game. We say that a property in p, ϱ is *robust* if whenever it is true in p, ϱ then for every sequence p_n, ϱ_n converging pointwise to p, ϱ and for n sufficiently large the property is true in p_n, ϱ_n.

COROLLARY 11.8.6. *Any strict inequality concerning equilibrium utility, probability of winning, or cost is robust.*

PROOF. Suppose not. Then there exists a subsequence in which $\Pi(p_n, \varrho_n, G_n)$ is false. Since the space of strategies is compact every subsequence contains a further subsequence that converges weakly to some G. By Theorem 11.8.1 G is an equilibrium and utility, winning probability, and cost converge. Hence as the strict inequality is presumed to be satisfied for G for all sufficiently large n it was satisfied for $\Pi(p_n, \varrho_n, G_n)$, a contradiction. □

An important implication of Theorem 11.8.1 and Corollary 11.8.6 is that if p_n converges to the all-pay auction holding fixed costs ϱ_k then utilities and the probability of winning approach those of the all-pay auction.

11.9. Rent Dissipation

As indicated, rent dissipation plays an important role in political economy. Although complete rent dissipation is often associated with symmetry and the all pay auction, interestingly symmetry, discontinuity, and mixed strategy equilibria are not needed for complete rent dissipation. Indeed any positive pure strategy pair can be turned into a pure strategy equilibrium with full dissipation. Specifically, if $p(b_k, b_{-k})$ is a contest success function with $p(0, b_{-k}) = 0$ and continuous for $(b_k, b_{-k}) \neq 0$, for example the Tullock function, and $\hat{b}_k, \hat{b}_{-k} > 0$ then there are cost functions $\varrho_k(b_k), \varrho_{-k}(b_{-k})$ such that $(\hat{b}_k, \hat{b}_{-k})$ is a pure strategy equilibrium with complete rent dissipation. An example is to take $\varrho_k(b_k) = p_k(b_k, \hat{b}_{-k})$ on $[0, 2\hat{b}_k]$ and $\varrho_k(b_k) = p_k(b_k, \hat{b}_{-k}) + b_k$ for $b_k > 2\hat{b}_k$.

Also important in the literature has been the weaker situation in which one contestant gets nothing - this is the case in every all pay auction, symmetric or not. It turns out that the possibility of a contestant getting nothing is quite exceptional. We say that a property is *generic* if it is robust and if for any $p, \varrho_1, \varrho_{-1}$ for which it is not true there is a sequence p_n, ϱ_{kn} converging pointwise to p, ϱ_k in which it is true.

We formally define properties corresponding to dissipation:

(1) *no dissipation*: in equilibrium $\varrho_1(G_1) + \varrho_{-1}(G_{-1}) = 0$
(2) *partial dissipation*: in equilibrium $0 < \varrho_1(G_1) + \varrho_{-1}(G_{-1}) < 1$
(3) *some dissipation*: in equilibrium $0 < \varrho_1(G_1) + \varrho_{-1}(G_{-1})$
(4) *complete dissipation*: in equilibrium $\varrho_1(G_1) + \varrho_{-1}(G_{-1}) = 1$
(5) *γ-dissipation*: in equilibrium $\varrho_1(G_1) + \varrho_{-1}(G_{-1}) > \gamma$ where $0 \leq \gamma < 1$

Notice that complete dissipation means γ-dissipation for every $0 \leq \gamma < 1$. Moreover, contested equilibrium implies some dissipation. If in addition one contestant has greater success then there is partial dissipation. Recall that robustness and genericity concern a property that applies to all equilibria. We have

THEOREM 11.9.1. *Concerning rent dissipation:*
(i) there is a subset of contests with no dissipation that are robust
(ii) the entire set of contests with some (or partial) dissipation is robust
(iii) contests without complete dissipation are generic
(iv) contests with γ-dissipation are robust

PROOF. (i) The property of very high cost for k is $\varrho_k(b_k) > \sup_{b_{-k}} p(b_k, b_{-k}) - p(0, b_{-k})$ which is robust by Corollary 11.8.6. By Theorem 11.5.2 if both contestants have very high cost there is a unique peaceful equilibrium and hence no dissipation.

Part (ii) follows directly from Corollary 11.8.6 and the fact that some (partial) dissipation is defined by a strict cost inequality

For (iii) we show the slightly stronger result that both contestants getting positive utility is generic. Strict inequality concerning utility is robust by Corollary 11.8.6: this proves that both contestants getting positive utility is robust. We will show that for any p_0, ϱ_{k0} there is a sequence p_n, ϱ_{kn} converging uniformly to p_0, ϱ_{k0} in which each contestant gets positive utility in every equilibrium, and this will complete the proof.

For costs we take $\varrho_{kn} = \varrho_{k0}$. Then take $1 > \lambda_n > 0$ to be a sequence converging to zero and define

$$p_n(b_k, b_{-k}) = (1 - \lambda_n) p_0(b_k, b_{-k}) + \lambda_n \Phi(b_k - b_{-k})$$

where Φ is the standard normal cumulative distribution function. This obviously converges uniformly to $p_0(b_k, b_{-k})$. Moreover, for $0 \leq b_k \leq \overline{W}$ we have $p_n(b_k, b_{-k}) \geq \lambda_n \Phi(-\overline{W})$. Hence providing zero effort gets at least $\lambda_n \Phi(-\overline{W}) > 0$ so this is obtained by both contestants in any equilibrium.

The proof of (iv) follows from taking an anomalous subsequence and then finding one on which G_n converges. $\square$

Notice that (iii) states that complete dissipation is not robust and (iv) that contests near those with complete dissipation - so for example close to symmetric all pay - have nearly complete dissipation.

11.10. Extensions

There are a number of variations on the basic model that can be analyzed using the methods here.

11.10.1. Asymmetric Contests. So far we have discussed contests that are symmetric in the sense that the probability of winning depends on the efforts of the two contestants and not on their names. Not all contest success functions examined in the political economy literature satisfy this property. It may be, fore example, that one contestant has a advantage in providing effort. For example, we might let $h(b_1)$ be a strictly increasing continuous function with $h(b_1) \geq b_1$ and assume that the probability of 1 winning is $p(h_1(b_1), b_{-1})$. This represents the idea that effort by 1 is "worth more" than effort by -1: for example, in a political contest because 1 has a more appealing platform or more attractive candidate.

In this case, and in a variety of others, the asymmetric contest can be mapped to an equivalent symmetric contest. The key point is that the units in which effort is measured do not matter. While it might be natural from an economic point of view to identify effort with number of voters, hours devoted to the cause, or amount of money contributed, the model does not care about the units. Specifically, if we let $h(b)$ denote a continuous strictly increasing function with $h(0) = 0$, that is, a strictly increasing cost function, then the contest $p(h(b_k), h(b_{-k})), \varrho_k(h(b_k))$ is *equivalent* to the contest $p(b_k, b_{-k}), \varrho_k(b_k)$ in the sense that any equilibrium in one contest can be transformed to an equilibrium of the other contest with exactly the same probabilities of winning and costs. If an equilibrium strategy of the h contest is denoted by $G_{hk}(b_k)$ we can map the equilibrium strategies by $G_{hk}(h(b_k)) = G_k(b_k)$ and $G_{hk}(b_k) = G_k(h^{-1}(b_k))$. In particular, if for contestant -1 the cost $\varrho_{-1}(b_{-1})$ is strictly increasing, we can take $h(b_{-k}) = \varrho_{-1}^{-1}(b_{-k})$ in which case the cost function of -1 is linear and given by $\varrho_{h-1}(b_k) = b_k$. Notice the implication that the statement "cost is convex" has no real meaning in a contest: we can change the units of cost so as to make cost concave or convex and get an equivalent contest by suitably modifying the contest success function.

In a similar vein, when the probability of winning is asymmetric of the form $p(h_1(b_1), b_{-1})$ we can find a symmetric contest that is equivalent. Here we create a new contest with contest success given by $p(b_1, b_{-1})$ and redefine cost for 1 as $\varrho_1(h^{-1}(b_1))$ for $b_1 \geq h(0)$ and 0 otherwise. Since it may be that there is an advantage even at zero we need to allow the possibility that ϱ_1 is flat up to $h(0)$.

Not all asymmetric contests can be mapped back to the symmetric setting, however. For example, the model of Coate and Conlin (2004b) maps to a symmetric contest only if the parties are of equal expected size.

11.10.2. Resource Limits. We have ruled out resource limits, but we can approximate their effect by assuming that cost grows rapidly, and in particular becomes greater than the value of the prize, as the limiting effort level is approached. For these approximations our assumptions are satisfied so our results hold. In particular Levine and Mattozzi (2020) in their online appendix compare two continuous participation cost functions: $D(\varphi)$, $\tilde{D}(\varphi)$ where for some $\overline{\varphi} < 1$ and $\varphi \leq \overline{\varphi}$

we have $D(\varphi) = \tilde{D}(\varphi)$ while for $\overline{\varphi} < \varphi \leq 1$ we have $D(\varphi) < \tilde{D}(\varphi)$. The cost function $D(\varphi)$ is bounded, but they allow $\tilde{D}(1) = \infty$. They note that this property is preserved when monitoring costs are taken account of and prove that equilibrium in the all-pay auction are close when $\overline{\varphi}$ is close to 1.

More generally, a model with a truncated effort level is equivalent to a model in which cost is discontinuous at the truncation point, jumping to a level greater than the value of the prize. Specifically, we now wish to consider the possibility that ϱ_k instead of being continuous on the whole support, it is continuous on $[0, \overline{b}_k]$ where $\overline{b}_k > 0$, $\varrho_k(\overline{b}_k) = \overline{\varrho}_k < 1$, and for $b_k > \overline{b}_k$ we have $\varrho_k(b_k) = \varrho_{Max} > 1$.

There are two cases. It may be that resource constraints do not bind in the sense that desire to bid $B_k \leq \overline{b}_k$. In this case nothing changes. If resource constraints do bind our our arguments on cost advantage that require each player be able to employ the strategy of the other fail. However, the result on strong cost advantage can be saved. Define the willingness to bid as $W_k = \min\{\overline{b}_k, B_k\}$ where now $\varrho_k(B_k) = 1$ defines the desire to bid rather than the willingness to bid.

THEOREM 11.10.1. *In a contested equilibrium 1 has greater success if* 1 *has a strong cost advantage.*

PROOF. If 1 has a strong cost advantage then there is a $\hat{b}_1$ with $\varrho_1(\hat{b}_1) < p(\hat{b}_1, W_{-1}) - p(W_{-1}, W_{-1}) = p(\hat{b}_1, W_{-1}) - 1/2$. Note in particular that in order for this to hold it must be that $\overline{b}_1 > W_{-1}$ or there can be no such $\hat{b}_1$.

Then $p(\hat{b}_1, W_{-1}) - \varrho_1(\hat{b}_1) > 1/2$. Observe that $G_{-1} \leq W_{-1}$ so $p(\hat{b}_1, W_{-1}) \leq p(\hat{b}_1, G_{-1})$. Finally, from optimality

$$p(G_1, G_{-1}) - \varrho_1(G_1) \geq p(\hat{b}_1, G_{-1}) - \varrho_1(\hat{b}_1) \geq p(\hat{b}_1, W_{-1}) - \varrho_1(\hat{b}_1) > 1/2$$

which as both contestants cannot have a utility greater than 1/2 implies greater success. □

If either their are no binding resource limits or the contest success function itself is continuous Levine and Mattozzi (2019) in their Appendix on resource limits show that the robustness results continue to hold. This leaves the issue of robustness when both p and ϱ are discontinuous, and here we can go no further. The following example that upper hemi-continuity of the equilibrium correspondence can fail in that case.

EXAMPLE 11.10.2. Let the contest success function be that of the all-pay auction, and fix a cost function for both contestants that is linear with constant unit marginal cost up to a resource limit of $\overline{b}_k > 0$. We normalize the value of the prize to be 1, assume that $\overline{b}_k + \overline{b}_{-k} = 1/2$ and let $\varrho_{Max} = 2$. This means that both contestants want to violate their resource limits.

Suppose first that $\overline{b}_1 > \overline{b}_{-1}$. In this case contestant 1 receives a utility of at least 3/4 and contestant -1 gets a utility of 0. Moreover, it is well-known that in the unique equilibrium effort is uniform in $(0, \overline{b}_{-1})$ and that -1 has an atom at zero and 1 has an atom at $\overline{b}_{-1}$. The implication of the non-trivial atom at $\overline{b}_{-1}$ means, however, that the tie-breaking rule that each contestant has an equal chance of winning in case of a tie is not consistent with equilibrium. If that is the tie-breaking rule, then -1 should chooses $\overline{b}_{-1}$, guaranteeing at least a 50% chance of winning, and so earning at least $1/2 - 1/4 > 0$ rather than zero as the equilibrium requires. In fact the tie-breaking rule at $\overline{b}_{-1}$ must favor contestant 1 at least to the extent that contestant -1 cannot profit from that effort level. In other words: when both p and ϱ are discontinuous we must allow for endogenous tie-breaking rules.

Second, consider what happens as we pass through the point of symmetry. For $\bar{b}_1 > \bar{b}_{-1}$ contestant 1 earns at least 3/4 and contestant -1 earns nothing; at $\bar{b}_1 = \bar{b}_{-1}$ both contestants earn 1/4, while for $\bar{b}_1 < \bar{b}_{-1}$ contestant 1 earns 0 and contestant -1 earns at least 3/4. In other words: both the individual and aggregate payoff are discontinuous as we pass through the point of symmetry.

Finally, suppose that we approximate the discontinuous cost functions by functions that are linear up to $\bar{b}_k - \epsilon$ then rise steeply to ϱ_{Max} as $\bar{b}_k$ is approached. Levine and Mattozzi (2020) show that in this case as long as $\bar{b}_1 \neq \bar{b}_{-1}$ payoffs are well-behaved in the limit. However, this is not the case when there is symmetry. If $\bar{b}_1 = \bar{b}_{-1}$ then with continuous cost there is complete rent dissipation: both players get zero. However, in the limit both contestants get $1/2 - 1/4$ so we have equilibria with complete rent dissipation converging to one where both contestants get a positive rent.

11.11. Random Turnout Models

Often contest success functions are derived from more primitive considerations. Jia et al. (2013), for example, derive contest success functions from a set of axioms. Of particular importance are models of random turnout. Conflict resolution models are agnostic as to where the uncertainty comes from: it may come from random turnout - or it might come from something else entirely, such as a decision by a court about what votes to count. Random turnout models are specific - and turnout is observed, so random turnout models make predictions about the distribution of turnout that conflict resolution models do not. As turnout is observed, to the extent that random turnout is important (and not court decisions) this greater precision is a good thing. However some care must be taken in case other forms of randomness should be important.

The idea behind random turnout is that $b_k$represents intended effort while actual effort is a random function of of intended effort. This can be due to random shocks to voter preferences, such as in Coate and Conlin (2004b) or from random shocks to the size of the group as in Feddersen and Sandroni (2006). Since random shocks to voter preferences lead to random shocks to the number of actual voters we focus on random shocks to group size.

Must a Self-Organizing Group Know How Large It Is? In the theory of social mechanisms a fundamental idea is that people to not act individually but act collectively as members of a group. Does it make sense to think that such a group does not know how large it is, that is, who its members are? For a small group, such as a committee, surely group members know how many there are. For large groups no matter how well organized this need not be the case. For example, farmers form a well-organized lobby, set collective objectives and cajole one another with peer pressure to engage in lobbying. It is not that all farmers know each other or operate in a single collective: rather they interact through a connected network as in Section 9.4. As a result nobody can really say for sure exactly how many farmers there are in the farm lobbying organization.

A second consideration is that even if the number of group members is known, not all may be available to act. We have modelled individual shocks that increase participation costs, but there can also be aggregate shocks - for example if a highway is blocked by a storm in a particular neighborhood on election day people none of the people in that neighborhood will be available to vote.

To illustrate the idea, suppose that an election is determined by the greatest number of votes cast and that the number of votes cast is a random function of the level of effort. This might be because in a finite population independent draws by voters of their type results in randomness in the number of votes, or because of other random factors such as weather the impact on voter turnout. Specifically, if the number of votes cast is $v_k = b_k + \epsilon_k$ then the probability that 1 wins the election is $\Pr\{b_1 + \epsilon_1 > b_{-1} + \epsilon_{-1}\}$ or $\Pr\{\epsilon_{-1} - \epsilon_1 < b_1 - b_{-1}\}$. If ϵ_1 and ϵ_{-1} are iid then $\epsilon_k - \epsilon_{-k}$ is symmetrically distributed, and indeed this can be the case even if ϵ_1 and ϵ_{-1} are correlated. This leads to *difference* model $p(b_k, b_{-k}) = P(b_k - b_{-k})$.

Difference models have been widely used, for example by Hirshleifer (1989), Shachar and Nalebuff (1999), Herrera et al. (2008), in part because they are well behaved with respect to a linear advantage $h(b_1) = h_0 + b_1$. The special case of a *quasi-linear* function $P(b_k - b_{-k})$ which is linear when it is not 0 or 1 has been studied extensively. Che and Gale (2000) give a relatively complete analysis when cost is linear and explicitly compute the equilibria, which all have finite support. Ashworth and Mesquita (2009) extend that analysis to the case where one contestant has linear cost and the other has a linear effort advantage.

A natural alternative to additive shocks in the case of effort are multiplicative shocks. Here $\nu_k = \epsilon_k b_k$ where ϵ_k is a positive random variable. In this case the probability of 1 winning the election is given by $\Pr\{\epsilon_1 b_1 > \epsilon_{-1} b_{-1}\}$ or $\Pr\{\epsilon_{-1}/\epsilon_1 < b_1/b_{-1}\}$. Provided that ϵ_{-1}/ϵ_1 and ϵ_1/ϵ_{-1} have the same distribution, which is certainly the case if ϵ_1 and ϵ_{-1} are iid this leads to $p(b_k, b_{-k}) = P(b_k/b_{-k})$, that is, a *ratio* model.

Like difference models ratio models have been widely used. For example Shachar and Nalebuff (1999) assume that $\epsilon_L = 1 - \epsilon_S$ where ϵ_S is normally distributed (although this does not respect the fact that ϵ should be non-negative). Coate and Conlin (2004b) assume that ϵ_k are independently drawn from beta distributions of which the uniform is a special case.[11]

[11] However, they assume that the parameters of the beta depend on the group size so their model does not map to a conflict resolution function.

11.11.1. Difference Versus Ratio Models. Most of the models in the literature are either of the ratio or difference form. Notice that ration models must be discontinuous at zero, which is not the case for difference models. Despite some controversy over which is more appropriate the two are not so different. For starters, in both Hirshleifer (1989) and Tullock (1967) if β is large both are similar to an all-pay auction. Since the equilibrium of the all-pay auction is unique 11.8.1 shows that the equilibrium strategies for both models must be similar in the sense that the difference between any two must converge weakly to zero.

The models are more similar than that, however. If in a ratio model we allow a small effort for free, so that $\varrho_k(b_k) = 0$ for $0 \leq b_k \leq b$ for some $b > 0$ then since the strategy of providing less than b units of effort is weakly dominated (strictly if p is strictly increasing) we may as well assume that the contestants make effort only $b_k \geq b$ and we can then change the units so that b in the original units corresponds to 0 in the new units, that is, $h(b_k) = b_k + b$ (see for example Amegashie (2006)), in which case we have for the ratio model

$$P\left(\frac{b_k + b}{b_{-k} + b}\right).$$

Having eliminated by a small perturbation the discontinuity at 0 we can now change units again with $h(b_k) = b\,(\exp(b_k) - 1)$ to convert the ratio model into

$$P\left(\frac{\exp(b_k)}{\exp(b_{-k})}\right),$$

an equivalent difference model.

11.11.2. The Tullock Function. Recall that with multiplicative shocks

$$p(b_k, b_{-k}) = \Pr\{\epsilon_{-1}/\epsilon_1 < b_1/b_{-1}\}.$$

We are especially interested in the Tullock function

$$p(b_k, b_{-k}) = \frac{b_k^\beta}{b_k^\beta + b_{-k}^\beta}$$

as this is convenient and widely used. Is there an underlying random turnout model that gives rise to this distribution? This question is partially answered by the following result concerning the logistic distribution:

THEOREM 11.11.1. *Suppose $x = \log \epsilon_{-1} - \log \epsilon_1$ has the logistic distribution with zero mean and scale parameter $1/\beta$ given by the density*

$$\beta \frac{e^{-\beta x}}{(1 + e^{-\beta x})^2}.$$

Then $p(b_k, b_{-k}) = P(b_k/b_{-k})$ is Tullock with parameter β.

PROOF. The logistic cdf is given by

$$\frac{1}{1 + e^{-\beta x}}$$

hence

$$\begin{aligned} p(b_1, b_{-1}) &= \Pr\{\log \epsilon_{-1} - \log \epsilon_1 < \log b_1 - \log b_{-1}\} \\ &= \frac{1}{1 + (b_{-1}/b_1)^\beta} = \frac{b_1^\beta}{b_1^\beta + b_{-1}^\beta} \end{aligned}$$

□

This is useful in itself because it tells us the probability distribution of the ratio of votes. Is there, however, a distribution for $\epsilon_1, \epsilon_{-1}$ that gives rise to this distribution for the ratio ϵ_1/ϵ_{-1}? There are in fact several: here are two. Suppose υ is uniform on $[0,1]$ and we set $\epsilon_1 = \upsilon^{1/\beta}$ and $\epsilon_{-1} = (1-\upsilon)^{1/\beta}$. The result follows from the fact that $\log\epsilon_1 - \log\epsilon_2$ is known to be a standard logistic. Here the two shocks are perfectly correlated. As it is also the case that if $\upsilon_1, \upsilon_{-1}$ are independent exponential random variable with mean 1 again $\log\epsilon_1 - \log\epsilon_2$ is know to be standard logistic. Hence we may take $\epsilon_k = \upsilon_k^{1/\beta}$ and again generate a Tullock contest success function.

11.12. Appendix: Continuity

11.12.1. Mathematical Preliminaries. We use the standard order on $\Re^M$ so that $x \geq x'$ means that this is true for each component. Suppose that X is a compact rectangle in $\Re^M$, that $f_n(x), f_0(x)$ are uniformly bounded non-decreasing real valued functions on X Denote by D the set of discontinuities of $f_0(x)$ and by $\overline{D}$ the closure of D.

LEMMA 11.12.1. *Suppose that $D_o \supset \overline{D}$ is an open subset of X. If for all $x \in X$ we have $f_n(x) \to f_0(x)$ then f_n converges uniformly to f on $X \backslash D_o$.*

PROOF. If $X\backslash D_o$ is empty this is true trivially. Otherwise as $X\backslash D_o$ is compact if the theorem fails there is a sequence $x_n \in X\backslash D_o$ with $x_n \to x \in X\backslash D_o$ and $f_n(x_n) \to z \neq f_0(x)$. There are two cases as $z < f_0(x)$ and $z > f_0(x)$. Denote the bottom corner of X as y_0 and the top corner as y_1. Notice that since D_o is open and contains the closure of D, then x has an open neighborhood in which f_0 is continuous.

If $z < f_0(x)$ and $x \neq y_0$ since f_0 is continuous near x there is a $y < x$ with $f_0(y) > z$ and an N such that for $n > N$ we have $x_n > y$. Since f_n is non-decreasing $f_n(x_n) \geq f_n(y)$. Hence $z \geq f_0(y)$ a contradiction. If $x = y_0$ then $f_n(y_0) \to f_0(y_0)$ while $f_n(x_n) \geq f_n(y_0)$. Taking limits on both sides we get $z \geq f_0(y_0)$ a contradiction

If $z > f_0(x)$ and $x \neq y_1$ we have $y > x$ such that $f_0(y) < z$ and an N such that for $n > N$ we have $x_n < y$. This gives $f_n(x_n) \leq f_n(y)$ implying $z \leq f_0(y)$ a contradiction. If $x = y_1$ we have $f_n(x_1) \to f_0(x_1)$ and $f_n(x_n) \leq f_n(x_1)$ and taking limits on both sides we get $z \leq f_0(x_1)$ a contradiction. □

We say that an open set D_o *encompasses* f_0 if there is a closed set $D_1 \subset D_o$ such that the interior of D_1 contains $\overline{D}$. Let $\overline{D}_o$ denote the closure of D_o.

THEOREM 11.12.2. *Suppose that the probability measures μ_n converge weakly to μ_0. If there is a sequence of sets D_a^m, D_g^m with $D_a^m \cup D_g^m$ encompassing f_0 such that $\limsup_m \limsup_n \sup_{x\in\overline{D}_a^m} |f_n(x) - f_0(x)| = 0$ and $\limsup_m \limsup_n \mu_n(\overline{D}_g^m) = 0$ then $\lim \int f_n d\mu_n = \int f_0 d\mu_0$.*

PROOF. By Urysohn's Lemma there are continuous functions $0 \leq g^m(x) \leq 1$ equal to 1 for $x \in X\backslash D_0^m$ and equal to zero for $x \in D_1^m$. Setting $D_o^m = D_g^m \cup D_a^m$

$$|\int f_n d\mu_n - \int f_0 d\mu_0| \leq |\int g^m f_n d\mu_n - \int g^m f_0 d\mu_0| +$$

$$+|\int(1-g^m)f_n d\mu_n - \int(1-g^m)f_0 d\mu_0| \leq$$

$$\leq |\int g^m f_n d\mu_n - \int g^m f_0 d\mu_0| + |\int_{\overline{D}_0^m} f_n d\mu_n - \int f_0 d\mu_0|.$$

If ϕ_n, ϕ_0 are real numbers and m_n, m_0 are non-negative real numbers we have the inequality $|\phi_n m_n - \phi_0 m_0| \leq |\phi_n - \phi_0|(m_n + m_0)$ so

$$|\int f_n d\mu_n - \int f_0 d\mu_0| \leq |\int g^m f_n d\mu_n - \int g^m f_0 d\mu_0| + \int_{\overline{D}_o^m} |f_n - f_0| d(\mu_n + \mu_0).$$

First we show that $\int_{\overline{D}_o^m} |f_n - f_0| d(\mu_n + \mu_0) \to 0$. Let $\overline{f} = \sup |f_k(x)|$. we have

$$\int_{\overline{D}_o^m} |f_n - f_0| d(\mu_n + \mu_0) \leq \int_{\overline{D}_a^m} |f_n - f_0| d(\mu_n + \mu_0) + \int_{\overline{D}_g^m} |f_n - f_0| d(\mu_n + \mu_0)$$

$$\leq \sup_{x \in \overline{D}_a^m} |f_n(x) - f_0(x)| + \overline{f}\left(\mu_n(\overline{D}_g^m) + \mu_0(\overline{D}_g^m)\right).$$

The first term converges to 0 by hypothesis. For the second, as $\overline{D}_g^m$ is closed and μ_n converges weakly to μ_0 we have $\mu_0(\overline{D}_g^m) \leq \limsup \mu_n(\overline{D}_a^m)$ so

$$\limsup_n \overline{f}\left(\mu_n(\overline{D}_g^m) + \mu_0(\overline{D}_g^m)\right) \leq 2\overline{f} \limsup_n \mu_n(\overline{D}_g^m)$$

giving the first result. Second, write

$$|\int g^m f_n d\mu_n - \int g^m f_0 d\mu_0| \leq |\int g^m |f_n - f_0| d\mu_n + |\int g^m f_0 d\mu_0 - \int g^m f_0 d\mu_n|.$$

Since $g^m f_0$ is continuous by construction we have $\lim_n |\int g^m f_0 d\mu_0 - \int g^m f_0 d\mu_n| = 0$ by weak convergence of μ_n to μ_0.

Finally, we show that $\lim_n |\int g^m |f_n - f_0| d\mu_n = 0$. Denote by D_1^{mo} the interior of D_1^m and $X_1^m = X \backslash D_1^{mo}$. By Lemma 11.12.1 $|f_n(x) - f_0(x)| \leq \epsilon_n^m$ for $x \in X_1^m$ where $\lim_n \epsilon_n^m = 0$. As $g^m(x) = 0$ for $x \in D_1^m \supset D_1^{mo}$ we have $g^m|f_n - f_0| \leq \epsilon_n^m$ so that $\int g^m |f_n - f_0| d\mu_n \leq \epsilon_n$. □

Recall that $\overline{D}$ denote the closure of D.

THEOREM 11.12.3. *Suppose that* X *is a compact rectangle in* $\Re^M$*, that* $f_n(x), f_0(x)$ *are uniformly bounded non-decreasing real valued functions on* X*, that* $f_n(x) \to f_0(x)$ *and that the probability measures* μ_n *converge weakly to* μ_0*. If* $\mu_0(\overline{D}) = 0$ *then* $\lim \int f_n d\mu_n = \int f_0 d\mu_0$*.*

PROOF. Take the sets $D_o^m = D_g^m$ to be the open $\epsilon_m \to 0$ neighborhoods of $\overline{D}$ and take $D_a^o = \emptyset$. We may take D_1^m sets to be the closed $\epsilon/2$ neighborhoods of $\overline{D}$: this clearly contains $\overline{D}$ in its interior and is contained in D_o^m. Take D_2^m to be the open $2\epsilon_m$ neighborhoods of D: as these contain $\overline{D}_o^m$ is suffices to show that $\limsup_m \limsup_n \mu_n(D_2^m) = 0$. Since D_2^m is open and μ_n converges weakly to μ we have $\limsup_n \mu_n(D_2^m) \leq \mu_0(D_2^m)$, so we need only prove $\limsup_m \mu_0(D_2^m) = 0$. Since $\cap_m D_2^m = \overline{D}$ we have $\lim_m \mu_0(D_2^m) = \mu_0(\overline{D}) = 0$. □

11.12.2. Upper Hemi-Continuity of the Equilibrium Correspondence. We now consider a *convergence scenario.* Here $p_n(b_1, b_{-1}) \to p_0(b_1, b_{-1})$, $\varrho_{jn}(b_j) \to \varrho_{j0}(b_j)$ is a sequence of contests on W. We take G_{1n}, G_{-1n} to be equilibria for n converging weakly to G_{10}, G_{-10} with μ_{jk} the corresponding measures. We say that the convergence scenario is *upper hemi-continuous* if $p_n(G_{jn}, G_{-jn}) \to p_0(G_{j0}, G_{-j0})$, $\varrho_{jn}(G_{jn}) \to \varrho_{j0}(G_{j0})$ for both j and G_{10}, G_{-10} is an equilibrium for $p_0(b_1, b_{-1}), \varrho_{j0}(b_j)$.

THEOREM 11.12.4. *If $p_n(G_{jn}, G_{-jN}) \to p_0(G_{j0}, G_{-j0})$ for both j then the convergence scenario is upper hemi-continuous.*

PROOF. By Theorem 11.12.3 $\varrho_{jn}(G_{jn}) \to \varrho_{j0}(G_{j0})$ on the relevant domain $0 \le b_j \le W$. This shows that $u_{jn}(G_{jn}, G_{-jn}) \to u_{j0}(G_{j0}, G_{-j0})$. Next consider j deviating to $b_j \in [0, W]$. Suppose first that b_j is an atom of G_{-j0}. Then this is not a best response. Suppose second that b_j is not an atom of G_{-j0}. Hence the function of b given by $p_0(b_j, b)$ has measure zero with respect to G_{-j0}. If follows from Theorem 11.12.3 that $p_n(b_j, G_{-jn}) \to p_0(b_j, G_{-j0})$, so also $u_{jn}(b_j, G_{-jn}) \to u_{j0}(b_j, G_{-j0})$. If $b_j$was a profitable deviation, that is, $u_{j0}(b_j, G_{-j0}) > u_{j0}(G_{j0}, G_{-j0})$, it follows by the standard argument that for sufficiently large n we would have $u_{jn}(b_j, G_{-jn}) > u_{jn}(G_{jn}, G_{-jn})$ contradicting the optimality of G_{jn}. □

In what follows all sequences are of strictly positive numbers.

LEMMA 11.12.5. *If $\gamma^m \to 0$ then there are sequences $J^n, H^m \to 0$ such that on $[0, W + 2\max\gamma^m]$ we have $\max_{b \in [0,W]} \varrho_{jn}(b + 2\gamma^m) - \varrho_{jn}(b) \le J^n + H^m$.*

PROOF. By Lemma 11.12.1 we have ϱ_{jn} converging uniformly to ϱ_{j0} so that

$$\max_{b \in [0,W]} \varrho_{jn}(b + 2\gamma^m) - \varrho_{jn}(b) \le \max_{b \in [0,W]} \varrho_{j0}(b + 2\gamma^m) - \varrho_{j0}(b) + J_{jn}$$

Since ϱ_{j0} is uniformly continuous on compact intervals $\max_{b \in [0,W]} \varrho_{j0}(b + 2\gamma^m) - \varrho_{j0}(b) \le H_{jm}$. Then take $J^n = \max J_{jn}, H^m = \max H_{jm}$. □

LEMMA 11.12.6. *Fix sequences $\gamma^m, \theta^m \to 0$. Then there exists a sequence $u^n \to 0$ and $\gamma^m \ge \omega^m$ such that for $0 \le b_{-j} - b \le \omega^m$:*

(i) If $p(b + \gamma^m) - 1/2 < \theta^m$ then $\sup_{0 \le b_k - b \le \omega^m} |p_n(b_j, b_{-j}) - p_0(b_j, b_{-j})| \le 2\theta^m + u^n$.

(ii) If $p(b + \gamma^m) - 1/2 \ge \theta^m$ then $p_n(b + \gamma^m + \omega^m, b_{-j}) - 1/2 \ge \theta^m/2 - u^n$.

PROOF. We may apply Theorem 11.12.2 to the functions $p_n(b_j, -x_{-j}), p_0(b_j, -x_{-j})$ on the rectangle $[0, W] \times [-W, 0]$ with $D_o = \{(b_j, x_{-j}) \,||b_j + x_j| < \gamma^m\}$ to conclude that $p_n(b_j, -x_{-j})$ converges uniformly to $p_0(b_j, -x_{-j})$ there. Hence there exists a constant u^m such that for $b_j - b_{-j} \ge \gamma^m$ we have $|p_n(b_j, b_{j-1}) - p_0(b_j, b_{j-1})| \le u^n$.

Fix b. For (i) Take $\omega^m = \gamma^m$. Take $0 \le b_k - b \le \omega^m$. Observe that

$$p_0(b_j, b_{-j}) \le p_0(b + \omega^m, b) < 1/2 + \theta^m.$$

Since $b + \omega^m - b \ge \gamma^m$ we also have $|p_n(b + \omega^m, b) - p_0(b + \omega^m, b)| \le u^n$ this implies

$$p_n(b_j, b_{-j}) \le 1/2 + \theta^m + u^n.$$

Reversing the role of j and $-j$ we see that

$$|p_0(b_j, b_{-j}) - 1/2| < \theta^m,\ |p_n(b_j, b_{-j}) - 1/2| < \theta^m + u^n.$$

Hence $|p_n(b_j, b_{-j}) - p_0(b_j, b_{-j})| < 2\theta^m + u^n$.

For (ii), observe that $p_0(b_j, b_{-j})$ is uniformly continuous on $b_j - b_{-j} \geq \gamma^m$. Hence we may find a $\omega^m > 0$ which without loss of generality we may take to be smaller than γ^m such that for $|b_{-j} - b| \leq \omega^m$ we have $|p_0(b_j, b_{-j}) - p_0(b_j, b)| < \theta^m/2$. Since $p_n(b + \gamma^m + \omega^m, b_{-j})$ is non-increasing in b_{-j} we put this all together:

$$p_n(b+\gamma^m+\omega^m, b_{-j}) \geq p_n(b+\gamma^m+\omega^m, b+\omega^m) \geq p_0(b+\gamma^m+\omega^m, b+\omega^m) - u^n$$

$$\geq p_0(b+\gamma^m+\omega^m, b) - \theta^m/2 - u^n \geq p(b+\gamma^m) - \theta^m/2 - u^n \geq 1/2 + \theta^m/2 - u^n.$$

□

LEMMA 11.12.7. *For any $\gamma^m \to 0$ there are sequences $J^n, H^m \to 0$ such that for any θ^m and $\omega^m \leq \gamma^m$ and any b with $p_n(b+\gamma^m+\omega^m, b_{-j}) - 1/2 \geq \theta^m/2 - u^n > 0$ for all $0 \leq b_{-j} - b \leq \omega^m$ we have*

$$\min_j \mu_{jn}([b, b+\omega^m]) \leq \frac{J_n + H_m}{\theta^m/2 - u^n}$$

PROOF. Given $\gamma^m \to 0$ choose the sequences J^n, H^m by Lemma 11.12.5.

Define $m_j \equiv \mu_{jn}([b, b+\omega^m])$. If for one j we have $m_j = 0$ then certainly the inequality holds. Otherwise, consider that if each j plays μ_{jn}/m_j in $[b, b+\omega^m]$ then one of them must have probability no greater than $1/2$ of winning. Say this is j. Consider the strategy for j of switching from μ_{jn} to $\hat{\mu}_{jn}$ by not providing effort in $[b, b+\omega^m]$ and instead providing effort with probability m_j at $b+\gamma^m+\omega^m$. This results in a utility gain of at least

$$m_{-j}\left(\theta^m/2 - u^n\right) - \left(\varrho_{jn}(b+\gamma^m+\omega^m) - \varrho_{jn}(b)\right)$$

$$\geq m_{-j}\left(\theta^m/2 - u^n\right) - \left(\varrho_{jn}(b+2\gamma^m) - \varrho_{jn}(b)\right) \geq m_{-j}\left(\theta^m/2 - u^n\right) - \left(J^n + H^m\right).$$

As the utility gain cannot be positive, this implies $0 \geq m_{-j}\left(\theta^m/2 - u^n\right) - \left(J^n + H^m\right)$ giving the desired inequality. □

THEOREM 11.12.8. *Convergence scenarios are upper hemi-continuous.*

PROOF. By Theorem 11.12.4 it suffices to show $p_n(G_{jn}, G_{-jn}) \to p_0(G_{j0}, G_{-j0})$.

Observe that $p_n(b_j, b_{-j}), p_0(b_j, b_{-j})$ are non-decreasing in the first argument and non-increasing in the second so that the functions on the rectangle $[0, W] \times [-W, 0]$, given by $f_k(x) \equiv p_k(x_j, -x_{-j})$, are uniformly bounded. Define $\mu_n = \mu_{1n} \times \mu_{-1n}$ and $\mu_0 = \mu_{10} \times \mu_{-10}$. From Fubini's Theorem μ_n converges weakly to μ_0. so Theorem 11.12.2 applies if we can show how to construct the sets D_a^m, D_g^m.

Fix a sequence $\gamma^m \to 0$. Choose sequences J^n, H^m by Lemma 11.12.7 and choose $\theta^m \to 0$ so that $H^m/\theta^m \to 0$. Then choose $u^n \to 0$ and $\omega^m \leq \gamma^m$ by Lemma 11.12.6.

We cover the diagonal with open squares of width ω^m. Specifically, for $\ell = 1, 2, \ldots, L$ we take the lower corners κ_ℓ of these squares to be $0, 2\omega^m/3, 4\omega^m/3, \ldots$ until the final square overlaps the top corner at (W, W). There are two types of squares: a-squares where $p(\kappa_\ell + \gamma^m) - 1/2 < \gamma^m$ and g-squares where $p(\kappa_\ell + \gamma^m) - 1/2 \geq \gamma^m$.

We take D_a^m to be the union of the a-squares and D_g^m to be the union of the g-squares. Then for each square ℓ we may take a closed square with the same corner but 3/4rths the width and define D_1 to be the union of these squares. Then $D_o^m = D_a^m \cup D_b^m \supset D_1 \supset \overline{D}$ so that indeed D_o^m encompasses p_0.

Since D_a^m is the union of a-squares, by Lemma 11.12.6 (i) we have $\sup_{x\in\overline{D}_a^m}|f_n(x)-f_0(x)|\leq 2\theta^m+u^n$, so indeed $\limsup_m \limsup_n \sup_{x\in\overline{D}_a^m}|f_n(x)-f_0(x)|=0$ as required by Theorem 11.12.2.

For a g-square ℓ we have $0\leq b_{-j}-b\leq\omega^m$ so by Lemma 11.12.6 $p_n(b+\gamma^m+\omega^m,b_{-j})-1/2\geq\theta^m/2-u^n$. Then by Lemma 11.12.7

$$\min_j \mu_{jn}([\kappa_\ell,\kappa_\ell+\omega^m])\leq\frac{J_n+H_m}{\theta^m/2-u^n}.$$

We now add up over the g-squares four times, once for the odd numbered ones and once for the even numbered ones. This assures that each sum is over disjoint squares. In each case we first add those for which $j=1$ has the lowest value of $\mu_{jn}([\kappa_\ell,\kappa_\ell+\omega^m])$ and once for $j=-1$. In each set of indices Λ we get a sum

$$\sum_{\ell\in\Lambda}\mu_{jn}([\kappa_\ell,\kappa_\ell+\omega^m])\mu_{-jn}([\kappa_\ell,\kappa_\ell+\omega^m])\leq$$

$$\frac{J_n+H_m}{\theta^m/2-u^n}\sum_{\ell\in\Lambda}\mu_{-jn}([\kappa_\ell,\kappa_\ell+\omega^m])\leq\frac{J_n+H_m}{\theta^m/2-u^n}.$$

This gives a bound

$$\mu_n(\overline{D}_g^m)\leq\frac{J_n+H_m}{\theta^m/2-u^n}.$$

We then have

$$\limsup_n \mu_n(\overline{D}_g^m)\leq\frac{H_m}{\theta^m/2}$$

and since we constructed the sequences so that $H^m/\theta^m\to 0$ the result now follows from Theorem 11.12.2. $\square$

11.12.3. Smoothing Conflict Resolution Functions.

THEOREM 11.12.9. *If p,ϱ_j is a contest on W then there is a sequence of well-behaved contests p_n,ϱ_{jn} on W with $p_n(b_j,b_{-j})\to p(b_j,b_{-j}),\varrho_{jn}(b_j)\to\varrho_j(b_j)$ for every $(b_1,b_{-1})\in[0,W]\times[0,W]$.*

To prove this theorem we first state and prove

LEMMA 11.12.10. *Suppose that $p_n(b_j,b_{-j})\to p_0(b_j,b_{-j})$ and $p_{mn}(b_j,b_{-j})\to_m p_n(b_j,b_{-j})$. Then there is $M(n)$ such that $p_{M(n)n}(b_j,b_{-j})\to p_0(b_j,b_{-j})$.*

PROOF. Define $d(p,q)=\inf\{\gamma|\sup_{|b_j-b_{-j}|\geq\gamma}|p(b_j,b_{-j})-q(b_j,b_{-j})|\leq\gamma\}$. Then $d(p,q)=0$ if and only if $p=q$, $d(p,q)=d(q,p)$ and $d(p,q)+d(q,r)\leq 2\max\{d(p,q),d(q,r)\}$. Moreover, $d(p_n,p_0)\to 0$ if and only if $p_n(b_j,b_{-j})\to p_0(b_j,b_{-j})$. Let $\epsilon_n\to 0$ and take $M(n)$ such that for $m\geq M(n)$ we have $d(p_{mn},p_n)<\epsilon_n$. Then $d(p_{M(n)n},p_0)\leq 2\max\{\epsilon_n,d(p_n,p_0)\}\to 0$. $\square$

We now prove Theorem 11.12.9.

PROOF. By Lemma 11.12.10 we can do the perturbations sequentially.

Step 1: Perturb p to get it strictly increasing with strictly positive infimum: take $p_n(b_j,b_{-j})=(1-\lambda_n)p(b_j,b_{-j})+\lambda_n\Phi(b_j-b_{-j})$ where Φ is the standard normal cdf.

Step 2: Given p strictly increasing and positive perturb it to get it strictly increasing, positive and C^2. Let $g_n(x_j|b_j) = (1/W)h_n(x_j/W|b_j)$ where $h_n(\bullet|b_j)$ is the Dirichlet distribution with parameter vector

$$8n^3\Big[(1-\frac{1/n}{2\sqrt{2}})(b_j/W)+\frac{1/n}{2\sqrt{2}}\frac{1}{2}\Big], 8n^3\Big[(1-\frac{1/n}{2\sqrt{2}})(1-b_j/W)+\frac{1/n}{2\sqrt{2}}\frac{1}{2}\Big].$$

This is C^∞ in b_j and $g_n(0|b_j) = g_n(W|b_j) = 0$ and taking

$$p_n(b_j, b_{-j}) \equiv \int_0^\infty p(x_j, x_{-j})g_n(x_j|b_j)g_n(x_{-j}|b_{-j})dx_j dx_{-j}$$

this is certainly strictly positive and C^2. To see that it is strictly increasing observe that increasing b_j increases $g_n(x_j|b_j)$ in first order stochastic dominance. Finally, letting $\tilde{x}_j$ be the random variable with density $g_n$following the Web Appendix of Dutta et al. (2018) we show that Chebyshev's inequality implies $\Pr(|\tilde{x}_j - b_j| > 1/n) \leq 1/n$ so that we have pointwise convergence at every continuity point of p. Pointwise convergence on the diagonal is by definition.

To use Chebyshev's inequality, observe that $\tilde{x}_j$ has mean

$$\overline{x}_j = (1-\frac{1/n}{2\sqrt{2}})(b_j/W)+\frac{1/n}{2\sqrt{2}}\frac{1}{2}\Big].$$

Since the covariances of the Dirichlet are negative, $E|\tilde{x}_j - \overline{x}_j|^2$ is bounded by the sum of the variances and we may apply Chebyshev's inequality to find

$$\Pr[|\tilde{x}_j - \overline{x}_j| > 1/(2n)] \leq E|\tilde{x}_j - \overline{x}_j|^2/(1/(2n))^2$$

By the standard Dirichlet variance formula we have

$$\frac{E|\tilde{x}_j - \overline{x}_j|}{(1/(2n))^2} \leq \frac{1}{(1/(2n))^2}\frac{\left(8n^3\right)^2}{\left(8n^3\right)^2\left(8n^3+1\right)} \leq \frac{1}{8n^3(1/(2n))^2} = \frac{1}{2n}$$

We also have $|\bar{x}_j - b_j/W| \leq \frac{1/n}{2\sqrt{2}}\frac{1}{2} \leq 1/(2n)$; then $|\tilde{x}_j - b_j/W| > (1/n)$ implies $|\tilde{x}_j - \overline{x}_j| > \epsilon/2$; hence $\Pr(|\tilde{x}_j - b_j/W| > 1/n) \leq Pr[\Pr(|\tilde{x}_j - \overline{x}_j| > 1/(2n))] \leq 1/(2n) \leq 1/n$.

Step 3: Given p strictly increasing, positive and C^2 perturb it to get it strictly increasing, positive on $[0, W] \times [0, W]$ and real analytic in an open neighborhood. By Whitney (1934) Theorem 1 we can extend p to be C^1on all of R^2. Take an open neighborhood $\mathcal{W}$ of $[0, W] \times [0, W]$ so that p is strictly positive there. By Whitney (1934) Lemma 5 for each $\epsilon > 0$ we can find a real analytic function $q(b_j, b_{-j})$ with $|q - p| < \epsilon$ and $|Dq - Dp| < \epsilon$ on the closure of $\mathcal{W}$. Then define $Q(b_j, b_{-j}) = q(b_j, b_{-j})/(q(b_j, b_{-j}) + q(b_{-j}, b_j))$.

Remark: The case of ϱ_j is similar but easier. In the final step the real analytic function $q_j(b_j)$ is not necessarily zero at zero so we define $Q_j(b_j) = q_j(b_j) - q_j(0)$. $\square$

11.12.4. Continuity For Ties.

LEMMA 11.12.11. *Suppose that either G_{-j} does not have an atom at b or p is continuous at (b, b). Then $p(b_j, G_{-j})$ as a function of b_j is right continuous at $b_j = b$.*

PROOF. Let $b_j^n \downarrow b$ and write

$$p(b_j^n, G_{-j}) = \int p(b_j^n, b_{-j})dG_{-j}(b_{-j})$$

$$= \int_{|b_{-j}-b|>\epsilon} p(b_j^n, b_{-j}) dG_{-j}(b_{-j}) + \int_{|b_{-j}-b|\leq\epsilon} p(b_j^n, b_{-j}) dG_{-j}(b_{-j}).$$

For the first term from Theorem [monotone-uniform]

$$\int_{|b_{-j}-b|>\epsilon} p(b_j^n, b_{-j}) dG_{-j}(b_{-j}) \to \int_{|b_{-j}-b|>\epsilon} p(b, b_{-j}) dG_{-j}(b_{-j})$$

$$\leq \int_{b_{-j}\neq b} p(b, b_{-j}) dG_{-j}(b_{-j}).$$

Hence there is a sequence $\epsilon^n \to 0$ such that

$$\limsup \int_{|b_{-j}-b|>\epsilon^n} p(b_j^n, b_{-j}) dG_{-j}(b_{-j}) \leq \int_{b_{-j}\neq b} p(b, b_{-j}) dG_{-j}(b_{-j}).$$

If G_{-j} does not have an atom at b

$$\int_{|b_{-j}-b|\leq\epsilon^n} p(b_j^n, b_{-j}) dG_{-j}(b_{-j}) \leq \int_{|b_{-j}-b|\leq\epsilon^n} dG_{-j}(b_{-j}) \to 0.$$

Hence $\lim_{n\to\infty} p(b_j^n, G_{-j}) \leq p(b, G_{-j})$.

If p is continuous at (b, b) and letting μ_{-j} be the measure corresponding to G_{-j}

$$\int_{|b_{-j}-b|\leq\epsilon^n} p(b_j^n, b_{-j}) dG_{-j}(b_{-j}) \to p(b, b)\mu_{-j}(b).$$

Hence $\lim_{n\to\infty} p(b_j^n, G_{-j}) \leq \int_{b_{-j}\neq b>\epsilon} p(b, b_{-j}) dG_{-j}(b_{-j}) + p(b, b)\mu_{-j}(b) = p(b, G_{-j})$.

Since by monotonicity $p(b_j^n, G_{-j}) \geq p(b, G_{-j})$ right continuity follows from $\lim_{n\to\infty} p(b_j^n, G_{-j}) \leq p(b, G_{-j})$. □

11.12.5. Resource Limits. A *resource constrained contest* on W is a contest success function $p(b_j, b_{-j})$ together with a pair of cost functions $\varrho_j(b_j)$ that satisfy the definition of being a contest except that p is required to be continuous and we allow the possibility that ϱ_j instead of being continuous on the entire support is continuous on $[0, \overline{b}_j]$ where $\overline{b}_j > 0$, $\varrho_j(\overline{b}_j) = \overline{\varrho}_j < 1$, and for $b_j > \overline{b}_j$ we have $\varrho_j(b_j) = \varrho_{Max} > 1$. Our goal is to prove:

Theorem 11.12.12. *Suppose $p_n(b_1, b_{-1}) \to p_0(b_1, b_{-1})$, $\varrho_{jn}(b_j) \to \varrho_{j0}(b_j)$ for $b_j \neq \overline{b}_{j0}$ are a sequence of resource constrained contests in W, that G_{1n}, G_{-1n} are equilibria for n converging weakly to G_{10}, G_{-10}. Then $p_n(G_{jn}, G_{-jn}) \to p_0(G_{j0}, G_{-j0})$, $\varrho_{jn}(G_{jn}) \to \varrho_{j0}(G_{j0})$ for both j and G_{10}, G_{-10} is an equilibrium for $p_0(b_1, b_{-1})$, $\varrho_{j0}(b_j)$.*

Proof. If ϱ_{j0} is continuous then $\varrho_{jn}(b_j) \to \varrho_{j0}(b_j)$ for all b_j there is nothing new to be proven. We take then the discontinuous case. There are two new things that must be shown. First, we must show that if a deviation to $\overline{b}_{j0}$ against G_{j0} is profitable then, because we do not have pointwise convergence at $\overline{b}_{j0}$, there is another deviation that is also profitable. Second, we must show that $\varrho_{jn}(G_{jn}) \to \varrho_{j0}(G_{j0})$.

The first is simple: if we take a sequence $b_{jm} \to \overline{b}_{j0}$ strictly from below, the continuity of p_0, ϱ_{j_0} imply that $u_{j0}(b_{jm}, G_{-j}) \to u_{j0}(\overline{b}_{j0}, G_{-j})$ so that for large enough m the deviation $b_{jm} \neq b_{j0}$ is also profitable.

To prove the second we first choose $0 < \epsilon < (\varrho_{Max} - 1)/2$. We observe that for each n (including $n = 0$) the fact that ϱ_{jn} is weakly decreasing and left continuous means that $\{b_j | \varrho_{jn}(b_j) \leq \overline{\varrho}_{j0} + \epsilon\} = [0, b_{jn}(\epsilon)]$ and $\{b_j | \varrho_{jn}(b_j) > \overline{\varrho}_{j0} + \epsilon\} =$

$(b_{jn}(\epsilon), W]$ where it is apparent that $b_{j0}(\epsilon) = \overline{b}_{j0}$. Moreover, we can show that $\lim_n b_{jn}(\epsilon) = \overline{b}_{j0}$. To see that for any $\gamma > \overline{b}_{j0}$ we have $\lim_n \varrho_{jn}(\gamma) = \varrho_{Max}$ implying $\limsup b_{jn}(\epsilon) \leq \gamma$. For any $\gamma < \overline{b}_{j0}$ we have $\lim_n \varrho_{jn}(\gamma) \leq= \varrho_{j0}(\gamma) \leq \overline{\varrho}_{j0}$ implying $\liminf_n b_{jn}(\epsilon) \geq \gamma$.

Second, since p_0 is continuous, pointwise convergence of p_n to p_0 implies uniform convergence and since $\mathcal{W}$ is compact, p_0 is uniformly continuous. It follows that $\Delta(\epsilon) = \inf\{0 \leq b_j^1 - b_j^2 | p_n(b_j^2, b_{-j}) - p_n(b_j^1, b_{-j}) \leq \epsilon\}$ is positive.

Third, we show that for sufficiently large n we have

$$\mu_{jn}((b_{jn}(\epsilon), \overline{b}_{j0} + \Delta(\epsilon/2)/2]) = 0.$$

Suppose that $b_j \in (b_{jn}(\epsilon), \overline{b}_{j0} + \Delta(\epsilon/2)/2])$. Then $\varrho_{jn}(b_j) \geq \overline{\varrho}_{j0} + \epsilon$ while $\varrho_{jn}(\overline{b}_{j0} - \Delta(\epsilon/2)/2]) \leq \varrho_{j0}(\overline{b}_{j0} - \Delta(\epsilon/2)/2]) + \eta_n$ where $\eta_n \to 0$. Since $b_j - (\overline{b}_{j0} - \Delta(\epsilon/2)/2]) \leq \Delta(\epsilon/2)$ it follows that $p_n(b_j, G_{-j}) - p_n(\overline{b}_{j0} - \Delta(\epsilon/2)/2], G_{-j}) \leq \epsilon/2$, while $\varrho_{jn}(b_j) - \varrho_{jn}(\overline{b}_{j0} - \Delta(\epsilon/2)/2]) \geq \epsilon - \eta_n$. Hence for $\eta_n < \epsilon/2$ it is not optimal to play b_j.

Fourth, we show that for sufficiently large n we have $\mu_{jn}((b_{jn}(\epsilon), W]) = 0$. To do so we need only show that for sufficiently large n we have $\mu_{jn}((\overline{b}_{j0} + \Delta(\epsilon/2)/2, W]) = 0$. Since $\varrho_{jn}(\overline{b}_{j0} + \Delta(\epsilon/2)/2) \to \varrho_{Max}$ for all sufficiently large n we have $\varrho_{jn}(\overline{b}_{j0} + \Delta(\epsilon/2)/2) > 1$ and since ϱ_{jn} is non-decreasing $\varrho_{jn}(b_j) > 1$ for all $b_j \geq \overline{b}_{j0} + \Delta(\epsilon/2)/2$. Of course it cannot be optimal to play such an b_j.

Fifth we show that $\mu_{j0}((\overline{b}_{j0}, W]) = 0$. This follows from the fact that it is the countable union of the sets

$$(\overline{b}_{j0} + |b_{jn}(\epsilon) - \overline{b}_{j0}|, W] \subset (b_{jn}(\epsilon), W].$$

Sixth, we construct approximating functions $\tilde{\varrho}_{jn}$. Since $\varrho_{j0}$is continuous on $[0, \overline{b}_{j0}]$ we may choose $\gamma < \overline{b}_{j0}$ so that $\varrho_{j0}(\overline{b}_{j0}) - \varrho_{j0}(\gamma) < \epsilon$. Then for $b_j \leq \gamma$ we take $\tilde{\varrho}_{jn}(b_j) = \varrho_{jn}(b_j)$ and for $b_j > \gamma$ we take $\tilde{\varrho}_{jn}(b_j) = \varrho_{jn}(\gamma)$. Certainly then $\tilde{\varrho}_{jn}$ is non-decreasing and converges pointwise to the non-decreasing function $\tilde{\varrho}_{j0}$. It follows that the convergence is uniform, hence $\tilde{\varrho}_{jn}(G_{jn}) \to \tilde{\varrho}_{0n}(G_{j0})$.

Seventh, we bound

$$|\tilde{\varrho}_{jn}(G_{jn}) - \varrho_{jn}(G_{jn})|$$

$$\begin{aligned}
&\leq \int_{[0,\gamma]} |\tilde{\varrho}_{jn}(b_{jn}) - \varrho_{jn}(b_{jn})|\, dG_{jn} + \int_{(\gamma, b_{jn}(\epsilon)]} |\tilde{\varrho}_{jn}(b_{jn}) - \varrho_{jn}(b_{jn})|\, dG_{jn} \\
&\quad + \left| \int_{(b_{jn}(\epsilon), W]} (\tilde{\varrho}_{jn}(b_{jn}) - \varrho_{jn}(b_{jn}))\, dG_{jn} \right| \\
&= \int_{(\gamma, b_{jn}(\epsilon)]} |\tilde{\varrho}_{jn}(b_{jn}) - \varrho_{jn}(b_{jn})|\, dG_{jn} \\
&\leq \sup_{(\gamma, b_{jn}(\epsilon)]} |\tilde{\varrho}_{jn}(b_{jn}) - \varrho_{jn}(b_{jn})| \\
&= \varrho_{jn}(b_{jn}(\epsilon)) - \varrho_{jn}(\gamma) \\
&\leq \left|\varrho_{jn}(b_{jn}(\epsilon)) - \varrho_{j0}(\overline{b}_{j0})\right| + \left|\varrho_{j0}(\overline{b}_{j0}) - \varrho_{j0}(\gamma)\right| + |\varrho_{j0}(\gamma) - \varrho_{jn}(\gamma)| \\
&\leq 2\epsilon + \eta_n
\end{aligned}$$

where $\eta_n \to 0$.

Finally, we put this together to see that for all $0 < \epsilon < 1/2$ and sufficiently large n we have

$$|\varrho_{jn}(G_{jn}) - \varrho_{j0}(G_{j0})| \leq |\tilde{\varrho}_{jn}(G_{jn}) - \tilde{\varrho}_{j0}(G_{j0})| + 4\epsilon + 2\eta_n.$$

It follows that $\limsup |\varrho_{jn}(G_{jn}) - \varrho_{j0}(G_{j0})| \leq 4\epsilon$. This proves the result. $\square$

11.12.6. Finite Support.

THEOREM 11.12.13. *Suppose that $\varrho_1(b_1) = 0$ for $0 \leq b_1 \leq \underline{w}_1$ and if $\underline{w}_1 > 0$ we require that $p(b_j, b_{-j})$ is strictly increasing in the first argument (so in particular in any equilibrium $\mu_1([0, \underline{w}_1)) = 0$). Suppose as well that $\varrho_j(W) > 1$. If $p(b_j, b_{-j}), \varrho_j(b_j)$ have real analytic extensions to an open neighborhood of $[\underline{w}_1, W] \times [0, W]$ then every equilibrium has finite support.*

PROOF. Take $\underline{w}_{-1} = 0$ and consider

$$U_j(b_j) \equiv \int_{\underline{w}_j}^{W} p(b_j, b_{-j}) dG_{-j}(b_{-j}) - \varrho_j(b_j).$$

We first show that this is real analytic in an open neighborhood of $[\underline{w}_j, W]$. For ϱ_j this is true by assumption so we show it for the integral

$$P_j(b_j) \equiv \int_{\underline{w}_j}^{W} p(b_j, b_{-j}) dG_{-j}(b_{-j}).$$

In Ewerhart (2015) the extensibility properties of $p(b_j, b_{-j})$ were known. Here we must establish them. Let $\mathcal{W}$ be the open neighborhood of $[\underline{w}_1, W] \times [\underline{w}_{-1}, W]$ in which p is real analytic. Then for each point $b \in \mathcal{W}$ the function p has an infinite power series representation with a positive radius of convergence $r_1, r_{-1}$for b_1, b_{-1} respectively. Hence the extension of p to a function of two complex variables has the same radius of convergence there. Take an open square around b_j in the complex plane small enough to be entirely contained in the circle of radius $\min\{r_1, r_{-1}\}$ and lying inside of $\mathcal{W}$. The product of these squares is an open cover of the compact set $[\underline{w}_1, W] \times [\underline{w}_{-1}, W]$, hence has a finite sub-cover. Choose the smallest square from this finite set, say with length $2h$. Hence $p(b_j, b_{-j})$ is complex analytic in the domain $((\underline{w}_1 - h, W + h) \times (-h, +h)) \times \big((\underline{w}_{-1} - h, W + h) \times (-h, +h)\big)$.

The remainder of the proof follows Ewerhart (2015) in showing that we may extend $P_j(b_j)$ to a complex analytic function in the domain $(\underline{w}_j - h, W + h) \times (-h, +h)$. As this is a convex domain, take a triangular path Δ in this domain and integrate

$$\oint_{\Delta} P_j(b_j) = \oint_{\Delta} \int_{\underline{w}_j}^{W} p(b_j, b_{-j}) dG_{-j}(b_{-j}).$$

Everything in sight is bounded so we may apply Fubini's Theorem and interchange the order of integration to find

$$\oint_{\Delta} P_j(b_j) = \int_{\underline{w}_j}^{W} \left(\oint_{\Delta} p(b_j, b_{-j}) \right) dG_{-j}(b_{-j}).$$

By Cauchy's Integral Theorem since p is analytic $\oint_\Delta p(b_j, b_{-j}) = 0$. Hence $\oint_\Delta P_j(b_j) = 0$ so by Morera's Theorem $P_j(b_j)$ is analytic, and in particular real analytic when restricted to $(\underline{w}_j - h, W + h) \times 0$.

Hence the gain from deviating to b_j is given by a real analytic function $U_j(b_j) - \max_j U_j(\tilde{b}_j)$. That implies it is either identically zero or has finitely many zeroes. We can rule out the former case since $\max_j U_j(\tilde{b}_j) \leq 1$ and $\varrho(W) > 1$. Hence G_j must place weight only on the finitely many zeroes. $\square$

CHAPTER 12

Group Size In All Pay Contests

We previously saw how group size impacted all-pay auctions by advantaging the large group when monitoring costs are high and the small group when they are low. Using the tools developed in Chapter 11 we can extend this analysis to more general all-pay contests. The results of the analysis will be applied to data on corruption and federalism, and we shall see that the predictions of the model are to a good extent confirmed in the data. For completeness we reiterate the model of the previous section.

12.1. The Model

Two contestants $k = L, S$ for large and small compete for a prize worth $V_k > 0$ to each. As before we denote by $\eta_k > 0$ the size of group k where $\eta_s + \eta_L = 1$. Each contestant chooses an effort level $b_k \geq 0$. The probability of contestant k winning the prize is given by a *contest success function* $0 \leq p(b_k, b_{-k}) \leq 1$ that is symmetric in the sense that it depends on the efforts of the two contestants and not on their names. As before the contest success function is assumed to be continuous for $b_k \neq b_{-k}$, non-decreasing in b_k, and it must satisfy the adding-up condition $p(b_k, b_{-k}) + p(b_{-k}, b_k) = 1$. We allow for a discontinuous jump in the winning probability when we move away from $b_k = b_{-k}$, but require that when there is a tie the probability of winning is $1/2$. In this model there are resource limits: $b_k \leq \eta_k$.

Both groups have the same per capita cost C which is a continuous, for large enough effort levels differentiable, and non-decreasing function of per capita effort with $C(0) = 0$. The cost of effort is then

$$\varrho_k(b_k) = \eta_k C(b_k/\eta_k)/V_k.$$

We continue to assume for some B_k we have $\varrho_k(B_k) = 1$ and if $b_k > B_k$ then $\varrho_k(b_k) > 1$. To avoid degeneracy we assume that for contestant -1 the cost function $\varrho_{-1}(\cdot)$ is strictly increasing at the origin.

12.2. Success With a Common Prize

We first assume that there is a common prize, that is, $V_k = V$ and ask when the larger and when the smaller group has the greatest success.

We say that per capita cost is *asymptotically convex* if $\lim_{\varphi \to \infty} C'(\varphi) = \infty$. As a specific example, take the homogeneous cost function from from Section 8.6 $C(\varphi) = \xi\varphi^\alpha$ which is convex and asymptotically convex if $\alpha > 1$.

THEOREM 12.2.1. *If per capita cost is asymptotically convex and for some* $\hat{b} > 0$ *and* $\lambda > 1$ *we have* $\lambda C(\hat{b}) \leq V \left(p(\hat{b}, 0) - 1/2\right)$ *then a large enough group has greater success.*

There can be no analogous result for asymptotically concavity and small groups: if a group get small the resource constraint binds with increasing severity until eventually it will be outbid by the large group. Notice the condition that for some $\hat{b} > 0$ and $\lambda > 1$ we have $\lambda C(\hat{b}) \le V\left(p(\hat{b},0) - 1/2\right)$. For any insensitive contest this is satisfied since for $\hat{b} > 0$ we have $p(\hat{b},0) > 1 - q$ where $q < 1/2$, so the condition is satisfied for all sufficiently small $\hat{b}$. Roughly speaking what the condition rules out is contest success functions that increase too slowly when the opponent makes no effort.

The big picture is that the basic feature of convexity gives an advantage to the large group remains true for a broad range of contests.

PROOF. If per capita cost is asymptotically convex we calculate desire to pay B_S from $\eta_S C(B_S/\eta_S) = V$. As $\eta_L \to 1$ so $\eta_S \to 0$ we see that we must have $C(B_S/\eta_S) \to \infty$. In particular this means that $B_S/\eta_S \to \infty$, so the resource constraint must bind and $W_S = \eta_S \to 0$.

As $\eta_L \to 1$ we have have $\varrho_L(\hat{b}) = \eta_L C(\hat{b}/\eta_L)/V \to C(\hat{b})/V$, so in particular $\varrho_L(\hat{b}) \le \sqrt{\lambda}C(\hat{b})/\mathrm{V}$. Hence $\varrho_L(\hat{b}) < \sqrt{\lambda}\varrho_L(\hat{b}) \le p(\hat{b},0) - 1/2$. For $W_S = \eta_S/V < \hat{b}$ we have $p(\hat{b}, W_S)$ continuous so for sufficiently small W_S it must be that $\varrho_L(\hat{b}) < p(\hat{b}, W_S) - 1/2$. Hence L has a strong cost advantage so greater success. □

We now specialize to the homogeneous cost function $C(\varphi) = V\xi\varphi^\alpha$ where $\xi = \alpha/(V(1+\alpha))$. First we establish a useful fact:

LEMMA 12.2.2. *Suppose that $V < \eta_S\alpha/(1+\alpha)$ then the resource constraints do not bind*

PROOF. We have

$$\varrho_k(b) = \xi\eta_k(b/\eta_k)^\alpha = \xi\eta_k^{1-\alpha}b^\alpha = (\eta_k/\eta_{-k})^{1-\alpha}\varrho_{-k}(b).$$

The desire to bid is $\xi\eta_k(B_k/\eta_k)^\alpha = 1$ so that $B_k = \eta_k^{(\alpha-1)/\alpha}\left(V(1+\alpha)/\alpha\right)^{1/\alpha} < \eta_k$. This inequality binds only for $k = S$ and can be written as stated □

THEOREM 12.2.3. *Suppose that $V < \eta_S\alpha/(1+\alpha)$. If $\alpha > 1$ equilibrium is contested and L has greater success. If $\alpha < 1$ and*

$$V > \frac{1}{\sup_{b\le\eta_S} b^{-\alpha}\left(p(b,0) - 1/2\right)}\eta_S^{1-\alpha}\alpha/(1+\alpha)$$

then S has greater success.

Notice that for the all-pay auction $\sup_{b\le\eta_S} b^{-\alpha}\left(p(b,0) - 1/2\right) = \infty$ so the second condition is certainly satisfied. There are two parts of this: first, we rule out very large prizes so that resource constraints do not bind on the small group and we can apply results about uniform cost advantage. Given this convexity favors the large group just as for the all-pay auction. For concavity to favor the small group as in the all-pay case, we require the prize not be too large. If it is not then the small group will choose not to bid at all, and there will be a peaceful equilibrium.

PROOF. First, if $\alpha > 1$ so that $C(\varphi)$ is convex then L has a uniform cost advantage, while if $\alpha < 1$ so that $C(\varphi)$ is concave then S has a uniform cost advantage.

Second, we must check that the equilibrium is not peaceful. In the convex case $\alpha > 1$. Since $\varrho_k(b)$ has zero slope at the origin the low cost assumption is satisfied by both so L has low cost hence greater success.

In the concave case $\alpha < 1$ we require that the small group S has low cost, otherwise there will be a peaceful equilibrium. We need $\xi\eta_S^{1-\alpha}b^\alpha < p(b,0) - 1/2$ for some $b \leq \eta_S$, or $\xi < \eta_S^{\alpha-1}b^{-\alpha}\left(p(b,0) - 1/2\right)$. The sufficient condition can be written as

$$V > \frac{1}{\sup_{b\leq\eta_S} b^{-\alpha}\left(p(b,0)-1/2\right)}\eta_S^{1-\alpha}\alpha/(1+\alpha).$$

in which case S has greater success. □

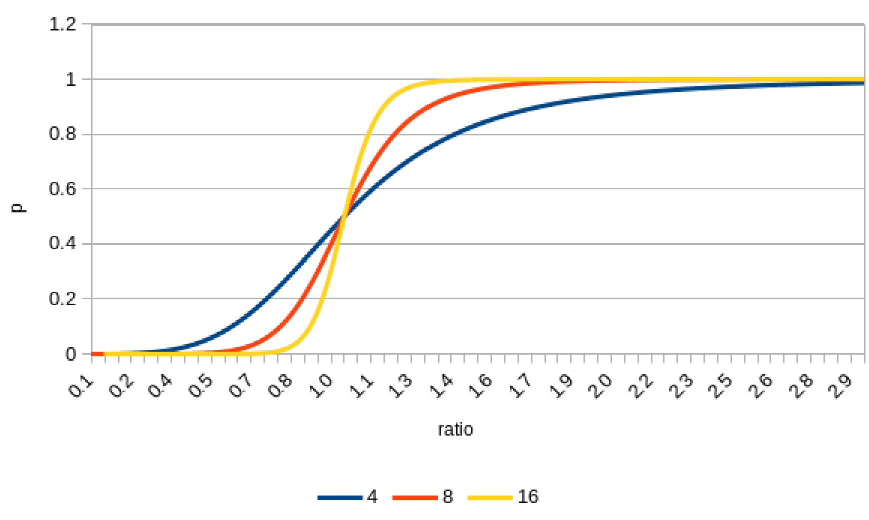
12.3. Mixed Tullock

We continue to assume homogeneous cost, but specialize further to the Tullock function. Note that in this case there is no question of a peaceful equilibrium, it must be contested, so we require only $V < \eta_S\alpha/(1+\alpha)$. The Tullock function as we observed above is left insensitive. We can show

Theorem 12.3.1. *If $\beta > 2\alpha$ then the objective function is generalized convex.*

The condition $\beta > 2\alpha$ that gives the all-pay auction result is not a strong one. If we ignore monitoring costs and take the quadratic benchmark case $\alpha = 2$ then the result holds for $\beta > 4$. Below the Tullock function is plotted with the probability of k winning against the ratio b_k/b_{-k} for three different values $\alpha = 4, 8, 16$.

Figure 12.3.1. The Tullock Function



As can be seen, even with $\beta = 16$ there is a substantial amount of noise - the steep part of the curve is scarcely vertical - and with $\beta = 4$ the noise is implausibly large for a situation such as voting: providing twice the effort leads to a 5% chance of failure.

PROOF. The objective function is

$$u_k(b_k, b_{-k}) = \frac{b_k^\beta}{b_k^\beta + b_{-k}^\beta} - \xi\eta_k^{1-\alpha} b_k^\alpha.$$

The condition for generalized convexity is that is for $0 < b_k \le b_{-k} \le \eta_L$ and $u_k(b_k, b_{-k}) \ge u_k(0, b_{-k}) = 0$ we have the derivative $D_k u_k(b_k, b_{-k}) > 0$ which is to say

$$\beta \frac{b_k^{\beta-1} b_{-k}^\beta}{\left(b_k^\beta + b_{-k}^\beta\right)^2} - \alpha\xi\eta_k^{1-\alpha} b_k^{\alpha-1} > 0$$

or equivalently

$$\frac{\beta}{\alpha} \frac{b_k^\beta b_{-k}^\beta}{\left(b_k^\beta + b_{-k}^\beta\right)^2} - \xi\eta_k^{1-\alpha} b_k^\alpha > 0. \tag{12.3.1}$$

We may compute

$$\frac{\beta}{\alpha} \frac{b_k^\beta b_{-k}^\beta}{\left(b_k^\beta + b_{-k}^\beta\right)^2} - \xi\eta_k^{1-\alpha} b_k^\alpha = \frac{\beta}{\alpha} \frac{b_k^\beta b_{-k}^\beta}{\left(b_k^\beta + b_{-k}^\beta\right)^2} + u_k(b_k, b_{-k}) - \frac{b_k^\beta}{b_k^\beta + b_{-k}^\beta}$$

Since we may assume $u_k(b_k, b_{-k}) \ge 0$ and $b_{-k}^\beta / \left(b_k^\beta + b_{-k}^\beta\right) \ge 1/2$ we have

$$\frac{\beta}{\alpha} \frac{b_k^\beta b_{-k}^\beta}{\left(b_k^\beta + b_{-k}^\beta\right)^2} + u_k(b_k, b_{-k}) - \frac{b_k^\beta}{b_k^\beta + b_{-k}^\beta} \ge \frac{\beta}{2\alpha} \frac{b_k^\beta}{b_k^\beta + b_{-k}^\beta} - \frac{b_k^\beta}{b_k^\beta + b_{-k}^\beta}$$

which is strictly positive for $b_k > 0$ if $\beta > 2\alpha$. □

From Theorem 11.7.1, Lemma 12.2.2 and the results on the all-pay auction Theorem 12.3.1 implies

COROLLARY 12.3.2. *Suppose $V < \eta_S\alpha/(1+\alpha)$ and $\beta \ge 2\alpha$. For $\alpha < 1$ the small group is advantaged and gets $1 - (\eta_S/\eta_L)^{1-\alpha}$ and the large group zero, while $\alpha > 1$ implies the large group is advantaged and gets $1 - (\eta_S/\eta_L)^{\alpha-1}$ and the small group gets zero. The advantaged group wins with probability $p_{-d} = 1 - (1/2)(\eta_d/\eta_{-d})^{\alpha-1}$ and rent dissipation is $(\eta_d/\eta_{-d})^{\alpha-1}$.*

Notice that once $\beta > 2\alpha$ it is irrelevant.

More is known about mixed equilibrium. The main unanswered question is whether equilibrium is unique. The equilibrium strategies for one equilibrium were found by Ewerhart (2015) using complex analytic methods he developed. He solved for a symmetric equilibrium with $\alpha = 1$ but this is valid for any $\beta/\alpha > 2$ by using the transformation $\tilde{b} = \xi\eta_k^{1-\alpha} b^\alpha$ similar to the proof of Theorem 12.4.1 below. It can be converted to an equilibrium for any asymmetric contest by the Alcalde and Dahm (2007) procedure we used in the proof of Theorem 11.7.1 above: the advantaged contestant uses the same strategy as in the symmetric game, while the disadvantaged contestant provides zero effort with probability $1 - (\eta_d/\eta_{-d})^{\alpha-1}$ and otherwise uses the same strategy as in the symmetric game.

The equilibrium is complicated: there are a countable number of mass points with mass accumulating at zero, although zero of course cannot be a mass point. For a range of values of β/α Ewerhart (2015) gives an approximate numerical solution by truncating the number of mass points to ten, which may be useful in applied research.

We should note one important point: for large α the inverse transformation $b = (1/2)^{(\alpha-1)/\alpha}(\tilde{b}/\xi)^{1/\alpha}$ maps most of the probability mass near one 1/2.

12.4. Convex Cost Pure Tullock

Theorem 12.3.1 tell us that if $\beta > 2\alpha$ the contest has a strong equivalence to an all-pay auction. We now give some well known opposite results for $\beta \leq \alpha$

THEOREM 12.4.1. *If $\beta \leq \alpha$ there are only pure strategy equilibria.*

PROOF. The objective function is

$$u_k(b_k, b_{-k}) = \frac{b_k^\beta}{b_k^\beta + b_{-k}^\beta} - \xi \eta_k^{1-\alpha} b_k^\alpha.$$

We change units of bids defining $\tilde{b} = \xi b^\alpha$ and study the problem in these units. Set $\gamma = \beta/\alpha$. We have

$$u_k(\tilde{b}_k, \tilde{b}_{-k}) = \frac{\tilde{b}_k^\gamma}{\tilde{b}_k^\gamma + \tilde{b}_{-k}^\gamma} - \eta_k^{1-\alpha} \tilde{b}_k.$$

The key fact is that for $\gamma \leq 1$, that is $\beta \leq \alpha$, the Tullock function is strictly concave in $\tilde{b}_k$ for $\tilde{b}_{-k} > 0$, while for $\tilde{b}_{-k} = 0$ it is weakly concave. If $-k$ has an atom at zero then there can be no equilibrium. This implies that for any equilibrium mixed strategy of the opponent the objective function is strictly concave, and so the only best response is a pure strategy, hence no mixed strategy equilibrium.

To show the Tullock function is strictly concave we simply compute the second derivative. The first derivative is

$$\gamma \frac{b_k^{\gamma-1} b_{-k}^\gamma}{\left(b_k^\gamma + b_{-k}^\gamma\right)^2}$$

so the second derivative is

$$\gamma \frac{(\gamma-1) b_k^{\gamma-2} b_{-k}^\gamma \left(b_k^\gamma + b_{-k}^\gamma\right) - 2\gamma b_k^{\gamma-1} b_{-k}^\gamma b_k^{\gamma-1}}{\left(b_k^\gamma + b_{-k}^\gamma\right)^3}$$

$$= \gamma \frac{(\gamma-1) b_k^{2\gamma-2} b_{-k}^\gamma + (\gamma-1) b_k^{\gamma-2} b_{-k}^{2\gamma} - 2\gamma b_k^{2\gamma-2} b_{-k}^\gamma}{\left(b_k^\gamma + b_{-k}^\gamma\right)^3}$$

$$= \gamma \frac{(-\gamma-1) b_k^{2\gamma-2} b_{-k}^\gamma + (\gamma-1) b_k^{\gamma-2} b_{-k}^{2\gamma}}{\left(b_k^\gamma + b_{-k}^\gamma\right)^3}$$

which is strictly negative for $\gamma \leq 1$ and $b_k, b_{-k} > 0$. □

THEOREM 12.4.2. *If $V < \eta_S \alpha/(1+\alpha)$ and $\beta \leq \alpha$ there is a unique equilibrium, it is in pure strategies, and the winning probabilities are*

$$p_k = \frac{1}{1 + (\eta_{-k}/\eta_k)^{\frac{\beta(\alpha-1)}{\alpha}}}$$

the equilibrium strategies are

$$b_k = \eta_k^{\frac{\alpha-1}{\alpha}} \left(\frac{\beta}{\xi\alpha} p_k (1-p_k) \right)^{1/\alpha}$$

and rent dissipation is $(\beta/\alpha)p_k(1-p_k)$.

PROOF. We prove uniqueness by computation.

We know there is an equilibrium and from Theorem 12.4.1 it is in pure strategies. We know from Lemma 12.2.2 that the resource constraints do not bind. We know that neither contestant can have an atom at zero, meaning that the equilibrium must be interior. Hence the first order conditions must be satisfied. Making use of the derivative in equation 12.3.1 we solve the first order condition as

$$b_k = \eta_k^{\frac{\alpha-1}{\alpha}} \left(\frac{\beta}{\xi\alpha} \frac{b_k^\beta b_{-k}^\beta}{\left(b_k^\beta + b_{-k}^\beta\right)^2} \right)^{1/\alpha}.$$

This gives the ratio

$$\frac{b_k}{b_{-k}} = \left(\frac{\eta_k}{\eta_{-k}}\right)^{\frac{\alpha-1}{\alpha}}$$

and we can plug that into the Tullock function to find the probability that contestant k wins is

$$p_k = \frac{1}{1 + (\eta_{-k}/\eta_k)^{\frac{\beta(\alpha-1)}{\alpha}}}$$

with the corresponding effort levels being determined by plugging back into the first order condition.

To find rent dissipation we plug back into the objective function to find $u_k(b_k, b_{-k}) = p_k - \frac{\beta}{\alpha} p_k(1-p_k)$, add across players and subtract from one. $\square$

12.5. Intermediate Tullock

When $\beta > 2\alpha$ we have equivalence to the all-pay auction. When $\beta \leq \alpha$ we have a unique pure strategy equilibrium. What happens in the intermediate case $1 < \beta/\alpha \leq 2$. Here the answer has been worked out and we summarize what it is. First, Feng and Lu (2017) show that the equilibrium is unique. Second, Nti (1999), Nti (2004) gives conditions under which it is in pure strategies: it is of course the same equilibrium described in Theorem 12.4.2. Third, Wang (2010) finds the equilibrium when the Nti condition fails.

First we describe Nti's condition for pure strategy equilibrium. Let $\lambda = (\eta_{-d}/\eta_d)^{\alpha-1} \geq 1$ denote the relative cost advantage of the advantaged contestant $-d$, and $\gamma = \beta/\alpha$. Let $\overline{\gamma}$ be the solution to $\lambda^\gamma(\gamma-1) = 1$. Since the LHS is increasing in γ this has a unique solution. Then there is a unique pure strategy equilibrium when $\gamma \leq \overline{\gamma}$. When $\overline{\gamma} < \gamma \leq 2$ there is a unique equilibrium in which the advantaged contestant plays a pure strategy and the disadvantaged contestant mixed between providing no effort and providing a single fixed level of effort.

The Semi-mixed Equilibrium. Define $\mu = (\gamma - 1)^{-1/\gamma}$. Note that $\mu > 1$. From Wang (2010) the disadvantaged contestant uses effort

$$b_d = \left(\frac{1}{\xi \eta_d^{1-\alpha}} \frac{1}{\gamma} \right)^{1/\alpha}$$

with probability μ/λ, less than 1 by Nti's condition, and otherwise makes no effort, while the advantaged player employs a pure strategy outbidding the disadvantaged player with $b_{-d} = \mu^{1/\alpha} b_d$.

Most interesting from our point of view is to understand how $\overline{\gamma}$ depends on λ. If $\lambda = 1$ so the two contestants are equal this is $\overline{\lambda} = 2$, and in the symmetric case we go directly from pure strategy to all-pay equivalence. Using the implicit function theorem

$$\frac{d\overline{\gamma}}{d\lambda} = -\frac{\gamma \lambda^{\gamma-1}(\gamma - 1)}{\lambda^{\gamma} + \lambda^{\gamma}(\gamma - 1)\log \lambda} < 0$$

since we are interested only in the case $\gamma > 1$ and $\lambda > 1$. In other words, symmetry favors the pure strategy equilibrium, while asymmetry favors the mixed equilibrium.

12.6. Is It Pure or Is It Mixed?

We would like to know what are β and α and whether or not we are likely to see mixed or pure equilibria. To assess this we analyzed election data. The idea is to estimate both the mixed equilibrium model assuming $\beta > 2\alpha$ and pure equilibrium model assuming $\beta \leq \alpha$. Note that these results pure strategy are valid also for $1 < \beta/\alpha \leq 2$ provided that the most asymmetric election is not "too" asymmetric.

We assume that there is a common value satisfying $V < \eta_S \alpha/(1 + \alpha)$. We assume as well that $\alpha > 1$, that is, monitoring costs are not so great as to lead to concavity in the cost function. This is strongly favored by the data as we will see.

12.6.1. The Data and Estimation Technique. To do an empirical analysis we need many similar elections. We chose to analyze California elections for the U.S. House of Representatives in 2016 and 2020. These are both presidential election years, so avoid the issue of lower turnout in off year elections. California was not a swing state in the Presidential election since it was relatively certain that the Democratic candidate would win California's electoral votes. Hence the local issues were important for bringing voters to the polls. Moreover these issues seem small enough that it is plausible that the small party is not willing to turn out all their voters, that is, $V < \eta_S \alpha/(1 + \alpha)$. The congressional districts are all of similar population. Finally, while there are many offices and issues on California ballots it is reasonable to suppose that the primary issue was whether the Democratic or Republican candidates and proposals would win: the winner of the House race seems a reasonable indication of this.

There are 53 Congrssional districts in California, so 106 observations in total.[1] The California Secretary of State provides detailed information about these elections, including the winning candidate and party, and party registrations, which we take as an indicator of party size. Specifically, for party size we used the most recent voter registration data prior to the election, including only Democratic and Republican registrations and ignoring independents (in some cases many) and other parties (few). The average Democratic party registration was 46% and Republican 25%. There were 7 upset elections where the party with the fewer registrations won the election.

We employed this data by using the respective models to compute log-likehood functions. These functions and the corresponding statistical analysis is reported for each case.

12.6.2. Mixed Equilibrium. First we assume that $\beta > 2\alpha$. This implies a mixed equilibrium in which the winning probabilities from Corollary 12.3.2 are given by $p_L = 1 - (1/2)(\eta_S/\eta_L)^{\alpha-1}$. Notice that this depends only on α, so we are entitled only to a lower bound on $\beta > 2\alpha$.

The log-likelihood function is plotted below in Figure 12.6.1. The maximum occurs where $\alpha = 6$, and this implies $\beta > 10$. Values of α below about 3.5 can be rejected by the asymptotic chi-squared test with one degree of freedom at the 95% level. This lower bound gives $\beta > 7$. The upper confidence bound is $\alpha = 12$ with corresponding $\beta \geq 24$. In other words we can be relatively confident that $\beta \geq 7$ and α is between 3.5 and 7.

The maximized value of the log-likelihood function is -18.71.

12.6.3. Pure Equilibrium. Now we assume that $\beta \leq \alpha$. From Theorem 12.4.2 the winning probability is given by

$$p_L = \frac{1}{1 + (\eta_S/\eta_L)^{\frac{\beta(\alpha-1)}{\alpha}}}.$$

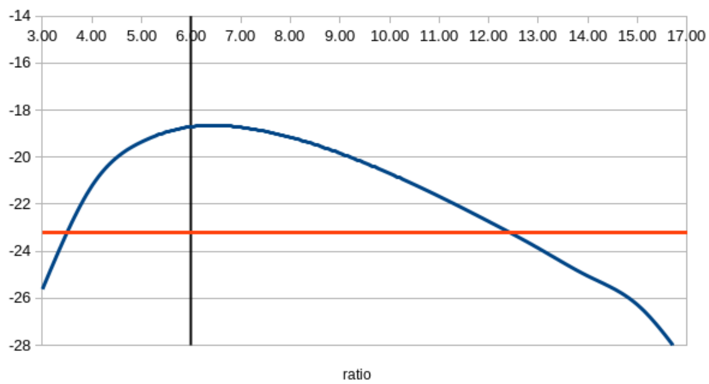
In particular this depends only on the parameter $\gamma \equiv \beta(\alpha-1)/\alpha$. The log-likelihood function is plotted below in Figure 12.6.2. The maximum occurs where $\gamma = 7$ and values below about 3.5 and above 11 can be rejected by the asymptotic chi-squared test with one degree of freedom at the 95% level.

Observe that $\beta = \gamma\alpha/(\alpha - 1)$. The RHS is increasing in α and asymptotes to 1. On the other hand, α cannot fall below β, so the least value of α is given by $\alpha = \gamma - 1$. In other words $\alpha \geq \gamma - 1$ and $\gamma \leq \beta \leq \gamma(\gamma - 1)/(\gamma - 2)$. For the maximum likelihood estimate of this gives $\alpha \geq 6$ and $7 \leq \beta \leq 8.4$. For the lower end of the confidence interval this is $\alpha \geq 2.5$ and $3.5 \leq \beta \leq 5.8$. For the upper end $\alpha \geq 10$ and $11 \leq \beta \leq 12$. In particular we can be relatively confident that $\alpha \geq 2.5$ and $3.5 \leq \beta \leq 12$.

The maximized value of the log-likelihood function is -18.95.

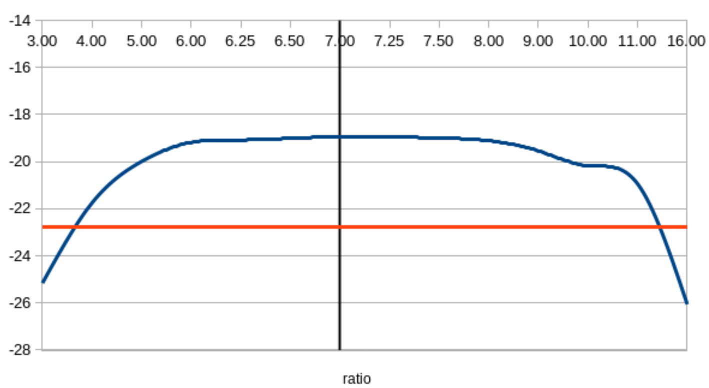
[1]Coate and Conlin (2004b) have data on 363 liquor referenda in Texas, but there is no equivalent of party registration to measure party size for these referenda and they are scattered over twenty years. Shachar and Nalebuff (1999) have data on 539 statewide Presidential elections but this data is over forty years and good voter registration is not available for all states. Both of those studies include a variety of nuisance variables in an effort to create comparability between the elections. Roughly, we accept a smaller sample size in order to have a cleaner dataset.

FIGURE 12.6.1. Mixed Log-likelihood Function



The horizontal axis is α, the vertical axis is the log likelihood with the blue line being the data and the red line the maximum minus the 95% cutoff for a chi-squared with 1 degree of freedom of 3.841. The vertical line marks the maximum of the likelihood function.

FIGURE 12.6.2. Pure Log-likelihood Function



The horizontal axis is $\gamma = \beta(\alpha - 1)/\alpha$, the vertical axis is the log likelihood with the blue line being the data and the red line the maximum minus the 95% cutoff for a chi-squared with 1 degree of freedom of 3.841. The vertical line marks the maximum of the likelihood function.

12.6.4. Discussion. Roughly speaking there is not much difference between the mixed and pure model. The log likelihood is very slightly higher for the mixed than for the pure strategy model, but the difference is so small (0.24) that we must conclude that both models fit the data about equally well. There is no evidence to tell us whether $\beta > 2\alpha$ or $\alpha \leq \beta$: both are consistent with the data. Never-the-less there is broad consistency between the parameters from the two different models: α between 3.5 and 7 and β between 7 and 12 is consistent with the confidence bands of both models.

A key point to emphasize is that the estimates of α are fairly large. The mixed model suggests 6 and the pure model suggests at least this amount. In other words the cost function is flat and then rises quite steeply. There are several reasons we believe this may be the case. First, there are typically many committed voters, so we would expect the cost function to be initially quite flat. Second of all, we assumed that V_k was the same for both parties and independent of party size. There is reason to think this might not be the case. In general a party with an attractive or incumbent candidate is likely to have a larger V_k then a party with a weak challenger. Moreover, a weak party will have trouble convincing a good candidate to run and is less likely to be fielding an incumbent. Consequently it may be that $V_L/V_S = (\eta_L/\eta_S)^\mu$ for some $\mu > 0$. This is equivalent to increasing α, that is further enhancing the size advantage of a party.

Finally a cautionary note. We can rule out with a high degree of confidence $\alpha = 2$ the quadratic case. This unfortunately is a staple of applied and empirical work - it is employed, for example, by Coate and Conlin (2004b). In the future it might be a better practice to estimate α.

12.7. Corruption and Federalism

There is an alternative interpretation of the conflict resolution function discussed in Herrera et al. (2015). Rather than a probability of winning the prize V it may be that $p(b_k, b_{-k})$ represents a deterministic share of the prize received by each group. For example, think of of a number of regions. In a federal system each region would be separately governed: an election would determine how many districts each group controls. By contrast there could be a central system in which the winner controls all the districts. In this case rather than thinking of $1/\beta$ as the level of exogenous uncertainty we would think of it as the level of federalism. In particular we might think of $\beta = 1$ where the prize is proportional to effort as federalism and the auction where the winner takes all as centralism.

We now want to ask how the size of the prize affects the relative advantages of federalism and centralism. If the prize is low, resource constraints do not bind, and centralism leads to more rent seeking than federalism. On the other hand when the stakes are high centralism is not so bad because neither side can bid terribly high. To get a handle on this, let us take the symmetric linear case $\eta_S = \eta_L = 1/2$, that $\alpha = 1$ and compare federalism with $\beta = 1$ to centralism with $\beta = \infty$. Define the advantage of federalism $A(V)$ as the difference between utility (unnormalized) under federalism and that under centralism as a function of the size of the prize V.

Theorem 12.7.1. *We have for*

(1) $V < 1/2$ that $A(V) = V/2$
(2) $1/2 < V < 1$ that $A(V) = 1/2 - V/2$
(3) $V > 1$ that $A(V) = 0$

Proof. From Theorem 12.4.2 under federalism $b_k = (1/4)(1/\xi)$ for $b_k \leq 1/2$ and $b_k = 1/2$ for $\xi < 1/2$. Utility for each side is $1/2 - \xi b_k$ giving a combined utility of $1/2$ for $\xi > 1/2$ and $1 - \xi$ for $\xi < 1/2$. Unnormalizing utility multiplying through by V and using $\xi = 1/(2V)$ this gives $V/2$ for $V < 1$ and $V - 1/2$ for $V > 1$.

Under centralism from Theorem 12.3.2 for $V < \eta_S\alpha/(1+\alpha)$, which is to say $V < 1/4$ utility is 0. For $(1/2V) > 1/4$ both contestants to randomize uniformly over $[0, 1/2]$ with a spike at $1/2$, still giving a utility of 0. For $V > 1/2$ both use the pure strategy of $1/2$ and joint utility is $V - 1/2$. □

The key point is that for low V there is very little advantage of federalism, but it rises, hitting a peak where $V = 1/4$. Then it declines linearly, with centralism being advantageous when $V = 1/2$ and reaching 0 again when $V = 1$ and remaining constant after. However: besides the size of the prize there are several reasons one may be superior to the other:

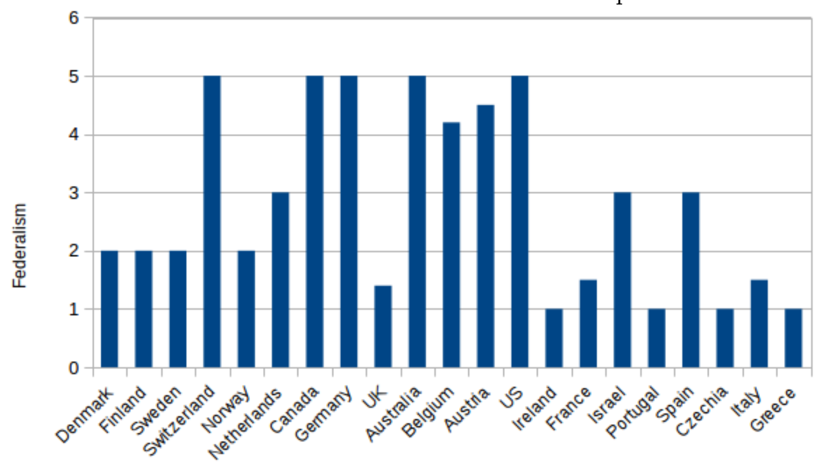
- Risk: individuals and parties are generally risk averse. A 50% share of a prize (federalism) is generally regarded as superior to a 50% chance of a prize (centralism).
- Increasing returns to scale: there may be economies of scale in government administration - this is particularly likely to be the case in defense and security. One hundred different armies are not generally as effective as one army one hundred times as large.

If on the balance these forces favor centralism then we might expect that for small stakes as well as large centralism is advantageous, while for intermediate stakes federalism would be advantageous. If we think that real political systems - perhaps because of evolutionary pressure - have efficiency properties we might ask if it is indeed true that it is for countries with intermediate stakes that we see federalism?

To answer whether countries with intermediate stakes are more federalist we must measure both stakes and federalism. Fortunately Lijphart (2012) provides us with the latter. A natural measure of stakes is the level of corruption of a country: in a very corrupt country the winner can seize a greater prize than in a system where the government is professional and (as many economists seem to believe) benevolent. In Figure 12.7 we plot federalism against corruption for a group of "similar" countries.

Strikingly the most countries with the most intermediate levels of corruption seem to have the highest levels of federalism: the highest bars are in the middle. Five of the six least corrupt countries have a level of federalism of 2-3. The eight most corrupt countries have a level of federalism of 1-3: both extremes are quite centralized as the theory predicts. By contrast six of the seven "intermediate" countries all have federalism levels of 4 or more.

FIGURE 12.7.1. Federalism Versus Corruption



This is a subset of IMF "advanced countries" excluding those with a population less than one million (for which the meaning of "federalism" is unclear) and for which Lijphart (2012) provides an index of federalism. The vertical axis reports the index for the period 1981-2010. The horizontal axis orders countries in order of corruption as measured by the Transparency International corruptions perception index from 2016.

CHAPTER 13

Pivotality, the Anti-folk Theorem and the Paradox of Voting

Let us go back to the basic model of a political contest with group members organized into two parties with the outcome determined by majority voting. The structure of the model of individual behavior studied in Chapter 4 is that introduced by Palfrey and Rosenthal (1985) to analyze voting: each voter has a independent and randomly determined cost of participation which is negative for at least some committed voters. We can describe the model of behavior used by Palfrey and Rosenthal (1985) as rational selfish behavior: voters are rational and care only about their own utility. Absent peer enforcement mechanisms, the incentive to vote is the chance to shift the outcome of the election from unfavorable to favorable and by doing so claim a share of the prize. The key factor in determining individual behavior, then, is the probability that the voter will be *pivotal* meaning that the election is decided by a single vote - otherwise the decision of the voter to participate does not matter. This is a special case of the more general problem of contributing to a public good where the punishment for failing to contribute is a collective punishment - either all group members are punished, or none at all. In the case of voting the punishment is the loss of the election, a cost borne by all group members.

Pivotality is controversial because both computations of equilibrium and empirical studies of the probability of being pivotal indicate that in large elections there is so little incentive to vote that to a good approximation only committed voters will turn out.[1] This, however, has the consequence that turnout should be independent of strategic considerations such as the importance of the election and there is strong evidence that this is not the case. This apparent contradiction in the context of voting has been termed "the paradox of voter turnout." It has motivated the large literature of which this book is part investigating models that feature, at least superficially, something other than selfish rational behavior. Models of ethical or altruistic voters study rational voters who are not selfish - either for ethical or other reasons their preferences are other-regarding and they care about the consequences of their actions for other voters. We study voters who are both selfish and rational - but we recognize that groups and political parties are not blank slates, but rather are based on social networks which have the ability to provide incentive to group members through punishments and rewards.

This does not mean we should reject pivotality or models that study pivotality. As Levine and Palfrey (2007) show, the simple Palfrey and Rosenthal (1985) model

[1] In settings where partial exclusion from the benefits of a public good are possible the incentive for voluntary contribution may be much greater. For issues such as tax changes or farm subsidies where exclusion is difficult this would have little relevance, but if the rewards are government jobs, contracts or other spoils then it maybe. See, for example, Nitzan and Ueda (2011).

finds strong support in the laboratory. That does not make it relevant to large elections - but does make it relevant to smaller elections - including elections that take place within juries, committees, and legislative bodies. However, just because pivotality is important in these smaller elections does not mean that peer incentives are not important as well.

Here we are going to examine the closely connected issues of pivotality and collective punishment. Roughly speaking we will find that when groups are large collective punishment does not work well in either theory or practice - and we will also find common ground for models that incorporate both pivotal considerations and individual punishments by peers.

13.1. How Relevant is Pivotality in Large Elections?

Not surprisingly, the probability of being pivotal in large elections is very low as documented by Mulligan and Hunter (2003) and Shachar and Nalebuff (1999). Based on Shachar and Nalebuff (1999)'s calculations on the probability of casting a pivotal vote in presidential elections, in all but a few states a rational voter for whom it cost \$1 to vote would have to win a prize larger than the wealth of the wealthiest person in the world for it to be worth voting. We can also point to more refined modelling and calculations: Coate et al. (2008) show that in a sample of Texas liquor referendums, elections are much less close than what would be predicted by the pivotal voter model, and Coate and Conlin (2004b) show that the ethical voter model better fits the data than the model of pivotal voters.

A less discussed but also important issue is the scaling of elections. As the electorate grows the probability of being pivotal declines, and so should voter turnout. If there are committed voters, then in large elections basically only they will turnout, so voter turnout will be flat. If there are no committed voters but some voters with very low costs, so that the density of voting costs is positive at 0 then voter turnout should be roughly proportional to the inverse cube root of the number of voters $N^{-1/3}$.

Pivotal Turnout in Large Elections. Following Penrose (1946) and Chamberlain and Rotschild (1981) suppose that there is a two party election in which each party has N voters who independently turn out to vote with probability ϕ. Let t_i be 0 or 1 as the voter voted or not. Then the vote differential is $\sum_{i=1}^{N} t_i - \sum_{j=1}^{N} t_j$. According to the central limit theorem this should be approximately normally distributed with mean 0 and variance $2N\phi(1-\phi)$. The probability q that this lies between -1 and 1 (that is everyone is pivotal) is to a good approximation twice the height of the density at 0, that is to say, $q = 2/\sqrt{4N\pi\phi(1-\phi)}$. In a pivotal voting model, the voters who cast their votes are those for whom the cost of voting $c < qB$ where B is the individual benefit of winning the election. Hence for the percentage of voters who turn out is approximately $\phi = G(qB)$ where G is the cdf of voting costs, or $\phi = g(0)qB$ where g is the density of voting costs. Hence pivotality should satisfy $q = \sqrt{1/(N\pi\phi(1-\phi))} \approx \sqrt{1/(N\pi\phi)} = \sqrt{1/(N\pi g(0)qB)}$. and we may solve to find $q = (1/(N\pi g(0)qB)^{1/3}$ and turnout $\phi = g(0)B\,(1/(N\pi g(0)qB)^{1/3}$.

How do elections scale? Below we present data for post-war national elections in consolidated democracies with per capita income above the world average and voluntary voting, and examine how voter turnout depends on the size of the country,

TABLE 13.1.1. Turnout Across Countries

Country	% Turnout	Voting Age Population	Period
Iceland	88.33	148,068	1946-2013
New Zealand	83.06	2,077,795	1946-2014
Norway	78.71	2,892,801	1945-2013
Denmark	83.52	3,575,272	1945-2015
Sweden	82.63	6,134,212	1948-2014
Hungary	65.91	7,919,118	1990-2014
Czech Republic	72.19	9,011,908	1990-2013
Canada	65.45	16,428,034	1945-2011
United Kingdom	71.57	41,111,434	1945-2015
Japan	68.26	75,404,908	1946-2014
United States	55.74	210,541,626	1948-2012

Turnout and population data are averages in the post-war period of OECD countries with voluntary voting and Freedom House Index of political freedom below 3. We included Denmark, the UK, and Sweden and excluded the eurozone since in the latter substantial power has passed to the EU itself, so that the significance of "national" elections is different than in fully sovereign nations - in particular for the smaller EU nations. However, including the rest of the EU does not alter the overall picture. Data is taken from http://www.idea.int.

We see that there is a group of small countries with voting age population ranging from 140,000 to 10 million with high voter turnout of 65% to 89% and a group of large countries with population ranging from 16 million to 211 million with lower voter turnout ranging of 55% to 72%. Within these groups of countries there is very little variation or evidence of negative correlation between size and turnout.

While it is true that the group of smaller countries generally have higher turnout than the larger countries, the large and the small groups turnout is quite homogeneous while population varies by a factor of nearly 10 or more - this data is in no way consistent with scaling by the cube root of the population. In fact, it is not even consistent with a monotone relation between turnout and population, which is the main prediction of the pivotal voter model. A similar picture emerges if we turn attention to the dynamics of voter turnout in advanced democracies: turnout declined on average by a mere 10% in the past 50 years in the face of a voting age population which more than doubled.[2]

We should mention that there are other factors that militate against the pivotal voter and in favor of the idea the monitoring and social closeness matter. Empirical analysis by Gray and Caul (2000) relates post-war turnout decrease with the decline of mobilizing actors such as labor parties and trade unions. See also Knack (1992)

[2]In fact in Denmark and Sweden turnout increased by 3% and 6%, respectively.

on the decline of American voter turnout and its relation to a weakened enforcement of social norms.[3]

13.2. The Anti-Folk Theorem and Collective Punishment

The empirical failure of the pivotal voting model in large elections reflects a deep theoretical problem sometimes called the anti-folk theorem. Empirically it seems that incentives for individuals to contribute do not come entirely from the possibility of defeat. Hence our premise is that groups overcome public goods problems by punishing individuals for failing to adhere to social norms.

It is undeniable that in addition to the individual incentive of facing peer punishment there are incentives to participate due to collective punishments. Moreover, real groups face repeated public goods problems and an alternative and widely used model to that of peer punishment is one in which incentives for public goods contributions are due to the possibility of retaliation over future public goods. For example, one reason to adhere to the social norm of voting might be that individual group members understand that if they do not vote the social norm will break down and in the future elections will be lost due to low turnout. Leaving aside the fact that in practice groups generally punish individuals for failing to adhere to social norms, an important reason that we do not examine these types of schemes is that they depend on collective punishments and because in large groups collective punishments do not work. This is an old observation in the literature on oligopoly and the prisoner's dilemma - dating back at least to Radner (1980)'s work on the subject. Radner shows that in a repeated oligopoly game cooperation breaks down as the number of competing firms grows large. This is the conclusion of an extensive literature including results by Green (1980), Sabourian (1990), and Levine and Pesendorfer (1995). More recently Sugaya and Wolitzky (2021) show that without personalized punishment cooperation breaks down when there are many players, and characterize how much personalization is needed to support cooperation. We base our presentation on the simpler model of Fudenberg et al. (1998).

To understand the problem with collective punishments, let us examine a group composed of identical members $i = 1, 2, \ldots, N$ who can participate at a cost of D. There is a noisy signal of whether or not a member has participated with the bad signal generated for sure if the member did not participate and with probability $\pi > 0$ if the member did.

We now want to explore what happens when the only available punishment is a collective punishment P - that is, individuals cannot be punished, either the entire group must be punished or nobody at all. This corresponds to a situation where social sanctions are not available and, for example, the only punishment for not voting is that the party loses or that other group members withhold their votes in a future election. More broadly it corresponds to a situation where the punishment for failing to adhere to a social norm is the breakdown of the social

[3] It is somewhat ironic given the limited empirical support that pivotal voter models have been widely used, for example the Poisson model of Myerson (1998), to study information aggregation in elections. While this makes perfectly good sense in small elections the application to information aggregation in large elections must be subject to some doubt. The voluntary public goods provision model has also been used, for example, to study the impact of group size. For example Esteban and Ray (2001) use a voluntary donation model. In this model per capita effort goes to zero as absolute group size grows. Moreover if the size of both groups is big enough then the larger group is always more likely to win.

norm resulting in a collective punishment for the entire group. This is the type of mechanism studied by Wolitzky (2013), by Acemoglu and Wolitzky (2015) and by Ellison (1994). Ellison recognizes that - as we are about to show - such schemes do not work well when the group is large - and indeed the same observation motivated the Kandori (1992) model of social norms from which our own peer punishment model is derived.

We suppose that the group must pick a mechanism consisting of a punishment size P and a rule for determining punishment based on signal profiles in an effort to enforce compliance with the social norm for all members. Let us first consider schemes that determine whether or not to punish based on the number of bad signals. Let Q^1 denote the probability of punishment if all group members participate and let Q^0 denote the probability of punishment if all group members except one participate.

First consider the case where $\pi = 0$ so that there is no noise. In this case we can punish if any bad signals are received: then $Q^1 = 0$ and $Q^0 = 1$. If the punishment $P \geq D$ the social norm is incentive compatible: if everyone participates all pay the cost of D. If any single member deviates they save the cost of participation but certainly receive a punishment at least equal to this. Hence each individual member is pivotal: if any one violates the social norm the agreement breaks down, so none do so.

So far so good, unfortunately this simple mechanism of triggering punishment if any bad signals occur breaks down badly if there is noise, that is if $\pi > 0$. Since the probability of no bad signal being received if everyone participates is $(1-\pi)^N$ we have $Q^0 - Q^1 \leq 1 - \left[1-(1-\pi)^N\right] = (1-\pi)^N$. Incentive compatibility demands that $(Q^0 - Q^1)P = D$ so that the expected cost of punishment is $\left[1-(1-\pi)^N\right] P \geq \left[1-(1-\pi)^N\right] D/(1-\pi)^N$ which has the unfortunate feature that as $N \to \infty$ the expected cost of punishment becomes infinite. If we assume an upper bound $\overline{P}$ on how much punishment is possible then the incentive constraint implies $P \geq D/(1-\pi)^N$ so that for large enough N the constraint $\overline{P} \geq D/(1-\pi)^N$ will be violated. The problem is that trying to punish on a single bad signal means that with noise and a large population the collective punishment is triggered almost for certain, and since you are going to be punished anyway, you might as well cheat.

Despite many proposals to induce effort provision in large populations with collective punishment through a clever choice of punishment rule - it cannot be done. One of the authors was briefly involved with a startup internet firm that believed otherwise - it is not easy to convince people who develop convoluted rules that they will not work. The natural thought is that the problem can be fixed by being more tolerant - recognizing that bad signals will be generated when everyone adheres to the social norm so perhaps we should punish only if a threshold fraction $\beta > 0$ is exceeded. For example, punish when twice the expected number of bad signals is observed, or something like that. We might suspect that this does not work since it will not work even in the case that $\pi = 0$: in that case everyone will cheat! To see if this can work in the case that $\pi > 0$, observe that for large N to a good approximation the distribution of the fraction of signals is normally distributed. Let

$$n^1 = \frac{\sqrt{N}(\beta - \pi)}{\sqrt{\pi(1-\pi)}}, n^0 = \frac{\sqrt{N}\left(\beta - ((N-1)\pi + 1)/N\right)}{\sqrt{\pi(1-\pi)}}$$

then from the central limit theorem for large N we have $Q^1 \approx \Phi\left(n^1\right), Q^0 \approx \Phi\left(n^0\right)$ where Φ is the standard normal distribution function. So to a good approximation[4]

$$Q^0 - Q^1 \approx (n^0 - n^1) C e^{-(1/2)\left(n^0 - n^1\right)^2} \leq (n^0 - n^1) C = C \frac{1-\pi}{\sqrt{N}\sqrt{\pi(1-\pi)}}.$$

This is again the inverse square root of N rule for being pivotal. We see immediately the problem: no matter what the choice of β as $N \to \infty$ we have $Q^0 - Q^1 \to 0$ and since the benefit of deviating remains at least $1 - (Q^0 - Q^1)\overline{P}$ once again for large N the social norm fails to be incentive compatible. By choosing a $\beta > 0$ we solve the problem of punishing too frequently - but at the expense of assuring that a deviation by a single individual has very little effect on the outcome. This failure is more general. For any mechanism where a collective punishment of $P \leq \overline{P}$ is determined only by the total number of bad signals Fudenberg et al. (1998) in their Lemma A give the generally valid bound $Q^0 - Q^1 \leq 2/(\sqrt{N}\min\{\pi, 1-\pi\})$, so incentive compatibility must fail for N sufficiently large.

Suppose we drop the assumption that the punishment must be a constant and that it depend only on the number of bad signals and that we allow general punishment schemes $\overline{P} \geq P(z) \geq 0$, or indeed a fixed[5] finite number of different kinds of punishments that have a different effect on different individuals. We could then, for example, base the punishment only on the behavior of a single member and give that member incentives to follow the social norm. Unfortunately with a collective punishment we cannot do this simultaneously for any substantial fraction of the group: Fudenberg et al. (1998) prove in their Proposition $1'$ that regardless of the punishment scheme the fraction of members participating must fall to zero as $N \to \infty$. This result depends only on the fact that there is a minimal amount of noise in observing individual behavior and that the group is limited to collective punishments - it does not rest on symmetry assumptions or specific details of the game.

Notice that it is possible to use collective punishment to punish a few key individuals - and if they have the ability to punish individually the larger number of other group members this can be bootstrapped into an effective incentive scheme. To take a simple example, we can imagine the CEO of a business firm who can punish individual workers, for example, by firing them or cutting their pay. The CEO may be tempted to shirk by not monitoring the employees - but then the firm will fail and everyone including the CEO will be punished. This collective punishment gives the CEO the needed incentive to monitor and punish workers. Even with limited supervisorial capacity - so that a supervisor can monitor only a few employees - a hierarchical organization of the type studied by Williamson is possible. However, while political organizations are hierarchical in practice and these types of incentives may be relevant for the upper echelon, in political organizations the rank and file - individual voters, or individual farmers in a farm lobby - cannot easily be punished or rewarded by the hierarchy, so that for these types of organizations some form of peer discipline must be at work.

---

[4] The constant C is of course $1/\sqrt{\pi}$ (where π is not a probability but the number $\pi \approx 3.14$). As the constant does not matter, to avoid having π appear in the same equation with two different meanings or to change notation for a few paragraphs, we just introduce the constant C.

[5] That is the number of different types of punishments does not grow with N.

13.3. Incentive Constraints with Pivotality

Just as peer punishment is important in large elections, pivotality is important in small elections and it would be useful to have a model that incorporates both features - the more so as the importance of pivotality will increase as monitoring costs grow large. Here we sketch out a simple version of such a model.

We examine the setting of a group in which group members independently draw types y_i uniformly distributed on $[0, 1]$. They may either contribute to a public good or not: if they do not they bear no cost and if they do they bear a cost $c(y^i) = y^i$. The fraction φ of the group who contribute is the probability that group members receive a prize $V > 1$.

As usual we analyze a social norm of participating for $y^i \leq \varphi$ together with a monitoring technology $\pi < \pi_1$ the probability of a bad signal for a non-contributor conditional on adhering to and violating the social norm and an endogenous punishment P. Now, however, we assume that the group has N members who share the prize V.

Group members for whom $y^i \leq V/N$ will wish to contribute regardless of incentives. Assume that $V/N < 1$ so that full voluntary participation is not incentive compatible. Then for $\varphi > V/N$ and $y^i \leq \varphi$ we have the incentive constraint

$$\varphi V - y^i \geq \frac{N-1}{N}\varphi V - \pi_1 P.$$

This incorporates the notion of pivotality in this context: if you do not participate the probability of getting the prize is reduced by $1/N$. If this holds with equality for the marginal type $y^i = \varphi$ then all remaining incentive constraints will be satisfied, so we find

$$P = \frac{\varphi - \varphi V/N}{\pi_1}.$$

We can compute directly the corresponding cost

$$C = \int_0^\varphi y^i dy^i + \pi(1-\varphi)P = \varphi^2/2 + (\pi/\pi_1)(1 - V/N)\varphi(1-\varphi)$$

and the optimal social norm is

$$\varphi = \max\left\{0, \frac{V - (\pi/\pi_1)(1 - V/N)}{1 - 2(\pi/\pi_1)(1 - V/N)}\right\}.$$

There is no deep message here: it is perfectly straightforward to incorporate the "pivotality," that is, the change in the probability of success due to individual action, into the incentive constraint, and we get the answer we expect: pivotality substitutes for punishment which is reduced by $\varphi V/(\pi_1 N)$ (recall from 4.2.1 that with $c(y) = y$ we have $P = \varphi/\pi_1$). As N grows, naturally this effect diminishes.

13.4. The Holdup Problem and the Tragedy of the Anticommons

A nice illustration of pivotality and uncertainty and one quite relevant to political economy, lobbying, and public policy is the classical holdup problem. This can be formulated as the problem that was faced by the Germans prior to the formation of the customs union in the 19th Century. Along the Rhine river ships carrying cargoes of varying values pass and there are many castles each of which can charge a toll to passing ships. This toll creates a public "bads" problem - charging a higher toll reduces the amount of shipping and lowers the profit of all castles. If the value of cargo is known with certainty then the monopoly price can be supported: each castle charges a fraction of the value of the cargo, and if any tries to raise their toll the ship will not pay and the sale is lost. In this case all castles are pivotal. Once again, however, with uncertainty, pivotality is lost and if there are enough castles shipping and castle revenue shrink to zero.

For concreteness, imagine that the value of a cargo ρ is uniformly distributed over $[0,1]$ and is known to the shipper. Along the river are a number N of castles with each castle i charging a fee p_i for passage. The castles do not know the value of ρ but only that they are drawn uniformly on $[0,1]$. If all other castles set the price p and a deviant castle charges the price p_i then the total cost of passage faced by the shipper is $(N-1)p+p_i$ and the shipper will operate only if this is less than or equal to ρ. Hence the expected revenue of a deviant castle is p_i times the probability of a cargo of value at least $(N-1)p+p_i$, that is, expected revenue is $(1-(N-1)p-p_i)p_i$. The optimal price to set is therefore determined by the first order condition $1-(N-1)p-2p_i=0$. Hence the symmetric Nash equilibrium of this game is at $p=1/(N+1)$. The equilibrium total price is $Np=N/(N+1)$ which makes traffic shrink to zero as N grows. Moreover, the total revenue received by all castles is $N(1-Np)p=N/(N+1)^2$- thus as $N\to\infty$ not only does shipping shrink away to zero, but so does the revenue of the castles.

Here the problem is that each castle by setting a high price imposes an externality on its neighbors by reducing shipping. This is like a public good problem - all castles would benefit if they colluded and agreed to the single price maximizing total revenue $(1-Np)Np$, that is $p=1/(2N)$ - and this is of course what the establishment of the customs union did. If there were no uncertainty then the monopoly price for the N castles is ρ and each castle could charge ρ/N - if any castle tried to raise the price the sale would be lost - without uncertainty each castle would be pivotal.

The bottom line is that a consequence of the anti-folk theorem - roughly, that collective punishment does not work for large groups - is that many small monopolies producing complementary goods are much worse than a single monopoly controlling all production. This idea has many applications. The presentation here is based on Boldrin and Levine (2005)'s application to the analysis of patent systems. If many different ideas are required to innovate then a strong patent system strangles innovation. The problem is that many independent patent holders each separately license all the ingredients needed to innovate. A similar problem can occur in construction or the opening of a new business. If permits from many different corrupt agencies are required then development will come to a halt: each corrupt official demands too high a bribe. Even if mere paperwork is required - if each agency benefits by a high paperwork requirement so as to get more resources from a central authority - this also can bring development to a halt. Yet another example can be found in the Chari and Jones (2000) analysis of pollution rights: if each individual property owner in a city owned air rights and any polluter had to get a permit from each property owner then there would be no pollution - and also no output. The broad problem of too many owners of complementary resources is called by Heller (2008) the tragedy of the anticommons, and his book documents numerous examples of gridlock brought about by the holdup problem.

CHAPTER 14

Repeated Play, Voluntary Fines and Collective Punishment[1]

So far we have treated the problem of group self-organization as a static one shot mechanism design problem. In reality groups play over long periods of time, so engage in a repeated rather than one-shot game. Nowhere is this more apparent than in modern cartel theory which focuses on the repeated play, strategies such as grim trigger, and results such as the folk theorem. Originally the focus was on strongly symmetric equilibrium as in Green and Porter (1984), Rotemberg and Saloner (1986) or Abreu et al. (1990). These grim trigger equilibria are problematic because they rely on collective punishment. A fundamental insight from Fudenberg and Maskin (1994)'s work on repeated games with imperfect information is that collective punishments such as price wars are inefficient in comparison to punishments that involve transfer payments from guilty to innocent: transfer payments provide incentives without diminishing overall cartel profits. This no doubt is why in practice cartels do not use price wars and collective punishments. However, the underlying repeated game equilibria are complex and in some ways do not reflect how cartels actually operate. In Levenstein and Suslow (2006)'s survey of the empirical literature we find "after the adoption of an international price-fixing agreement in the bromine industry, the response to violations in the agreement was a negotiated punishment, usually a side-payment between firms, rather than the instigation of a price war... As repeatedly discovered by these cartel members, the threat of Cournot reversion is an inefficient way to sustain collusion." Indeed one of the main conclusions of the survey is precisely the point that cartels do not generally use price wars or collective punishment to deter cheating but when possible use fines.

In this chapter we bring together several of our earlier themes. We know that fines are relatively efficient because the involve transfer payments rather than socially costly punishment. However: why do firms pay fines? Here we find an answer in the use of collective punishment to enforce the rules. While output and/or price are difficult to monitor whether or not firms pay their fines is easy to monitor. Hence there is little cost of using a collective punishment to enforce fines. The use of voluntary transfer payments are effective in the presence of noise and they convert a noisy signal of behavior into a sharp signal of adherence to the rules. Using this approach we introduce dynamics to the model and use the strongly symmetric equilibrium approach to show how the best (from the point of view of the group) equilibria of the repeated game are given by the solution of precisely the type of static mechanism design problem we have been studying. There is one proviso however: if group members are too impatient then this constrains the size

[1]This chapter is based on Levine (2021a).

of fines that can be used, and the type of punishment restrictions we studied earlier becomes endogenous depending on the discount factor.

In this chapter we focus on the case of a finite number of firms, so pivotality will play a (limited) role.

14.1. The Model

We study a dynamic Cournot industry with N identical firms with common discount factor $0 \leq \delta < 1$. As is standard in the repeated game literature we use average present value throughout. In period $t = 1, 2, \ldots$ firm i produces output $X \geq x_t^i \geq 0$. As usual average firm output is $\overline{x}_t$. The profit of firm i in period t is random and given by $\tilde{u}^i(\overline{x}_t, x_t^i)$ with a firm independent mean $u(\overline{x}_t, x_t^i)$. We assume that firm specific profits are not perfectly correlated so that $\tilde{u}^i(\overline{x}_t, x_t^i)$ is private information. Let us assume that the mean utility functions $u(\overline{x}_t, x_t^i)$ are smooth and sufficiently well behaved that there is a unique Cournot equilibrium at $\overline{x}^C$ and the standard regularity condition that the derivative of monopoly profit with respect to output is strictly negative at the Cournot equilibrium

The cartel that sets a common quota $\overline{x}^C \geq \varphi_t \geq 0$ for all of the identical firms at the beginning of each period. After this production takes place. For each firm an independent public binary signal is observed of whether $x_t^i \leq \varphi_t$, that is whether the quota was adhered to or not. As usual if the quota is adhered to a bad signal is generated with probability $\pi > 0$ and if it is not with probability $\pi_1 > \pi > 0$. We denote by Z denote the total number of bad signals in the industry. Finally, after signals are commonly observed, firms may optionally choose to pay fines P_t^i where a fraction $1 - \psi$ of the proceeds are distributed among the remaining firms and ψ of the fine is lost due to transaction costs with $0 \leq \psi < 1$.[2]

By restricting attention to perfect public strongly symmetric equilibrium we may use the results of Abreu et al. (1990) to give a simple characterization of the best agreement achievable by the cartel. An *agreement* consists of a quota φ, the rule that firms produce to quota[3] $x_t^i = \varphi$, a system of required fines $\overrightarrow{P}(Z)$ paid by firms with bad signals, termination (of the cartel) probabilities $Q(Z)$ and the rule that if any firm fails to pay a required fine termination (of the cartel) takes place with probability one.[4] A *strongly symmetric profile* is an agreement along with the rule that if termination (of the cartel) takes place each firm will produce at the Cournot level $\overline{x}^C$ forever and otherwise the agreement will continue for another period. Our notion of equilibrium is strongly symmetric subgame perfect equilibrium: we say that an agreement is *incentive compatible* if in the strongly symmetric profile every firm is willing to pay the fine and no firm wishes to deviate from the quota. Our goal is to characterize the *best agreement*: the incentive compatible agreement that yields the highest per firm profit.

Note that the assumption that firms face idiosyncratic prices together with perfect public equilibrium rules out the use of firm's own price information in cartel enforcement. While this a useful theoretical simplification it is also empirically relevant. Levenstein and Suslow (2006)'s survey of the empirical literature discusses

---

[2]So for, for example, if there are two firms and both are required to pay a fine f then each pays the fine to the other, paying P and receiving $(1 - \psi)P$ for a net per firm loss of ψP.

[3]Since we assume the quota is no greater than the Cournot equilibrium output $\varphi \leq \overline{x}^C$ no firm will produce less than the quota.

[4]Because the payment of fines is perfectly observed this is without loss of generality.

"Monitoring Output and Prices" but in fact only discusses the monitoring of output. In their discussion of "Cartel Breakdown" there is no indication that fluctuations in individual firm prices or market prices play a role in enforcement: in general it appears that the information used by cartels lies in evidence of adherence to cartel rules as in this model, not on data about market prices.

14.2. The Theorem

Denote the Cournot utility by $u^C = u(\overline{x}^C, \overline{x}^C)$. Define the greatest utility from deviating from a quota as $u^B(\varphi) = \max_{x^i} u\left(((N-1)/N)\varphi + x^i/N, x^i\right)$. A crucial aspect of the model is that the transactional loss from a fine ψ is smaller than 1 so at least some small part of the fine is received by cartel members. In other words, the social cost of using a fine as punishment is strictly less than the size of the punishment. By contrast, if the cartel is terminated, the social cost of the punishment is at least equal to the size of the punishment. This suggests that it is always better to use fines, and our first result shows that this is true: termination of the cartel should be used only to enforce the payment of fines.

THEOREM 14.2.1. *In any best agreement $Q(Z) = 0$.*

PROOF. Let $\overline{u}$ be the utility from the best agreement. Then $\overline{u} \geq u^C$ and if $\overline{u} = u^C$ then there is no loss of generality in taking $Q(Z) \equiv 0$. Hence to prove $Q(Z) = 0$ we may assume $\overline{u} > u^C$.

Define the collective punishment as $q(Z) = (\delta/(1-\delta))Q(Z)\left[\overline{u} - u^C\right]$. Suppose that $N-1$ firms adhere to the quota and let $\Pi(Z)$ be the probability that they generate exactly Z bad signals. Note that since the quota is no greater than the Cournot equilibrium output $\varphi \leq \overline{x}^C$ no firm will wish to produce less than the quota. Hence the only incentive constraint is that firms weakly prefer producing to quota to deviating to a higher output and receiving $u^B(\varphi)$ with a higher probability of punishment:

$$u(\varphi,\varphi) - \sum_{Z=0}^{N-1} \Pi(Z)\left[(1-\pi)q(Z) + \pi\left(\overrightarrow{P}(Z+1) + q(Z+1)\right)\right]$$

$$\geq u^B(\varphi) - \sum_{Z=0}^{N-1} \Pi(Z)\left[(1-\pi_1)q(Z) + \pi_1\left(\overrightarrow{P}(Z+1) + q(Z+1)\right)\right]$$

which may be written as

$$u^B(\varphi) - u(\varphi,\varphi) \leq (\pi_1 - \pi)\sum_{Z=0}^{N-1} \Pi(Z)\left[\overrightarrow{P}(Z+1) + (q(Z+1) - q(Z))\right]. \tag{14.2.1}$$

The incentive constraint for paying fines is

$$\overrightarrow{P}(Z) \leq (\delta/(1-\delta))(1-Q(Z))\left[\overline{u} - u^C\right] = (\delta/(1-\delta))\left[\overline{u} - u^C\right] - q(z). \tag{14.2.2}$$

Per firm profits when all firms adhere to the quota are

$$u(\varphi,\varphi) - \sum_{Z=0}^{N-1} \Pi(Z)\left[(1-\pi)q(Z) + \pi\left(\psi\overrightarrow{P}(Z+1) + q(Z+1)\right)\right]. \tag{14.2.3}$$

Start with an incentive compatible plan in which $Q(Z_0) > 0$ and consider increasing $\overrightarrow{P}(Z_0)$ by r and decreasing $q(Z_0)$ by r. The RHS of 14.2.1

$$(\pi_1 - \pi) \sum_{Z=0}^{N-1} \Pi(Z) \left[\overrightarrow{P}(Z+1) + (q(Z+1) - q(Z))\right]$$

$$+\mathbf{1}(Z_0 \leq N-1)(\pi_1 - \pi)\Pi(Z_0)r,$$

so the incentive constraint for adhering to the quota is satisfied. It is similarly clear that 14.2.2 remains satisfied.

Per firm profits are

$$u(\varphi,\varphi) - \sum_{Z=0}^{N-1} \Pi(Z) \left[(1-\pi)q(Z) + \pi \left(\psi \overrightarrow{P}(Z+1) + q(Z+1)\right)\right]$$

$$+\mathbf{1}(Z_0 \geq 1)\Pi(Z_0 - 1)\pi(1-\psi)r + \mathbf{1}(Z_0 \leq N-1)\Pi(Z_0)(1-\pi)r$$

which is strictly increasing in r. We conclude that $Q(Z) \equiv 0$. □

Define the monitoring cost $M(\varphi) \equiv [\psi\pi/(\pi_1 - \pi)]\left(u^B(\varphi) - u(\varphi,\varphi)\right)$. Because termination of the cartel is used only to enforce the payment of fines the punishment from termination imposes a simple constraint on the size of the fines used to enforce quotas. This enables us to reduce the dynamic problem of finding the best agreement to the simple static mechanism design problem of maximizing one-period utility net of a cost of enforcing the agreement.

THEOREM 14.2.2. *In any best agreement the quota $\hat{\varphi}$ is a solution of the static mechanism design problem* $\max_\varphi u(\varphi,\varphi) - M(\varphi)$ *subject to* $M(\varphi) \leq \psi\pi(\delta/(1-\delta))\left[u(\varphi,\varphi) - M(\varphi) - u(\overline{x}^C, \overline{x}^C)\right]$ *and any such solution is part of a best agreement in which* $\phi(Z) = \hat{\phi}(y)$.

PROOF. We continue to let $\overline{u}$ be the utility from the best agreement. Define $\Phi = \sum_{Z=0}^{N-1} \Pi(Z)\overrightarrow{P}(Z+1)$. Since $\Pi(Z)$ is the probability that $N-1$ firms generate exactly Z bad signals, we have $\sum_{Z=0}^{N-1} \Pi(Z) = 1$. Since $Q(Z) = q(z) = 0$ the objective function is to maximize $u(\varphi,\varphi) - \pi\psi\Phi$, the incentive constraint for adhering to the quota is $u^B(\varphi) - u(\varphi,\varphi) \leq (\pi_1 - \pi)\Phi$, and the incentive constraint for paying fines is $\overrightarrow{P}(Z) \leq (\delta/(1-\delta))\left[\overline{u} - u^C\right]$. The fine paying constraint may also be written as $\max_Z \overrightarrow{P}(Z) \leq (\delta/(1-\delta))\left[\overline{u} - u^C\right]$ from which it is clear that $\Phi \leq (\delta/(1-\delta))\left[\overline{u} - u^C\right]$, while conversely if that is the case then $\overrightarrow{P}(Z) = \Phi$ satisfies the constraint. Hence the constraint $\Phi \leq (\delta/(1-\delta))\left[\overline{u} - u^C\right]$ suffices. Since the objective function is decreasing in Φ the quota adherence constraint must hold with exact equality $u^B(\varphi) - u(\varphi,\varphi) = (\pi_1 - \pi)\Phi$ solving for Φ and plugging in then gives the result. □

This is similar to our standard mechanism design problem, but the size of the fines (punishments) is constrained by

$$M(\varphi) \leq \psi\pi(\delta/(1-\delta))\left[u(\varphi,\varphi) - M(\varphi) - u(\overline{x}^C, \overline{x}^C)\right]$$

. However, if the discount factor is sufficiently large then this constraint does not bind.

COROLLARY 14.2.3. *If we denote by* $\underline{M}$ *the minimum of* $M(\varphi)$ *over unconstrained solutions to the problem* $\max_\varphi u(\varphi,\varphi) - M(\varphi)$ *and* $\overline{U}$ *the corresponding maximized utility then for* $\overline{\delta} \equiv \overline{M}/(\underline{M} + \psi\pi(\overline{U} - u^C))$ *we have* $\overline{\delta} < 1$ *and for* $\delta \geq \overline{\delta}$ *utility from the optimal agreement is* $\overline{U}$*, that is, the constraint does not bind.*

PROOF. The only thing to be proven here is $\overline{U} > u^C$. If we set $\theta = \psi\pi/(\pi_1 - \pi)$ we can write the objective as $u(\varphi,\varphi) - \theta\left(u^B(\varphi) - u(\varphi,\varphi)\right)$. Consider the derivative with respect to φ at $\varphi = \overline{x}^C$. From the envelope theorem the derivative of the second part $\left(u^B(\varphi) - u(\varphi,\varphi)\right)$ is zero, so that the derivative is just that of $u(\varphi,\varphi)$, that is, the monopoly profit. But under our standard regularity conditions the derivative of monopoly profit with respect to output is strictly negative at the Cournot equilibrium so we are done. $\square$

CHAPTER 15

Sticky Adjustment and Reputation Traps[1]

Earlier, in Chapter 3, we discussed the fact that often social norms can be quite slow to change. One reason as discussed in that chapter is that it can be costly to reach new agreements on social norms. Here we observe that with reputational effects social norms may not change even when adjustment costs are negligible.

It is conventional to think that a good reputation is easy to lose and hard to gain. One reason we suspect this might be the case is that if you have a good reputation people will be eager to do business with you – hence if they are cheated it will quickly become known. On the other hand if you have a bad reputation few will do business with you so even if you are honest few will find out. Insofar as the importance of good social norms for economic success revolves around good treatment of immigrants and foreign investors there might well be a reputational effect. A region that has a reputation for poor treatment of foreigners is unlikely to get much immigration or foreign investment and so is unlikely to have thriving urban centers of production and innovation. It then becomes the case that even if treatment of outsiders is improved nobody is likely to find out. As a result the dysfunctional norm of cheating outsiders may become self-fulfilling - a kind of reputation trap. We may ask, for example, is the reason that Nigeria does not mimic Japan because it would be so costly to make the change? Or is partly due to a a reputation trap?

15.1. The Environment

As there is nothing in this theory special about groups: it would apply as well to an individual, we will simply study a standard dynamic game between a representative short-run player (2) - the foreign investors - and the long-run player (1) - the group investing in a social norm. Each period $t = 1, 2, \ldots$ a stage game is played. In the stage game the long-run player must first choose whether or not to provide effort - that is decide whether to invest in a social norm that protects outsiders. Let $a_1 \in \{0, 1\}$ denote the decision of the long-run player with 1 meaning to provide effort and the cost being ca_1 where $0 < c < 1$. This might include monitoring and punishment costs as well as other costs. The short-run player moves second and without observing the effort choice of the long-run player decides whether to enter $a_2 = 1$ or stay out $a_2 = 0$. The short-run player receives utility 0 for staying out, utility -1 for entering when no effort has been made and utility $V > 0$ for entering when effort is provided.

As is standard in the reputational literature there are privately known types of long-run player. Specifically there are three types $\tau \in \{b, n, g\}$ where g means "good" (a beneficial event), b means "bad" (an adverse event), and n means "normal."

[1]This chapter is based on Levine (2021b).

For example, a beneficial event might be an outside threat that makes it essential to adapt a considerate social norm - as in Belloc et al. (2016) - and an adverse event might be an occupation by extractive foreigners so that it is only sensible to be dishonest with them.

Player type is fixed during the lifetime of the player. The good and bad types are behavioral types: the good type always provides effort and the bad type never does. The stage game payoff of the normal type is given by $a_2 - ca_1$. Players care only about expected average utility during their lifetime.

The life of a long-run player is stochastic: with probability δ the player continues for another period, and with probability $1-\delta$ is replaced. This replacement is not observed by the short-run player. When a long-run player is replaced the type may change. The probability that type τ is replaced by a type $\sigma \neq \tau$ is $Q_{\tau\sigma}\epsilon/(1-\delta)$ where $Q_{\tau\sigma} > 0$. We are interested in the case in which types are persistent - that is, in which ϵ is small.

Choice of Units. The transition probabilities conditional on replacement are $Q_{\tau\sigma}\epsilon/(1-\delta)$. Since the replacement probability is $1-\delta$ that means that the transition probabilities to types $\sigma \neq \tau$ are $Q_{\tau\sigma}\epsilon$, that is, are assumed not to depend upon δ, but rather on ϵ. Roughly speaking this means that the amount of time until the ergodic distribution of the Markov process is a good approximation will be of order $1/\epsilon$.

At the beginning of each period a public signal z of what occurred in the previous period is observed and takes on one of three values: $1, 0, N$. If entry took place last period the signal is equal to last-period long-run player effort decision - so z is 0 or 1. If the short-run player stayed out last period then with probability $1 \geq \pi > 0$ the signal is equal to last period long-run player effort decision and with probability $1-\pi$ the signal is N. Here we are to think of "1" as a good signal (effort was observed), "0" as a bad signal (it was observed that there was no effort) and "N" as no signal. Note that in this chapter signals are about the behavior of the group and not of individual group members.

There are two features of this information technology. First, even when the short-run player stays out some information is generated. Second, when the short-run players enter information is perfect. Subsequently we will model more closely investment and information and demonstrate the robustness of our results when information upon entry is less than perfect.

The game begins with an initial draw of the public signal $z(1)$ and private type $\tau(1)$ from the common knowledge distribution $\mu_{z\tau}(1)$.

Players are only aware of events that occur during their lifetime. The long-run player also knows their own generation T.[2] Let h denote a finite history for a long-run player. A strategy for the normal type of long-run player is a choice of effort probability $\alpha_1(h, t, T)$ as a function of privately known history, calendar time, and generation T. A strategy for the short-run player is a probability of entering $\alpha_2(z, t)$ as a function of the beginning of period signal and calendar time.

We study Nash equilibria of this game and assume *generic cost* in the sense that

$$c \notin \left\{ \delta, \frac{\delta}{2-\pi}, \frac{\delta\pi}{1-\delta+\delta\pi}, \frac{\delta\pi(\pi-\delta\pi)}{(1-\delta\pi)(1-\delta)) + \delta\pi(\pi-\delta\pi)} \right\}.$$

[2] That is, how many replacement events have taken place since the beginning of the game.

15.2. Short-run Player Beliefs and Time Invariant Equilibrium

If players know calendar time they can use this information to coordinate their play in an implausible way. In particular, if short-run players stay out and no information is generated it eventually becomes likely that the long-run player has migrated back to a "normal" type. It is now possible for the short-run players and long-run player to coordinate. On a particular date it is common knowledge that if the long-run player is normal honest behavior will take place and that the short-run player will enter. This is then a self-fulfilling prophecy. It is not, however, a very compelling one: it requires that both players agree about the exact timing of events in the long-distant past and that they agree that "today is the day." To rule this out we assume that agents know only about events that took place during their lifetime and that short-run player strategies and beliefs are independent of calendar time. Notice that this same assumption is implicit in the definition of a Markov equilibrium, but is weaker since long-run player strategies may depend on the entire lifetime history of events as well as generation and calendar time.

For brevity all references to a decision problem of the long-run player should be understood to refer to the normal type. A strategy for a short-run player is a now a time invariant probability of entering $\alpha_2(z) \in [0,1]$ as a function of the beginning of period signal. Given such a strategy the normal type faces a well-posed Markov decision problem. It depends only on the probability α_2 with which the short-run player enters. Let $V(\alpha_2)$ denote the corresponding expected average value of utility. First period utility is $\alpha_2 - ca_1$. With probability δ the game continues and the probability of the next signal is $P(z'|z, a_1)$ where $P(1|z,1) = P(0|z,0) = \alpha_2(z) + (1 - \alpha_2(z))\pi$ and $P(N|z, a_1) = (1 - \alpha_2(z))(1 - \pi)$. Hence the Bellman equation is

$$V(\alpha_2) = \max_{a_1}(1-\delta)\left[\alpha_2 - ca_1\right] + \delta \sum_{z'} P(z'|z, a_1)V(\alpha_2(z')). \tag{15.2.1}$$

As usual, this has a unique solution. The set of best responses, for the normal type, then, is determined entirely by the current state through $\alpha_2(z)$. Hence at time t with signal z_t any best response of the normal type $\alpha_1(h_t, t, T_t)$ must lie in this set. Time invariant beliefs of the short-run player about the effort probability of the normal type, which we denote by $\alpha_1(z)$, are then a weighted average of these best responses - and so must also be a best response and lie in this set.

Dynamic Programming and the Bellman Equation. A basic tool for studying dynamic optimization problems is dynamic programming and the Bellman equation. While a full treatment is beyond the scope of this book the intuition is not. The idea is to break an infinite horizon optimization problem into a two period optimization problem by assuming that the solution starting in the second period is already known and using this to solve the problem in both periods. In our setting, the future play of the short-run players is entirely determined by their current play. If they are using the strategy $\alpha_2(z)$ then when the realization of z in period 1 is z' there would be a new optimization problem starting in period 2 starting with $\alpha_2(z')$. Suppose we thought that the highest obtainable payoff with this starting point is $V_2(\alpha_2(z;))$. In period 1 we would want to maximize the expected average present value of utility today and utility starting tomorrow

$$V_1(\alpha_2) = \max_{a_1}(1-\delta)\left[\alpha_2 - ca_1\right] + \delta \sum_{z'} P(z'|z, a_1)V_2(\alpha_2(z')).$$

That is, we would do the best possible today, presuming starting tomorrow we would obtain the best possible result. The Bellman equation says that because the problem is stationary, that is, the problem starting in period 2 is the same as that starting in the same position in period 1, we should have $V_1(\alpha_2) = V_2(\alpha_2)$. The underlying theory of dynamic programming shows that this has a unique solution that correctly characterizes the *value function* mapping starting points to best utility and that once $V(\alpha_2)$ is known the solution of the maximization problem in 15.2.1 is an optimal strategy for the long-run player.

Prior to observing the signal z_t the short-run player at time t has unconditional beliefs about the joint distribution $\mu_{z\tau}(t)$ from which the signal and type of the long-run player are drawn. After observing z_t short-run player beliefs about long-run player type are given by the conditional probability $\mu_{\tau|z_t}(t)$. This together with beliefs about the normal type effort $\alpha_1(z_t)$ determines $\mu^1(z_t, t)$ the overall beliefs about the probability of long-run player effort. The short-run player strategy $\alpha_2(z_t)$ must then be a best response to those beliefs.

The evolution of $\mu_{z\tau}(t)$ depends upon the initial condition $\mu_{z\tau}(1)$ and the beliefs of the short-run player about the probabilities $\alpha_1(z)$, $\alpha_2(z)$ with which earlier normal-type long-run and short-run players chose actions. It does not depend on the actual choice of those actions or the earlier signals, none of which are observed. This has two consequences. First, no action or deviation by the long-run player has any effect on the evolution of $\mu_{z\tau}(t)$. Second, the evolution of $\mu_{z\tau}(t)$ is deterministic as it does not depend on the stochastic realization of actions, signals or types. The stochastic nature of short-run player beliefs are due to the single stochastic variable they observe, the signal; that is, $\mu_{\tau|z_t}(t)$ is stochastic because z_τ is.

Since $\mu_{z\tau}(t)$ follows a deterministic law of motion, if we let $\overrightarrow{\mu}(t)$ denote the vector with components $\mu_{z\tau}(t)$ that law is $\overrightarrow{\mu}(t+1) = \mathbf{A}\overrightarrow{\mu}(t)$ where $\mathbf{A}$ is a Markov transition matrix the coefficients of which are determined by $\alpha_1(z), \alpha_2(z)$ and π, Q, ϵ.[3] To have an equilibrium with time invariant beliefs it must be that $\overrightarrow{\mu}(t+1) = \overrightarrow{\mu}(t)$ and this is true if and only if the initial condition $\mu_{z\tau}(1)$ is a stationary distribution of $\mathbf{A}$. For time invariance we cannot have arbitrary initial short-run player beliefs

[3]See the the Appendix to this chapter for the computation.

$\mu_{zt}(1)$, but only initial beliefs that are consistent with the strategies of the players and the passage of time.

We take our object of study, then, to be *time invariant equilibrium.* This is a Nash equilibrium in which the initial beliefs of the short-run players are determined endogenously to be the stationary distribution that arises from the equilibrium strategies. It is conveniently described as a triple $(\alpha_1(z), \alpha_2(z), \mu_{z\tau})$ where $\alpha_1(z)$ and $\mu_{z\tau}$ are time invariant beliefs of the short-run player and $\alpha_2(z)$ is the strategy of the short-run players. The conditions for equilibrium are that $\alpha_1(z)$ is a solution to the Markov decision problem induced by the short-run player strategy $\alpha_2(z)$, that $\mu_{z\tau}$ is a stationary distribution of the Markov transition matrix $\mathbf{A}$ determined by $\alpha_1(z), \alpha_2(z)$, and Q, ϵ, and that $\alpha_2(z)$ is a best response to beliefs about long-run player action $\mu^1(z)$ determined from $\alpha_1(z), \mu_{z\tau}$.

Let $z(h)$ be the most recently observed signal by the long-run player in the history h. We may conveniently summarize the discussion:

THEOREM 15.2.1. *If $(\alpha_1(z), \alpha_2(z), \mu_{z\tau})$ is a time invariant equilibrium then the strategies $\alpha_1(y, t, T) = \alpha_1(z(y)), \alpha_2(z, t) = \alpha_2(z)$ are a Nash equilibrium with respect to the initial condition $\mu_{z\tau}(1) = \mu_{z\tau}$. Conversely if $\alpha_1(h, t, T), \alpha_2(z, t)$ is a Nash equilibrium that satisfies the time invariant short-run player condition that the short-run player equilibrium beliefs $\alpha_1(z, t) = \alpha_1(z), \mu_{z\tau}(t) = \mu_{z\tau}$ and equilibrium strategy $\alpha_2(z, t) = \alpha_2(z)$, then $(\alpha_1(z), \alpha_2(z), \mu_{z\tau})$ is a time invariant equilibrium.*

Hereafter by equilibrium we mean time invariant equilibrium.

15.3. Characterization of Equilibrium

Our interest is in the existence of a reputation trap: that following an adverse event the long-run player will be trapped with a bad reputation until a beneficial event "comes to the rescue." Our main result characterizes when a trap does and does not occur. It shows that there is a single pure strategy equilibrium that is one of three types, and give conditions under which that equilibrium is unique. In reading the theorem, note that $1 - \delta + \delta\pi$ is a weighted average of 1 and π so is strictly greater than π.

THEOREM 15.3.1. *For given V, Q there exists an $\underline{\epsilon} > 0$ such that for all $\epsilon \in (0, \underline{\epsilon}\pi^2(1-\pi))$*

i. [bad] *If*

$$c > \delta$$

then there is a unique equilibrium, it is strict and in pure strategies, there is no effort by the normal type, and the short-run player enters only on the good signal.

ii. [trap] *If*

$$\delta > c > \delta\pi/(1 - \delta + \delta\pi)$$

then there is exactly one pure strategy equilibrium, it is strict, the normal type provides effort only on the good signal, and the short-run player enters only on the good signal. *If in addition $c > \delta/(1 + \delta(1 - \pi))$ this is the unique equilibrium.*

iii. [good] *If*

$$c < \delta\pi/(1 - \delta + \delta\pi)$$

then there is exactly one pure strategy equilibrium, the normal type always provides effort, and the short-run player enters only on the good signal.

Note that the boundary cases are ruled out by the generic cost assumption.[4]

This result is described in terms of the comparative statics of entry cost c: it shows how the set of equilibria changes as c is reduced. As all of the cutoffs $\delta\pi/(1-\delta+\delta\pi) = \delta\pi/(1-\delta(1-\pi))$ and $\delta/(1+\delta(1-\pi))$ are strictly increasing in δ the results may equally be described in terms of increasing the discount factor δ, with the (more complicated) cutoffs described in terms of c .

The proof is outlined below and the details can be found in the online appendix to Levine (2021b), reproduced as the Appendix to this Chapter. The result has two main parts: the characterization of pure strategy equilibria and the uniqueness of pure strategy equilibria. We will discuss each of these in turn.

The pure strategy equilibrium is relatively intuitive. The assumption that ϵ is small means that types are highly persistent so the short-run player does not put much weight on the possibility of the type changing. Given the possible strategies of the long-run player the signal 0 indicates either a bad type or a normal type who will not provide effort unless they anticipate entry by the short-run player. Hence it makes sense for the short-run player not to enter in the face of bad signal. Similarly the signal 1 indicates either a good type or a normal type who will provide effort if entry is anticipated, so it makes sense for the short-run player to enter in the face of a good signal.

More subtle is the inference of the short-run player when the signal N is observed. The short-run player can infer that the previous short-run player chose not to enter - hence must have received the bad signal or was in the same boat with the signal N. As a result while less decisive than the signal 0 the signal N also indicates past bad behavior by the long-run player, so staying out is a good idea.

For the long-run player the choice is whether to provide effort when entry is anticipated and when it is not. The difference between the two cases lies in the probability that effort results in a good reputation - that is in a good signal - which we may denote by $p = 1$ when entry is anticipated and $p = \pi$ when it is not. It is useful to consider the problem for general values of p: when the cost c is incurred there is a probability p of successfully establishing a good reputation and gaining $1 - c$ in the future and probability $1 - p$ of failing to establish a good reputation and starting over again. Here the expected average present value of the gain from effort is $\Gamma = -(1-\delta)c + \delta p(1-c) + \delta(1-p)\Gamma$ or

$$\Gamma = \frac{\delta p(1-c) - (1-\delta)c}{1-\delta(1-p)}.$$

If this is negative, that is $\delta p(1-c) < (1-\delta)c$, then it is best not to provide effort and conversely. Take first the case where information is revealed immediately, that is $p = 1$. This is the situation most conducive to effort. The condition for not wishing to provide effort is $c > \delta$ so when this is the case there will be no effort. This is a standard case, corresponding to part (i) of the Theorem in which the long-run player is impatient and does not find it worthwhile to give up c for a future gain of $1 - c$. In this case effort will be provided only occasionally during beneficial events when the good type provides effort for non-reputational reasons.

[4]There is also a fourth case: if $c < \delta/(1+\delta(1-\pi))$ and there are "enough" normal types then there are at least two mixed strategy equilibria. A discussion of this result can be found in the online appendix to Levine (2021b) reproduced as the Appendix to this Chapter.

When $c < \delta$ it is worth it to maintain a reputation when the short-run player enters as indeed in this case $p = 1$. The remaining question is whether it is also worth it to provide effort when the short-run player does not enter. In this case $p = \pi$, and the condition for effort is that given in (iii). If c is very small then it is worth providing effort even when the short-run player does not enter. This good equilibrium corresponds to the "usual" reputational case, for example in Kreps and Wilson (1982), Milgrom and Roberts (1982), Fudenberg and Levine (1989), Fudenberg and Levine (1992) or Mailath and Samuelson (2001). There the long-run player is always is willing to provide effort over the relevant horizon.[5] Here, as in Mailath and Samuelson (2001), occasionally an adverse event occurs and the bad type does not provide effort regardless of reputational consequences so there is no effort until another normal or good type arrives.

The new and the interesting case is the trap equilibrium in case (ii) where $\delta > c$ so the cost of effort is low enough to maintain a reputation, but $c > \delta\pi/(1-\delta+\delta\pi)$ so it is not worth it to try to acquire a reputation. Here we have strong history dependence. Depending on the history a normal type will be in one of two very different situations. A normal type that follows a history of good signals, will provide effort, have a good reputation and have a wealthy and satisfactory life with an income of $1-c$. A normal type that has the ill-luck to follow a history in which the last signal was bad or there was no signal will not provide effort, will have a (deservedly) bad reputation, and have an impoverished life with an income of 0. This is a reputational trap. The only difference between these normal types is an event that took place in the far distant past: did the last behavioral type correspond to an adverse or beneficial event? Looked at another way, adverse and beneficial events, rare as they are, cast a very long shadow. After a beneficial event there will be many lives of prosperous normal types - indeed until an adverse event occurs. Contrariwise, following an adverse event normal types will be mired in the reputation trap until they are fortunate enough to have a beneficial event.

Observe that $\delta\pi/(1-\delta+\delta\pi)$ is increasing in π so as π increases and news spreads quickly the range of costs for the reputation trap diminishes and we are more likely to see the "usual" good reputation case. More important, although we will defer discussion of mixed strategies, is the condition

$$\delta > c > \delta \max\left\{\frac{\pi}{1-\delta+\delta\pi}, \frac{1}{1+\delta(1-\pi)}\right\}$$

in which the trap equilibrium is the only equilibrium: that is, in this case not only does the pure strategy equilibrium constitute a trap but there is no other equilibrium. Here the crucial fact is that both $\pi/(1-\delta+\delta\pi)$ and $1/(1+\delta(1-\pi)$ are both strictly less than one, so there is always a range of costs c in which the trap is the unique equilibrium.

15.4. Outline of the Proof: Pure Strategies

The proof of the main theorem involves the interplay between the strategy of the long-run player and the beliefs of the short-run player. First, regardless of the strategy of the short-run player the long-run player must provide effort when entry

[5]In models without type replacement eventually effort stops and the equilibrium collapses permanently into a no effort trap. Mailath and Samuelson (2001) show that with type replacement there is always effort.

is is anticipated if she is willing to do so when entry is not anticipated. In addition that unless the short-run player enters on the good signal and stays out on the bad signal the long-run player should never provide effort. This information can be used to rule out many combinations of long-run and short-run player strategies.

The next series of steps are to characterize the ergodic beliefs of the short-run player about the long-run player. These beliefs do not depend on ϵ and are bounded away from zero. The key to showing that the unique equilibrium strategy of the short-run player is to enter only on a good signal is to characterize these beliefs conditional on the signal. Let $\overline{B}$ be the probability of effort that makes the short-run player indifferent to entering, that is, $\overline{B}V = (1 - \overline{B})$. Recall that $\mu^1(z)$ is the ergodic belief of the short-run player about the probability that the long-run player will provide effort. If $\mu^1(z) > \overline{B}$ it is strictly optimal to enter, and if it is less than this, strictly optimal to stay out. If we can show that

$$\mu^1(1) \geq 1 - K\frac{\epsilon}{\min\{\pi, 1-\pi\}}$$

and

$$\mu^1(0), \mu^1(N) \leq K\frac{\epsilon}{\min\{\pi, 1-\pi\}}$$

for some positive constant K depending only on Q then it follows that for

$$K\frac{\epsilon}{\min\{\pi, 1-\pi\}} < \min\{\overline{B}, 1-\overline{B}\}$$

it is strictly optimal for the short-run player to stay out on a bad or no signal and to enter on a good signal. This then gives the main theorem with $\underline{\epsilon} = \min\{\overline{B}, 1-\overline{B}\}/K$.

The derivation of the bounds requires several steps. First, to a good approximation the beliefs of the short-run player about the type of long-run player are the same at the beginning of a period where the type may have changed as they were at the end of the previous period. This enables us to compute approximate conditional beliefs about types and signals from the simpler problem in which types are persistent. We then want to apply Bayes law to compute the probability of types conditional on signals. To implement this we need to know a lower bound on the marginal probability of the signals: in the case of the good and bad signal this follows from the fact that the good and bad types are playing the good and bad action; this is shown using ergodic calculations. Next, we apply Bayes law for the special case in which the long-run player takes an action independent of signal (as is the case for the behavioral types).

At this point things have been narrowed down to three possible strategies for the long-run player and eight for the short-run. It is now possible to check each of the twenty four combinations to find the ergodic beliefs and show that the only best response for the short-run player to a best response of the long-run player is to enter on a good signal and stay out for all others. Fortunately many combinations can be checked at once.

Finally, now that we know the unique strategy of the short-run player, we must calculate the best response of the long-run player: this is the computation with Γ above.

15.5. Intuition of the Main Result: Mixed Strategies

The important result is that there is a range of c for which there is a reputation trap and also no other equilibria. Why must this be the case? The reason is that the equilibrium short-run player pure strategy of staying out on a bad or no signal $z \in \{0, N\}$and entering on a good signal $z = 1$ provides the greatest incentive for the normal type to provide effort. If $c > \delta$ this is not enough, so weakening the incentive to provide effort by mixing does not help and the only equilibrium is the one in which the normal type never provides effort.

In the Levine (2021b) it is shown that if the short-run player uses a pure strategy the long-run player must do so as well. To understand why the short-run player strategy must remain pure even for $c < \delta$ (but not too small) consider that at $c = \delta$ the normal type strictly prefers to not to provide effort on a bad or no signal and is indifferent to effort on a good signal. When c is lowered slightly the normal type now strictly prefers to provide effort on a good signal, while of course the strict preference on bad and no signals remain. Can there be an equilibrium in which the short-run player mixes only "a little?" That cannot happen on a bad or no signal since to get the short-run player to mix the normal type would have to mix "a lot" and this in turn would require the short-run player to mix "a lot."

What about the good signal? Here with c a little less than δ "a little" mixing by the short-run player gets the normal type back to indifference. Without types this can be an equilibrium - but not with types. The reason is tied to the ergodic distribution of types and signals. With the normal type providing no effort on a bad or no signal once those states are reached the normal type will no longer get the good signal. With the short-run player mixing on the good signal there is a positive probability that the normal type will get no signal: this "drains" the normal types from the good signal so that in the ergodic distribution of types and signals conditional on a good signal it is extremely likely the short-run player is facing a good type. Consequently, the short-run player will not mix on a good signal - rather the short-run player will enter for certain.

The conclusion is that mixed strategy equilibria require the short-run player to mix "a lot." Formally in any mixed equilibrium the short-run player must be at least as likely to enter on no signal as on a good signal. This provides substantially less incentive for the normal type to provide effort than the short-run player equilibrium pure strategy in which the short-run player is a lot less likely to enter on no signal than on a good signal. Hence the value of c that is low enough to provide adequate incentive for effort is higher for a pure strategy equilibrium than for any mixed strategy equilibrium.

15.6. The Olympic Effect

If this reputation trap is real is it possible to get out of it? For example, if Southern Italy is caught in a reputation trap, what might the central government of Italy or the EU do to help? One possibility is to subsidize the cost of investment: if the cost c is low enough then investment even with the bad signal will be profitable and - eventually - the trap will be escaped. Welfare analysis of the model, however, indicates that this is probably not a good idea. The long-run player already has the possibility of making the investment and finds it not worth while; if the money designated for an investment subsidy was instead given to the long-run player the long-run player would choose not to spend it on investment - and would be strictly better off.

The model, however, points to another possible direction: if π could be increased it would be much easier to escape the reputation trap. Here an outside agency might have an advantage over the long-run agent having, perhaps, greater influence on outsiders and information flow to outsiders. Large mega-sporting events such as a World Cup or the Olympics come to mind in this context. By bringing large numbers of outsiders a cultural change is publicized. Bearing in mind that these events are awarded many years before they take place there is increased incentive for institutional change. One reason cities and regions compete for these events is precisely in hopes of obtaining favorable publicity. We need to ask, however, has this ever worked as a means of escaping a reputation trap? Certainly to be effective the investment must actually take place – hence the Olympics in Athens in 2004 or in Rio in 2016 simply confirmed what everybody already believed about those cities. In this context it must also be emphasized that to be effective the increase in π must be large enough – it must cross the threshold for which it becomes profitable to invest on the bad signal.

In the case of mega-sporting events there is an empirical literature and there are positive examples. Not all of this literature is relevant: much of it focuses on narrow issues such as tourism and local tax revenue, and venues with good reputations (where there should be little or no effect) are lumped in with venues with bad reputations (where there might be an effect). For a good overview of this literature see Matheson (2006). Strikingly, there is evidence from Rose and Spiegel (2011) that mega-sporting events when they are combined with institutional change have a substantial effect on international trade. Examples include the Olympics awarded to China in 2001 combined with entering the WTO, the Olympics awarded to Italy in 1955 combined with a series of reforms culminating in joining the European Economic Community, the Japanese Olympics of 1964 combined with entry into the IMF and the OECD, the Olympics awarded to Spain in 1986 combined with entering the European Economic Community, the Korean Olympics of 1988 Games together with political liberalization, and the Mexican 1986 FIFA World Cup combined with entry into GATT. Two other non-sporting events that may have had a similar impact (but have not been studied empirically) are the World Exposition in Chicago in 1893 and the 1997 opening of the Guggenheim Museum in Bilbao giving rise to a revival of that city called in the popular press the "Bilbao effect.[6]"

15.7. Appendix: Proofs

15.7.1. Problem of the Long Run Player. We examine the problem of the normal type of long-run player. Recall the Bellman equation

$$V(\alpha_2) = \max_{a_1}(1-\delta)\left[\alpha_2 - ca_1\right] + \delta \sum_{z'} P(z'|z, a_1) V(\alpha_2(z')).$$

We may write this out as

$$V(\alpha_2) =$$
$$\max_{a_1}(1-\delta)\left[\alpha_2 - ca_1\right] + \delta\left[\left(\alpha_2 + (1-\alpha_2)\pi\right) V(\alpha_2(a_1)) + (1-\alpha_2)(1-\pi)V(\alpha_2(N))\right].$$

[6]See for example "The Bilbao effect: how Frank Gehry's Guggenheim started a global craze," *The Guardian*, October 1.

LEMMA 15.7.1. *The optimum for the normal type of long run-player depends on the state only through* α_2 *and one of three cases applies:*

(i) $V(\alpha_2(1)) - V(\alpha_2(0)) < c(1-\delta)/\delta$*: it is strictly optimal provide no effort in every state. In particular if* $\alpha_2(1) = \alpha_2(0)$ *this is the case.*

(ii) $V(\alpha_2(1)) - V(\alpha_2(0)) > c(1-\delta)/(\delta\pi)$*: it is strictly optimal to provide effort in every state*

Defining

$$\tilde{\alpha}_2 = \frac{1-\delta}{\delta(1-\pi)\left(V(\alpha_2(1)) - V(\alpha_2(0))\right)}c - \frac{\pi}{1-\pi}$$

(iii) it is strictly optimal to provide effort if $\alpha_2(z) > \tilde{\alpha}_2$ *and conversely. In particular the strategy* $\alpha_1(0) > \alpha_1(1)$ *is never optimal.*

In addition

(iv) if $\alpha_2(0) = 1$ *then it is strictly optimal provide no effort in every state.*

Finally, if the short-run player uses a pure strategy then the optimum of the long-run player is strict and pure.

PROOF. The argmax is derived from:

$$\max_{a_1} -(1-\delta)ca_1 + \delta\left(\alpha_2 + (1-\alpha_2)\pi\right)V(\alpha_2(a_1)).$$

The gain to no effort is

$$G(\alpha_2) = (1-\delta)c - \delta\left(\alpha_2 + (1-\alpha_2\pi)\right)\left[V(\alpha_2(1)) - V(\alpha_2(0))\right].$$

We then solve this equation form α_2 to see when effort is and is not optimal.

Turning to the details, it follows that no effort is strictly optimal if

$$V(\alpha_2(1)) - V(\alpha_2(0)) < \frac{1-\delta}{\delta\left(\alpha_2 + (1-\alpha_2)\pi\right)}c$$

and conversely.

The RHS is strictly decreasing function of α_2. The value $\tilde{\alpha}_2$ is the unique value for which the two sides are equal, so the results (i) to (iii) follow directly.

For part (iv) suppose that $\alpha_2(0) = 1$. Then in state 0 choosing $\alpha_1 = 0$ gives 1 in very period so $V(0) = 1$. Since that is the greatest possible one period payoff $V(\alpha_2(1)) \leq 1 = V(\alpha_2(0))$ so the result follows from (i).

Finally, we analyze best response of the long-run player when the short-run player uses a pure strategy. From (i) and (iv) if $\alpha_2(0) \geq \alpha_2(1)$ it is strictly best not to provide effort. That leaves only the case $\alpha_2(a_1) = a_1$, or rather two cases, depending on $\alpha_2(N)$. This is a matter of solving the Bellman equations for each case to determine the value of c (if any) there can be a tie. This are the "non-generic" values listed in the text.

Turning to the details, If the response is not strict the condition for the gain to no effort must be zero

$$V(\alpha_2(1)) - V(\alpha_2(0)) = \frac{1-\delta}{\delta\left(a_2 + (1-a_2)\pi\right)}c.$$

Observe this cannot be the case at both states a_2.

(a) The tie is for $a_2 = 1$

In this case we have

$$V(\alpha_2(1)) = V(\alpha_2(0)) + \frac{1-\delta}{\delta}c$$

Moreover since $a_1 = 0$ must solve the Bellman equation for $a_2 = 1$ we have $V(\alpha_2(1)) = (1-\delta) + \delta V(\alpha_2(0))$. Solving we find $V(\alpha_2(0)) = 1 - c/\delta$.

Since $a_1 = 0$ is optimal at $a_2 = 1$ it must be that $a_1 = 0$ is strictly optimal at $a_1 = 0$. Hence

$$V(\alpha_2(0)) = \delta\left[\pi V(\alpha_2(0)) + (1-\pi)V(\alpha_2(N))\right].$$

There are two sub-cases depending on whether $\alpha_2(N) = 0, 1$.

If $\alpha_2(N) = 0$ then $V(\alpha_2(0)) = \delta V(\alpha_2(0))$ implies $V(\alpha_2(0)) = 0$. Since we previously found $V(\alpha_2(0)) = 1 - c/\delta$ this implies that $c = 1/\delta$ which is ruled out by generic cost.

If $\alpha_2(N) = 1$ then we have

$$V(\alpha_2(0)) = \delta\left[\pi V(\alpha_2(0)) + (1-\pi)V(\alpha_2(0)) + (1-\pi)\frac{1-\delta}{\delta}c\right]$$

which we solve to find

$$V(\alpha_2(0)) = (1-\pi)c/\delta.$$

Again this must also be equal to $1 - c/\delta$ so we have $(1-\pi)c/\delta = 1 - c/\delta$ or $c = \delta/(2-\pi)$ also ruled out by generic cost.

(b) The tie is for $a_2 = 0$

In this case we have

$$V(\alpha_2(1)) = V(\alpha_2(0)) + \frac{1-\delta}{\delta\pi}c.$$

Moreover since $a_1 = 1$ is optimal for $a_2 = 0$ it must also solve the Bellman equation for $a_2 = 1$, that is,

$$V(\alpha_2(1)) = (1-\delta)(1-c) + \delta V(\alpha_2(1))$$

so that $V(1) = 1 - c$. Hence

$$V(\alpha_2(0)) + \frac{1-\delta}{\delta\pi}c = 1 - c,$$

or

$$V(\alpha_2(0)) = 1 - c - \frac{1-\delta}{\delta\pi}c.$$

Again, there are two sub-cases depending on whether $\alpha_2(N) = 0, 1$.

If $\alpha_2(N) = 0$ then again $V(\alpha_2(0)) = \delta V(\alpha_2(0))$ implies $V(\alpha_2(0)) = 0$, giving

$$c\left[\frac{1-\delta+\delta\pi}{\delta\pi}\right] = 1$$

which is ruled out by generic cost.

If $\alpha_2(N) = 1$ since $a_1 = 0$ is optimal at $a_2 = 0$ wand $V(\alpha_2(1)) = 1 - c$

$$V(\alpha_2(0)) = \delta\left[\pi V(0) + (1-\pi)(1-c)\right]$$

or

$$V(\alpha_2(0)) = \frac{1-\pi}{1-\delta\pi}(1-c).$$

This must be equal to

$$1 - c - \frac{1-\delta}{\delta\pi}c$$

and equating the two we find

$$\begin{aligned}
1-c-\frac{1-\delta}{\delta\pi}c &= \frac{1-\pi}{1-\delta\pi}(1-c)\\
c+\frac{1-\delta}{\delta\pi}c-\frac{1-\pi}{1-\delta\pi}c &= 1-\frac{1-\pi}{1-\delta\pi}\\
\frac{1-\delta}{\delta\pi}c+\frac{\pi-\delta\pi}{1-\delta\pi}c &= \frac{\pi-\delta\pi}{1-\delta\pi}\\
(1-\delta\pi)(1-\delta)c+\delta\pi(\pi-\delta\pi)c &= \delta\pi(\pi-\delta\pi)\\
c &= \frac{\delta\pi(\pi-\delta\pi)}{(1-\delta\pi)(1-\delta)+\delta\pi(\pi-\delta\pi)}
\end{aligned}$$

ruled out by the generic cost assumption. □

15.7.2. Ergodic Beliefs of the Short-Run Player. Next we examine the beliefs of the short-run player. For given pure strategies of both players the signal type pairs (z,τ) are a Markov chain with transition probabilities independent of δ and depending only on ϵ,π and the strategies of the two players. Excluding the state N in case the short-run player always enters the chain is irreducible and aperiodic so it has a unique ergodic distribution $\mu_{z\tau}$. We first analyze the marginals μ_τ and μ_z.

LEMMA 15.7.2. *The marginals μ_τ are independent of ϵ. Let $\underline{\mu}=\min_{\tau\neq n}\mu_\tau$. Then $\underline{\mu}>0$, $\mu_0,\mu_1\geq\pi\underline{\mu}$, if $\alpha_2(0)=\alpha_2(1)=1$ then $\mu_N=0$, otherwise if the short-run player plays a pure strategy then $\mu_N\geq(1-\pi)\underline{\mu}$.*

PROOF. The type transitions are independent of the signals, so we analyze those first. For $\epsilon>0$ we have $\mu_\tau>0$ since every type transition has positive probability. This ergodic distribution is the unique fixed point of the 3×3 transition matrix A, which is to say given by the intersection of the null space of $I-A$ with the unit simplex. Since $A=I+Q\epsilon$ it follows that it is given by the intersection of the null space of $Q\epsilon$ with the unit simplex. As the null space of $Q\epsilon$ is independent of ϵ the marginals μ_τ are independent of ϵ as well.

For the signals we have $\mu_1\geq\pi\mu_g$ and $\mu_0\geq\pi\mu_b$. If if $a_2(0)=a_2(1)=1$ then the state N is transient. If $\alpha_2(1)=0$ then $\mu_N\geq(1-\pi)\mu_g$ while if $\alpha_2(0)=0$ then $\mu_N\geq(1-\pi)\mu_b$. □

It will be convenient to normalize so that $\max(\mu_\sigma/\mu_\tau)Q_{\tau\sigma}=1$. Next we show how the conditional probabilities $\mu_{z|\tau}$ can be computed approximately by using the ergodic conditions for $\epsilon=0$.

LEMMA 15.7.3. *When $z=N$*

$$\mu_{N|\tau}=(1-\pi)\left(\sum_y(1-\alpha_2(y))\mu_{y|\tau}+\epsilon H_{N\tau}\right)$$

when $z\neq N$

$$\mu_{z|\tau}=$$

$$\sum_y 1\left((z=1)\alpha_1(\tau,y)+1(z=0)(1-\alpha_1(\tau,y))\right)\left[\alpha_2(y)+\pi(1-\alpha_2(y))\right]\mu_{y|\tau}+\epsilon H_{z\tau}.$$

where $|H_{z\tau}|\leq 2$ for all z.

PROOF. The idea is that the process for types is exogenous, so the stationary probabilities can be computed directly. This enables us to find a linear recursive relationship for the conditionals where the coefficients depend upon the strategies and the (already known) marginals over types. We then show that when ϵ is small to a good approximation we can do the computation for $\epsilon = 0$, that is, ignoring the type transitions, with the result above showing how good the approximation is for given ϵ.

For given strategies of the players define $P(z, \sigma|y, \tau)$ to be the conditional probability that $z_{t+1} = z, \sigma_{t+1} = \sigma$ conditional on $z_t = y, \tau_t = \tau$. We have

$$\mu_{z\tau} = \mu_{z|\tau}\mu_\tau = \sum_\sigma \sum_y P(z|y,\sigma)P(\tau|\sigma)\mu_{y\sigma}$$

$$= \sum_\sigma \sum_y P(z|y,\sigma)\Pr(\tau|\sigma)\mu_{y|\sigma}\mu_\sigma = \sum_\sigma P(\tau|\sigma)\mu_\sigma \sum_y P(z|y,\sigma)\mu_{y|\sigma}.$$

Since we know that $\mu_\tau > 0$ we may divide to find

$$\mu_{z|\tau} = \sum_\sigma P(\tau|\sigma)\frac{\mu_\sigma}{\mu_\tau} \sum_y P(z|y,\sigma)\mu_{y|\sigma}$$

$$= \sum_\sigma P(\tau|\sigma)\frac{\mu_\sigma}{\mu_\tau} \sum_y P(z|\alpha_2(y), \alpha_1(\sigma, y))\mu_{y|\sigma}.$$

Define $\overline{h}(\tau|\tau) = -\sum_{\sigma\neq\tau} Q_{\tau\sigma} = \left(P(\tau|\tau) - 1\right)/\epsilon$ and for $\tau \neq \sigma$ define $\overline{h}(\tau|\sigma)\epsilon = (\mu_\sigma/\mu_\tau)Q_{\sigma\tau} = P(\tau|\sigma)/\epsilon$. Observe that $\overline{h}$ depends only on Q and that

$$|\overline{h}(\tau|\sigma)| \leq \max\{2(\mu_\sigma/\mu_\tau)Q_{\tau\sigma}|\tau \neq \sigma\} = 2.$$

Then

$$\mu_{z|\tau} = \sum_y P(z|\alpha_2(y), \alpha_1(\sigma,\tau))\mu_{y|\tau} + \epsilon\sum_\sigma \overline{h}(\tau|\sigma)\sum_y P(z|\alpha_2(y), \alpha_1(\sigma, y))\mu_{y|\sigma}.$$

For $z = N$ this is

$$\mu_{N|\tau} = \sum_y (1-\pi)(1-\alpha_2(y))\mu_{y|\tau} + \epsilon\sum_\sigma \overline{h}(\tau|\sigma)\sum_y (1-\pi)(1-\alpha_2(y))\mu_{y|\sigma}$$

$$= (1-\pi)\left(\sum_y (1-\alpha_2(y))\mu_{y|\tau} + \epsilon H_{N\tau}\right).$$

For $z \neq N$ this is

$$\mu_{z|\tau} = \sum_y P(z|\alpha_2(y), \alpha_1(\sigma,\tau))\mu_{y|\tau} + \epsilon\sum_\sigma \overline{h}(\tau|\sigma)\sum_y P(z|\alpha_2(y), \alpha_1(\sigma, y))\mu_{y|\sigma}$$

$$\sum_y \left(1(z=1)\alpha_1(\tau, y) + 1(z=0)(1-\alpha_1(\tau,y))\right)\left[\alpha_2(y) + \pi(1-\alpha_2(y))\right]\mu_{y|\tau} + \epsilon H_{z\tau}.$$

In both cases $|H_{z\tau}| \leq 2$. □

To apply Bayes Law we will need to bound marginal probabilities of signals from below. The hard case is that of no signal where we must solve the equations for the conditionals simultaneously. Here we analyze the short-run pure strategy case. If the short-run player enters for both $z = 0, 1$ then no signals are unlikely as they are generated only from type transitions, so we rule that out.

LEMMA 15.7.4. *Suppose $\alpha_2(a_1) = 0$ for some $a_1 \in \{0, 1\}$. Then*

$$\mu_N \geq \frac{1-\pi}{2}\left(1 - \frac{4\epsilon}{\pi}\right)\underline{\mu}.$$

PROOF. Let τ be the type that plays a_1. We have

$$\mu_{a_1|\tau} = \sum_y \left[\alpha_2(y) + \pi(1-\alpha_2(y))\right]\mu_{y|\tau} + \epsilon H_{a_1\tau}$$

$$\mu_{N|\tau} = (1-\pi)\left(\sum_y (1-\alpha_2(y))\mu_{y|\tau} + \epsilon H_{N\tau}\right)$$

These imply the inequalities

$$\mu_{a_1|\tau} \geq \pi(1-\mu_{N|\tau}) + \left[\alpha_2(N) + \pi(1-\alpha_2(N))\right]\mu_{N|\tau} + \epsilon H_{a_1\tau}$$
$$\mu_{N|\tau} \geq (1-\pi)\left((1-\alpha_2(N))\mu_{N|\tau} + \mu_{a_1|\tau} + \epsilon H_{N\tau}\right).$$

Hence

$$\begin{aligned} &\mu_{N|\tau} \geq (1-\pi)\times \\ &\left((1-\alpha_2(N))\mu_{N|\tau} + \pi(1-\mu_{N|\tau}) + \left[\alpha_2(N) + \pi(1-\alpha_2(N))\right]\mu_{N|\tau} + \epsilon H_{N\tau} + \epsilon H_{a_1\tau}\right) \\ &= (1-\pi)\left(\pi + \left[\alpha_2(N) + (1+\pi)(1-\alpha_2(N)) - \pi\right]\mu_{N|\tau} + \epsilon H_{N\tau} + \epsilon H_{a_1\tau}\right) \\ &\geq (1-\pi)\left(\pi + (1-\pi)\mu_{N|\tau} + \epsilon H_{N\tau} + \epsilon H_{a_1\tau}\right). \end{aligned}$$

It follows that

$$\mu_{N|\tau} \geq \frac{1-\pi}{1-(1-\pi)^2}(\pi + 4\epsilon) = \frac{1-\pi}{(2-\pi)\pi}(\pi + 4\epsilon)$$

$$\geq \frac{1-\pi}{2-\pi}\left(1 - \frac{4\epsilon}{\pi}\right) \geq \frac{1-\pi}{2}\left(1 - \frac{4\epsilon}{\pi}\right).$$

The result now follows from $\mu_N \geq \mu_{N|\tau}\mu_\tau \geq \mu_{N|\tau}\underline{\mu}$. □

Finally we compute bounds on beliefs about types that play the same action independent of the signal. Here we combine bounds from the equations for the conditionals with Bayes Law.

LEMMA 15.7.5. *A long-run type τ that plays the pure action a_1 regardless of the signal has*

$$\mu_{\tau|-a_1} \leq \frac{2}{\underline{\mu}}\left(\frac{\epsilon}{\pi}\right)$$

and if $\alpha_2(1) = 1$ and $\alpha_2(0) = 0$ then a type τ that plays the action 1 regardless of signal has

$$\mu_{\tau|N} \leq \frac{8}{\left(1 - 4\left(\frac{\epsilon}{\pi}\right)\right)\underline{\mu}}\left(\frac{\epsilon}{\pi}\right).$$

PROOF. If long-run type τ plays the pure action a_1 from Lemma 15.7.3

$$\begin{aligned} &\mu_{-a_1|\tau} = \\ &\left(1(a_1=0)1(a_1=1) + 1(a_1=1)1(a_1=0)\right)\sum_y\right) \left[\alpha_2(y) + \pi(1-\alpha_2(y))\right]\mu_{y|\tau} + \epsilon H_{z\tau}. \\ &= \epsilon H_{-a_1\tau} \leq 2\epsilon. \end{aligned}$$

From Lemma 15.7.2 $\mu_{-a_1} \geq \pi \underline{\mu}$ and Bayes law then implies

$$\mu_{\tau|-a_1} \leq \frac{\epsilon 2}{\pi \underline{\mu}}.$$

For the second part we have from Lemma 15.7.3

$$\mu_{N|\tau} = (1-\pi)\left(\sum_y (1-\alpha_2(y))\mu_{y|\tau} + \epsilon H_{N\tau}\right)$$

$$\mu_{0|\tau} = \epsilon H_{0\tau}.$$

Hence

$$\mu_{N|\tau} = (1-\pi)\left(\mu_{0|\tau} + [1-\alpha_2(N)]\,\mu_{N|\tau}\right) + (1-\pi)\epsilon H_{N\tau}.$$

Plugging in

$$\mu_{N|\tau} = (1-\pi)\,[1-\alpha_2(N)]\,\mu_{N|\tau} + (1-\pi)\epsilon H_{0\tau} + (1-\pi)\epsilon H_{N\tau}$$
$$\mu_{N|\tau} \leq (1-\pi)\mu_{N|\tau} + (1-\pi)\epsilon H_{0\tau} + (1-\pi)\epsilon H_{N\tau}$$

so

$$\mu_{N|\tau} \leq \frac{(1-\pi)4\epsilon}{\pi}.$$

From Lemma 15.7.4

$$\mu_N \geq \frac{1-\pi}{2}\left(1-\frac{4\epsilon}{\pi}\right)\underline{\mu}.$$

Hence Bayes law implies

$$\mu_{\tau|N} \leq \frac{8\epsilon}{\pi\left(1-\frac{4\epsilon}{\pi}\right)\underline{\mu}}.$$

□

15.7.3. Short-Run Player Optimality. Recall that $\mu^1(z)$ is the probability of $a_1 = 1$ in state z and that $\overline{B} = 1/(V+1)$ is the critical value of $\mu^1(z)$ such that

Lemma 15.7.6. *If $\mu^1(z) > \overline{B}$ the short-run player strictly prefers to enter; if $\mu^1(z) < \overline{B}$ the short-run player strictly prefers to stay out, and if $\mu^1(z) = \overline{B}$ the short-run player is indifferent.*

We next show that it cannot be optimal for the short-run player always to enter. Set $B \equiv \underline{\mu}\min\{\pi, 1-\pi\}\min\{\overline{B}, 1-\overline{B}\}$.

Lemma 15.7.7. *For $\epsilon < (1/2)B$ always enter $a_2(z) = 1$ for all z is not an equilibrium.*

Proof. By Lemma 15.7.1 always enter implies no effort by the normal long-run player. As there are few good types at $z = 0$ we show that this forces the short-run player to stay out there so the short-run player should not in fact enter.

Turning to the details, Lemma 15.7.5 gives

$$\mu_{g|0} \leq \frac{2}{\underline{\mu}}\left(\frac{\epsilon}{\pi}\right).$$

Hence

$$\mu^1(0) \leq \frac{2}{\underline{\mu}}\left(\frac{\epsilon}{\pi}\right)$$

also. From Lemma 15.7.6 it follows that $\epsilon/\pi < \underline{\mu}\overline{B}/2$ implies $a_2(0) = 0$ a contradiction. □

LEMMA 15.7.8. *For $\epsilon < (1/16)B$ the strict equilibrium response to never provide effort is to enter only on $z = 1$ and do so with probability* 1.

PROOF. As the normal and bad types never provide effort the signal $z = 1$ implies a good type with high probability so the short-run player should enter there. This means that the long-run player can have the signal $z = 1, N$ only through a type transition. In particular the bad signal is dominated by normal and bad types so the short run player should stay out. This in turn means that most of the N signals are generated by normal and bad types, so the short-run player should stay out there too.

Turning to the details, from Lemma 15.7.5 no effort implies

$$\mu_{n|1} \leq \frac{2}{\underline{\mu}}\left(\frac{\epsilon}{\pi}\right)$$

and the same inequality holds for $\mu_{b|1}$. Hence

$$\mu_{g|1} \geq 1 - \frac{4}{\underline{\mu}}\left(\frac{\epsilon}{\pi}\right)$$

and by Lemma 15.7.6 $\epsilon/\pi < \underline{\mu}(1 - \overline{B})/4$ this forces $\alpha_2(1) = 1$. By the Lemma 15.7.5

$$\mu_{g|0} \leq \frac{2}{\underline{\mu}}\left(\frac{\epsilon}{\pi}\right)$$

so by Lemma 15.7.6 $\epsilon/\pi < \underline{\mu}\overline{B}/2$ we must have $\alpha_2(0) = 0$.

We may again apply the two Lemmas to conclude that

$$\mu_{g|N} \leq \frac{8}{\left(1 - 2\left(\frac{\epsilon}{\pi}\right)\right)\underline{\mu}}\left(\frac{\epsilon}{\pi}\right)$$

so that for

$$\epsilon/\pi < \max\left\{\underline{\mu}\overline{B}/16, 1/4\right\}$$

the short-run player must stay out on N as well.

All these responses are strict. □

LEMMA 15.7.9. *For $\epsilon < (1/16)B$ there is no equilibrium in which $\alpha_2(0) = 1$.*

PROOF. By Lemma 15.7.1 $\alpha_2(0) = 1$ implies never provide effort so by Lemma 15.7.8 $\alpha_2(0) = 0$ a contradiction. □

LEMMA 15.7.10. *For $\epsilon < (1/32)B$ the unique equilibrium response to always provide effort is to enter only on $z = 1$ and do so with probability* 1.

PROOF. This is basically the opposite of Lemma 15.7.8. Now at $z = 1$ there are mainly good and normal types so it is optimal for the short-run player to enter. While at $z = 0$ there are mainly bad types so it is optimal for the short-run player to stay out. Hence no-signal is generated by bad types from $z = 0$ so it is optimal for the short-run player to stay out there too.

Turning to the details, from Lemma 15.7.5

$$\mu_{g|0}, \mu_{n|0} \leq \frac{2}{\underline{\mu}}\left(\frac{\epsilon}{\pi}\right)$$

so

$$\mu_{b|0} \geq 1 - \frac{4}{\underline{\mu}}\left(\frac{\epsilon}{\pi}\right)$$

so by Lemma 15.7.6 $\epsilon/\pi < \underline{\mu}(1-\overline{B})/4$ implies $a_2(0) = 0$.

Apply the two Lemmas again to see that

$$\mu_{b|1} \leq \frac{2}{\underline{\mu}}\left(\frac{\epsilon}{\pi}\right)$$

so for $\epsilon/\pi < \underline{\mu}\overline{B}/2$ we have $a_2(1) = 1$.

Apply the two Lemmas a third time to see that

$$\mu_{g|N}, \mu_{n|N} \leq \frac{8}{\left(1 - 4\left(\frac{\epsilon}{\pi}\right)\right)\underline{\mu}}\left(\frac{\epsilon}{\pi}\right)$$

so that $\epsilon/\pi < \max\left\{\underline{\mu}\overline{B}/32, 1/8\right\}$ implying $a_2(N) = 0$.

All these responses are strict. $\square$

LEMMA 15.7.11. *If $\epsilon < (1/2)B$ and for some a_1 we have $\alpha_1(a_1) = a_1$ then $\alpha_2(a_1) = a_1$.*

PROOF. If $\alpha_1(0) = 0$ then from Lemmas 15.7.3 and 15.7.2 $\mu^1(0) = \mu_{0|g}\mu_g/\mu_0 = \epsilon H_{0g}\mu_g/\mu_0 \leq 2\epsilon/(\pi\underline{\mu})$. If $\alpha_1(1) = 1$ then $1 - \mu^1(1) = \mu_{1|b}\mu_b/\mu_1 = \epsilon H_{1b}\mu_b/\mu_1 \leq 2\epsilon/(\pi\underline{\mu})$. Hence for $\epsilon/\pi < \overline{B}\underline{\mu}/2$ it follows that $\alpha_2(a_1) = a_1$. $\square$

15.7.4. Uniqueness of Short-Run Pure Equilibria. We define an *equilibrium response* of the short-run player to a strategy of the long-run player to be a best response to $\mu_{z\tau}$ induced by the long-run player strategy and itself.

PROPOSITION 15.7.12. *There exists an $\underline{\epsilon} > 0$ depending only on V such that for any ϵ satisfying*

$$\underline{\epsilon} > \frac{\epsilon}{\underline{\mu}\min\{\pi, 1-\pi\}} > 0$$

in any short-run pure equilibrium the short-run player must enter on the good signal and only on the good signal. Moreover this is a strict equilibrium response.

PROOF. We rule out all other possibilities:

(a) A*lways enter $a_2(z) = 1$ for all z is not an equilibrium.* By Lemma 15.7.7

(b) *The unique equilibrium response to never provide effort is to enter only on* $z = 1$. From Lemma 15.7.7.

(c) *A equilibrium response requires $a_2(1) = 1, a_2(0) = 0$.* Any other strategy satisfies $a_2(0) \geq a_2(1)$. From Lemma 15.7.1 this implies no effort by the long-run player. Part (b) then forces $0 = a_2(0) < a_2(1) = 1$ a contradiction.

(d) *The unique equilibrium response to always provide effort is to enter only on* $z = 1$. From Lemma 15.7.10.

This leaves only the strategy $\tilde{a}$ in which the long-run player plays $a_1 = 1$ on entry and $a_1 = 0$ if the short-run player stays out. As we know that $\alpha_2(1) = 1, \alpha_2(0) = 0$ there are two possibilities $\alpha_2(N) = 1$ and $\alpha_2(N) = 0$. The former is ruled out because it leads to primarily bad types at $z = N$, and the latter is a strict best response by the short-run player because there are few good types at $z = N$.

Turning to the details, there is entry at $N, 1$ and not on 0 consequently there is effort on $N, 1$ and not on 0. From Lemma 15.7.3 we find:

$$\mu_{0|n} = \sum_y (1 - \alpha_1(n, y))\left[\alpha_2(y) + \pi(1 - \alpha_2(y))\right]\mu_{y|n} + \epsilon H_{0n} = \pi\mu_{0|n} + \epsilon H_{0n},$$

$$\mu_{N|n} = (1-\pi)\left(\sum_y (1-\alpha_2(y))\mu_{y|n} + \epsilon H_{Nn}\right) = (1-\pi)\mu_{0|n} + (1-\pi)\epsilon H_{Nn}.$$

The former implies

$$\mu_{0|n} \leq \frac{2\epsilon}{1-\pi}$$

so that the second implies

$$\mu_{N|n} \leq 2\epsilon + (1-\pi)\epsilon H_{Nn} \leq 4\epsilon.$$

From Lemma 15.7.4

$$\mu_N \geq \frac{1-\pi}{2}\left(1-\frac{4\epsilon}{\pi}\right)\underline{\mu}$$

so Bayes law gives

$$\mu_{n|N} \leq \frac{\pi}{1-\pi}\frac{8}{\left(1-4\left(\frac{\epsilon}{\pi}\right)\right)\underline{\mu}}\left(\frac{\epsilon}{\pi}\right).$$

Also by Lemma 15.7.5

$$\mu_{g|N} \leq \frac{8}{\left(1-4\left(\frac{\epsilon}{\pi}\right)\right)\underline{\mu}}\left(\frac{\epsilon}{\pi}\right).$$

Hence

$$\mu_{b|N} \geq 1 - \frac{16}{\left(1-4\left(\frac{\epsilon}{\pi}\right)\right)\underline{\mu}}\left(\frac{\epsilon}{1-\pi}\right)$$

from which the result follows. Note that it is only for this result that we require $\epsilon/(1-\pi)$ to be small as well as ϵ/π.

Finally we must show that $\alpha_2(N) = 0$ is in fact a strict equilibrium response for the short-run player. We have

$$\mu_{b|1}, \mu_{g|0} \leq \frac{2}{\underline{\mu}}\left(\frac{\epsilon}{\pi}\right)$$

$\mu_{b|1} = 0$ and $\mu_{g|0} = 0$ so is a strict best response to stay out in the former and enter in the latter. Finally Lemma 15.7.3 gives

$$\mu_{g|N} \leq \frac{8}{\left(1-4\left(\frac{\epsilon}{\pi}\right)\right)\underline{\mu}}\left(\frac{\epsilon}{\pi}\right)$$

implying for small ϵ/π it is strictly optimal for the short-run player to stay out on N. □

15.7.5. Mixing. Recall that all of the Lemmas concerning short-run optimality hold for $\epsilon \leq B/32$ (and the remaining Lemmas do not place restrictions on ϵ) where $B = \underline{\mu}\min\{\pi, 1-\pi\}\min\{\overline{B}, 1-\overline{B}\}$. Recall also the notion of a fundamental bound: it may depend on the fundamentals of the game π, V, δ, c but not on the type dynamics Q, ϵ. Define the fundamental bound $\overline{A} \equiv \pi^2(1-\pi)\min\{\overline{B}, 1-\overline{B}\}$ and observe that if $\epsilon \leq \underline{\mu}\overline{A}/32$ then also $\epsilon \leq B/32$. We shall assume $\epsilon \leq \underline{\mu}\overline{A}/32$ hereafter.

LEMMA 15.7.13. *There is no non-pure equilibrium with $\alpha_1(1) = 1$.*

PROOF. By Lemma 15.7.2 $\mu_{1|b} = \epsilon H_{1b} \le 2\epsilon$. Hence for $\epsilon < \overline{B}/2$ by Lemma 15.7.6 $\alpha_2(1) = 1$. Then by Lemma 15.7.2

$$\mu_{1|n} = \mu_{1|n} + \sum_{y \in \{0,N\}} \alpha_1(y)\left[\alpha_2(y) + \pi(1-\alpha_2(y))\right]\mu_{y|n} + \epsilon H_{z\tau}.$$

It follows that

$$\sum_{y \in \{0,N\}} \alpha_1(y)\mu_{y|n} \le 2(\epsilon/\pi) \text{ so } \max_{y \in \{0,N\}} \alpha_1(y)\mu_{y|n} \le 2(\epsilon/\pi).$$

Moreover for $z \in \{0, N\}$ we have $\mu_{z|g} = \epsilon H_{zg} \le 2\epsilon$. Hence

$$\mu^1(0) = \frac{\mu_{0|g}\mu_g + \alpha_1(0)\mu_{0|n}\mu_n}{\mu_0} \le 2(\epsilon/\pi)(\mu_g + \mu_n)/(\pi\underline{\mu}) \le 2(\epsilon/\pi)/(\pi\underline{\mu}).$$

So for $\epsilon/\pi^2 < \overline{B}\underline{\mu}/2$ (this is why π^2 appears in $\overline{A}$) by Lemma 15.7.6 we have $\alpha_2(0) = 0$. This implies by Lemma 15.7.4 that

$$\begin{aligned}\mu^1(N) = \frac{\mu_{N|g}\mu_g + \alpha_1(N)\mu_{N|n}\mu_n}{\mu_N} &\le 2(\epsilon/\pi)(\mu_g + \mu_n)/\mu_N \\ &\le \frac{8(\epsilon/\pi)}{(1-\pi)\left(1-\frac{4\epsilon}{\pi}\right)\underline{\mu}}.\end{aligned}$$

So when this is less than or equal $\overline{B}$ by Lemma 15.7.6 we have $\alpha_2(N) = 0$. For $\epsilon \le \overline{A}/8$ this is

$$\frac{16\epsilon}{\pi(1-\pi)\underline{\mu}} \le \overline{B}$$

so holds for $\epsilon < \underline{\mu}\overline{A}/16$ which was assumed. □

LEMMA 15.7.14. *In any equilibrium* $\alpha_1(0) = \alpha_2(0) = 0$.

PROOF. We already know this to be true in any pure equilibrium, so we may assume the equilibrium is not pure. From Lemma 15.7.11 if $\alpha_1(0) = 0$ then $\alpha_2(0) = 0$ so we may assume this is not the case, that is $\alpha_1(0) > 0$. From Lemma 15.7.13 we know that $\alpha_1(1) < 1$. It cannot be that the normal type is indifferent at both $z = 0, 1$ for then by Lemma 15.7.1 it must be that $\alpha_2(1) = \alpha_2(0) = \tilde{\alpha}_2$ so that $V_1 = V(\tilde{\alpha}) = V_0$ and that the normal type never provides effort in which case by Lemma 15.7.8 we would have a pure strategy equilibrium. Hence either the normal type strictly prefers to provide no effort at $z = 1$ and is willing to provide effort at $z = 0$ or the normal type is indifferent at $z = 1$ and strictly prefers to provide effort at $z = 0$. In either case from Lemma 15.7.1 we must have $\alpha_2(1) < \alpha_2(0)$.

The key point is that having the short-run player enter when there is no effort is kind of like winning the lottery - you get something for nothing. If that happens in the state 0 it is particularly good because you are guaranteed that you get to play again. Since $\alpha_2(1) < \alpha_2(0)$ we can write $\alpha_2(0) = \beta + (1-\beta)\alpha_2(1)$ where $\beta > 0$ meaning that in the state $z = 0$ there is a better chance of winning the lottery. We will use this to show that $V(\alpha_2(0)) \ge V(\alpha_2(1))$ so that never provide effort is optimal and the equilibrium must be pure by Lemma 15.7.8.

Specifically we compare $V(\alpha_2(0))$ to $V(\alpha_2(1))$. We may compute $V(\alpha_2(1))$ under the assumption that the normal type does not provide effort since this is optimal at $z = 1$. This gives

$$V(\alpha_1(1)) =$$

$$(1-\delta)\alpha_2(1) + \delta\left[\pi V(\alpha_2(0)) + (1-\pi)\left(\alpha_2(1)V(\alpha_2(0)) + (1-\alpha_2(1))V(\alpha_2(N))\right)\right]$$

We may compute a lower bound $V(\alpha_2(0))$ under the assumption that the normal type does not provide effort in the first period and optimizes afterwards. In this case

$$\begin{gathered} V(\alpha_2(0)) \geq (1-\delta)\alpha_2(0) \\ +\delta\left[\pi V(\alpha_2(0)) + (1-\pi)\left(\alpha_2(0)V(\alpha_2(0)) + (1-\alpha_2(0))V(\alpha_2(N))\right)\right] \\ = (1-\delta)\left(\beta + (1-\beta)\alpha_2(1)\right) \\ +\delta\left[\pi V(\alpha_2(0)) + (1-\pi)\times\right. \\ \left.\left((\beta+(1-\beta)\alpha_2(1)\right)V(\alpha_2(0)) + (1-(\beta+(1-\beta)\alpha_2(1)))V(\alpha_2(N)))\right] \\ = (1-\delta)\left(\beta + (1-\beta)\alpha_2(1)\right) \\ +\delta\left[\pi V(\alpha_2(0)) + (1-\pi)\times\right. \\ \left.\left((\beta+(1-\beta)\alpha_2(1)\right)V(\alpha_2(0)) + (1-\beta)\left(1-\alpha_2(1)\right)V(\alpha_2(N)))\right] \\ = (1-\delta)\beta + \delta\beta V(\alpha_2(0)) + (1-\beta)\alpha_2(1) \\ +\delta\left[(1-\beta)\pi V(\alpha_2(0)) + (1-\pi)\times\right. \\ \left.\left((1-\beta)\alpha_2(1)\right)V(\alpha_2(0)) + (1-\beta)\left(1-\alpha_2(1)\right)V(\alpha_2(N))\right] \end{gathered}$$

Using the expression for $V(\alpha_1(1))$ from above this gives

$$V(\alpha_2(0)) \geq (1-\delta)\beta + \delta\beta V(\alpha_2(0)) + (1-\beta)V(\alpha_1(1)).$$

Hence

$$V(\alpha_2(0)) \geq \frac{(1-\delta)\beta}{1-\delta\beta} + \frac{1-\beta}{1-\delta\beta}V(\alpha_1(1)),$$

Since $V(\alpha_1(1)) \leq 1$ this then implies $V(\alpha_1(0)) \geq V(\alpha_1(1))$ as advertised, $\square$

LEMMA 15.7.15. *In any non-pure equilibrium* $0 < \alpha_2(1) < 1$, $\alpha_1(N) > 0$, *and* $\alpha_2(N) \geq \alpha_2(1)$.

PROOF. First suppose that $\alpha_2(1) = 1$. Since the short-run player must be mixing and by Lemma 15.7.14 is not doing so at $z = 0$ the short-run player must be mixing at $z = N$, that is, that $0 < \alpha_2(N) < 1$. Lemma 15.7.13 implies that at $z = 1$ the normal type does not strictly prefer to provide effort. Since $\alpha_2(N) < \alpha_2(1)$ Lemma 15.7.1 implies that at $z = N$ normal type strictly prefers to provide no effort, so $\alpha_1(N) = 0$. Hence $\mu^1(N) = \mu_{N|g}\mu_g/\mu_N = \epsilon H_{0g}\mu_g/\mu_N$. As $\alpha_2(0) = 0$ by Lemma 15.7.14 it follows from Lemma 15.7.4 that

$$\mu^1(N) \leq \frac{4\epsilon}{(1-\pi)\left(1-\frac{4\epsilon}{\pi}\right)\underline{\mu}}$$

as the RHS this is less than $\overline{B}$ by assumption we have $\alpha_2(N) = 0$ a contradiction.

Next suppose that $\alpha_2(1) = 0$. By Lemma 15.7.14 we also have $\alpha_2(0) = 0$ so by Lemma 15.7.1 the long run player never provides effort. Hence $\alpha_2(1) > 0$ follows from Lemma 15.7.8, a contradiction. We have now shown strict mixing the the short-run player at $z = 1$.

Now we show that since the short-run player is strictly mixing at $z = 1$ then $\alpha_1(N) > 0$. Strict mixing by the short-run player at $z = 1$ implies from Lemma 15.7.6 $1 - \overline{B} = 1 - \mu^1(1) = \left([1-\alpha_1(1)]\mu_{1|n}\mu_n + \mu_{1|b}\mu_b\right)/\mu_1$. From Lemma 15.7.3 and Lemma 15.7.14 if $\alpha_1(N) = 0$ we have $\mu_{1|n} \leq \alpha_1(1)\mu_{1|n} + 2\epsilon$ and $\mu_{1|b} \leq 2\epsilon$. Hence by Lemma 15.7.2 $1 - \mu^1(1) \leq 2\epsilon/(\pi\underline{\mu})$, so for $2\epsilon/(\pi\underline{\mu}) < 1 - \overline{B}$ this is a contradiction.

Since $\alpha_2(N) > 0$ the normal type weakly prefers to provide effort at $z = N$. If $\alpha_2(1) > \alpha_2(N)$ by Lemma 15.7.1 this implies the normal type would strictly prefer to provide effort at $z = 1$ contradicting Lemma 15.7.13. $\square$

15.7.6. Signal Jamming. Define the *auxiliary system* with respect to $0 \leq \lambda, \gamma \leq 1$ as

$$V_1 = (1-\delta)\tilde{\alpha}_2 + \delta\left[(\tilde{\alpha}_2 + (1-\tilde{\alpha}_2)\pi)\, V_0 + (1-\tilde{\alpha}_2)(1-\pi)V_N\right]$$

$$V_N = (1-\gamma)(\lambda - c) + \gamma V_1$$

$$V_0 = \frac{\delta(1-\pi)}{1-\delta\pi} V_N.$$

Since in a mixed equilibrium we know from Lemma 15.7.13 that $\alpha_1(1) < 1$ so that at $z = 1$ the long-run player must be willing not to provide effort. This system corresponds to providing no effort at $z = 0, 1$. From the contraction mapping fixed point theorem this has a unique solution V_1, V_N, V_0. Define the function $\underline{\Delta}(\tilde{\alpha}_2) \equiv V_1 - V_0$.

Lemma 15.7.16. *We have*

$$V_1 = \frac{\delta(1-\pi)(1-\gamma)(\lambda-c) + (1-\delta)\left[1-\delta\pi - \delta(1-\pi)(1-\gamma)(\lambda-c)\right]\tilde{\alpha}_2}{(1-\delta\pi-\gamma\delta(1-\pi)) + \gamma\delta(1-\pi)(1-\delta)\tilde{\alpha}_2}$$

strictly increasing in $\tilde{\alpha}_2$.

Proof. Here we simply solve the linear system and determine the sign of the derivative of V_1.

Plugging V_0 into V_1

$$V_1 = (1-\delta)\tilde{\alpha}_2 + \delta\left[(\tilde{\alpha}_2 + (1-\tilde{\alpha}_2)\pi)\,\frac{\delta(1-\pi)}{1-\delta\pi} + (1-\tilde{\alpha}_2)(1-\pi)\right] V_N$$

Plugging in V_N

$$V_1 =$$

$$(1-\delta)\tilde{\alpha}_2 + \delta\left[(\tilde{\alpha}_2 + (1-\tilde{\alpha}_2)\pi)\,\frac{\delta(1-\pi)}{1-\delta\pi} + (1-\tilde{\alpha}_2)(1-\pi)\right]((1-\gamma)(\lambda-c) + \gamma V_1)$$

from which

$$V_1 = \frac{(1-\delta)\tilde{\alpha}_2 + \delta\left[(\tilde{\alpha}_2 + (1-\tilde{\alpha}_2)\pi)\,\frac{\delta(1-\pi)}{1-\delta\pi} + (1-\tilde{\alpha}_2)(1-\pi)\right](1-\gamma)(\lambda-c)}{1-\gamma\delta\left[(\tilde{\alpha}_2 + (1-\tilde{\alpha}_2)\pi)\,\frac{\delta(1-\pi)}{1-\delta\pi} + (1-\tilde{\alpha}_2)(1-\pi)\right]}$$

$$= \frac{(1-\delta)\tilde{\alpha}_2 + \delta\left[\Sigma\right](1-\gamma)(\lambda-c)}{1-\gamma\delta\left[\Sigma\right]}.$$

We have

$$\begin{aligned}
\Sigma &= (\tilde{\alpha}_2 + (1-\tilde{\alpha}_2)\pi)\,\frac{\delta(1-\pi)}{1-\delta\pi} + (1-\tilde{\alpha}_2)(1-\pi)\\
&= \tilde{\alpha}_2\frac{\delta(1-\pi)}{1-\delta\pi} + (1-\tilde{\alpha}_2)\frac{1-\pi}{1-\delta\pi}\\
&= \frac{1-\pi}{1-\delta\pi}\left(\delta\tilde{\alpha}_2 + 1 - \tilde{\alpha}_2\right)
\end{aligned}$$

$$= \frac{1-\pi}{1-\delta\pi}\left(1-(1-\delta)\tilde{\alpha}_2\right)$$

Plug back into $V_1$to find

$$V_1 = \frac{(1-\delta)\tilde{\alpha}_2 + \delta\left[\frac{1-\pi}{1-\delta\pi}\left(1-(1-\delta)\tilde{\alpha}_2\right)\right](1-\gamma)(\lambda-c)}{1-\gamma\delta\left[\frac{1-\pi}{1-\delta\pi}\left(1-(1-\delta)\tilde{\alpha}_2\right)\right]}$$

$$= \frac{(1-\delta)(1-\delta\pi)\tilde{\alpha}_2 + \delta\left[(1-\pi)\left(1-(1-\delta)\tilde{\alpha}_2\right)\right](1-\gamma)(\lambda-c)}{1-\delta\pi-\gamma\delta\left[(1-\pi)\left(1-(1-\delta)\tilde{\alpha}_2\right)\right]}$$

$$= \frac{\delta(1-\pi)(1-\gamma)(\lambda-c)+(1-\delta)\left[1-\delta\pi-\delta(1-\pi)(1-\gamma)(\lambda-c)\right]\tilde{\alpha}_2}{(1-\delta\pi-\gamma\delta(1-\pi))+\gamma\delta(1-\pi)(1-\delta)\tilde{\alpha}_2}.$$

The derivative $dV_1/D\tilde{\alpha}_2$ has the same sign as

$$\sigma = \left[1-\delta\pi-\delta(1-\pi)(1-\gamma)(\lambda-c)\right](1-\delta\pi-\gamma\delta(1-\pi)) - \delta^2(1-\pi)^2\gamma(1-\gamma)(\lambda-c)$$

$$= (1-\delta\pi)(1-\delta\pi-\gamma\delta(1-\pi))$$

$$-\delta(1-\pi)(1-\gamma)(\lambda-c)(1-\delta\pi-\gamma\delta(1-\pi)) - \delta^2(1-\pi)^2\gamma(1-\gamma)(\lambda-c)$$

$$= (1-\delta\pi)(1-\delta\pi-\gamma\delta(1-\pi))$$

$$-\delta(1-\pi)(1-\gamma)(\lambda-c)(1-\delta\pi) + \delta(1-\pi)(1-\gamma)(\lambda-c)\gamma\delta(1-\pi)$$

$$-\delta^2(1-\pi)^2\gamma(1-\gamma)(\lambda-c)$$

$$= (1-\delta\pi)\left[1-\delta\pi-(1-\pi)\delta\left(\gamma+(1-\gamma)(\lambda-c)\right)\right] > 0.$$

□

LEMMA 15.7.17. $\underline{\Delta}(\tilde{\alpha}_2)$ *is strictly increasing. There is a solution* $0 < \hat{\alpha}_2 < 1$ *to*

$$\underline{\Delta}(\tilde{\alpha}_2) = \overline{\Delta}(\tilde{\alpha}_2) \equiv \frac{1-\delta}{\delta\left(\tilde{\alpha}_2+(1-\tilde{\alpha}_2)\pi\right)}c,$$

it and only if

$$c < \delta\frac{(1-\delta\pi-\delta(1-\pi)\left[\gamma+\lambda(1-\gamma)\right])}{1-\delta\pi-\delta^2(1-\pi)},$$

in which case it is unique.

PROOF. Here solve V_0 as a function of V_1 from the system. We subtract this from V_1 and find that $\underline{\Delta}(\tilde{\alpha}_2)$ is strictly increasing in V_1. Hence we may apply Lemma 15.7.16. Since $\overline{\Delta}(\tilde{\alpha}_2)$ is decreasing there will be a unique intersection if and only if $\overline{\Delta}(0) > \underline{\Delta}(0)$ and $\overline{\Delta}(1) < \underline{\Delta}(1)$. By computation we show that the first condition is always satisfied and the second is the condition on c given as the result.

Turning to the details, we first find V_0

$$V_0 = \frac{\delta(1-\pi)}{1-\delta\pi}V_N = \frac{\delta(1-\pi)}{1-\delta\pi}\left((1-\gamma)(\lambda-c)+\gamma V_1\right).$$

Hence

$$\underline{\Delta}(\tilde{\alpha}_2) \equiv V_1 - V_0 = \left(1-\gamma\frac{\delta(1-\pi)}{1-\delta\pi}\right)V_1 - \frac{\delta(1-\pi)}{1-\delta\pi}\left((1-\gamma)(\lambda-c)\right)$$

$$\underline{\Delta}(\tilde{\alpha}_2) \equiv V_1 - V_0 = \frac{1}{1-\delta\pi}\left[(1-\delta\pi-\gamma\delta(1-\pi))V_1 - \delta(1-\pi)(1-\gamma)(\lambda-c)\right]$$

is strictly increasing in V_1 hence by Lemma 15.7.16 in $\tilde{\alpha}_2$.

The function

$$\overline{\Delta}(\tilde{\alpha}_1) = \frac{1-\delta}{\delta\left(\tilde{\alpha}_2 + (1-\tilde{\alpha}_2)\pi\right)} c$$

strictly decreasing in $\tilde{\alpha}_2$. Hence there is a solution $0 < \hat{\alpha}_2 < 1$ to $\underline{\Delta}(\tilde{\alpha}_2) = \overline{\Delta}(\tilde{\alpha}_2)$ if and only if $\overline{\Delta}(0) > \underline{\Delta}(0)$ and $\overline{\Delta}(1) < \underline{\Delta}(1)$ in which case it is unique. This gives the first result.

From Lemma 15.7.16 at $\tilde{\alpha}_2 = 0$ we have

$$V_1 = \frac{\delta(1-\pi)(1-\gamma)(\lambda-c)}{1-\delta\pi-\gamma\delta(1-\pi)}$$

so

$$\underline{\Delta}(0) = \frac{1}{1-\delta\pi}\left(\delta(1-\pi)(1-\gamma)(\lambda-c) - \delta(1-\pi)(1-\gamma)(\lambda-c)\right) = 0 < \frac{(1-\delta)c}{\delta\pi} = \overline{\Delta}(0).$$

Finally we study $\overline{\Delta}(1) < \underline{\Delta}(1)$. From Lemma 15.7.16

$$\begin{aligned} V_1 &= \frac{\delta(1-\pi)(1-\gamma)(\lambda-c) + (1-\delta)\left[1-\delta\pi-\delta(1-\pi)(1-\gamma)(\lambda-c)\right]}{1-\delta\pi-\gamma\delta(1-\pi)+\gamma\delta(1-\pi)(1-\delta)} \\ &= \frac{(1-\delta)(1-\delta\pi)+\delta^2(1-\pi)(1-\gamma)(\lambda-c)}{1-\delta\pi-\gamma\delta^2(1-\pi)}, \end{aligned}$$

so that

$$\begin{aligned} \underline{\Delta}(1) &= \frac{1}{1-\delta\pi}\Bigg[\frac{1-\delta\pi-\gamma\delta(1-\pi)}{1-\delta\pi-\gamma\delta^2(1-\pi)}\left[(1-\delta)(1-\delta\pi)+\delta^2(1-\pi)(\lambda-\gamma)(1-c)\right] \\ &\qquad -\delta(1-\pi)(\lambda-\gamma)(1-c)\Bigg] \\ &= \frac{1}{1-\delta\pi}\Bigg[\left(\frac{(1-\delta)(1-\delta\pi)\left(1-\delta\pi-\gamma\delta(1-\pi)\right)}{1-\delta\pi-\gamma\delta^2(1-\pi)}\right) \\ &\qquad + \left(\frac{\delta-\delta^2\pi-\gamma\delta^2(1-\pi)}{1-\delta\pi-\gamma\delta^2(1-\pi)} - 1\right)\delta(1-\pi)\left((1-\gamma)(\lambda-c)\right)\Bigg] \\ &= \frac{1}{1-\delta\pi}\Bigg[\left(\frac{(1-\delta)(1-\delta\pi)\left(1-\delta\pi-\gamma\delta(1-\pi)\right)}{1-\delta\pi-\gamma\delta^2(1-\pi)}\right) \\ &\qquad + \left(\frac{\delta-\delta^2\pi-1+\delta\pi}{1-\delta\pi-\gamma\delta^2(1-\pi)}\right)\delta(1-\pi)\left((1-\gamma)(\lambda-c)\right)\Bigg] \\ &= \frac{1}{1-\delta\pi}\Bigg[\left(\frac{(1-\delta)(1-\delta\pi)\left(1-\delta\pi-\gamma\delta(1-\pi)\right)}{1-\delta\pi-\gamma\delta^2(1-\pi)}\right) \\ &\qquad + \left(\frac{-(1-\delta)+\delta\pi(1-\delta)}{1-\delta\pi-\gamma\delta^2(1-\pi)}\right)\delta(1-\pi)\left((1-\gamma)(\lambda-c)\right)\Bigg] \\ &= \frac{(1-\delta)}{1-\delta\pi-\gamma\delta^2(1-\pi)}\left[(1-\delta\pi-\gamma\delta(1-\pi)) - \delta(1-\pi)(1-\gamma)(\lambda-c)\right] \\ &= \frac{(1-\delta)\left[1-\delta\pi-\gamma\delta+\gamma\delta\pi-\delta(1-\pi)(1-\gamma)(\lambda-c)\right]}{1-\delta\pi-\gamma\delta^2(1-\pi)}. \end{aligned}$$

Hence $\overline{\Delta}(1) < \underline{\Delta}(1)$ if and only if

$$\frac{(1-\delta)\left[1-\delta\pi-\gamma\delta+\gamma\delta\pi-\delta(1-\pi)(1-\gamma)(\lambda-c)\right]}{1-\delta\pi-\gamma\delta^2(1-\pi)} > \underline{\Delta}(1) = (1-\delta)c/\delta.$$

We rewrite this inequality

$$\delta\left(1-\delta\pi-\gamma\delta+\gamma\delta\pi-\delta(1-\pi)(1-\gamma)(\lambda-c)\right) > \left(1-\delta\pi-\gamma\delta^2(1-\pi)\right)c$$

$$\delta\left(1-\delta\pi-\gamma\delta(1-\pi)-\lambda\delta(1-\pi)(1-\gamma)\right)$$

$$> \left(1 - \delta\pi - \gamma\delta^2(1-\pi) - \delta^2(1-\pi)(1-\gamma)\right) c$$
$$\delta\left(1 - \delta\pi - \gamma\delta(1-\pi) - \lambda\delta(1-\pi)(1-\gamma)\right) > \left(1 - \delta\pi - \delta^2(1-\pi)\right) c$$
$$c < \delta\frac{(1 - \delta\pi - \gamma\delta(1-\pi) - \lambda\delta(1-\pi)(1-\gamma))}{1 - \delta\pi - \delta^2(1-\pi)}$$
$$c < \delta\frac{(1 - \delta\pi - \delta(1-\pi)\left[\gamma + \lambda(1-\gamma)\right])}{1 - \delta\pi - \delta^2(1-\pi)}.$$

This gives the final result. □

PROPOSITION 15.7.18. *If* $\epsilon < \underline{\mu}\pi^2(1-\pi)\min\{\overline{B}, 1-\overline{B}\}/32$ *and*

$$c \geq \delta\frac{1}{1+\delta(1-\pi)}.$$

all equilibria are in pure strategies.

PROOF. Suppose that $\alpha_1(z), \alpha_2(z)$ is a non-pure equilibrium. If the normal type is willing to provide effort at $z = 1$ we take $\hat{\alpha}_2 = \alpha_2(1)$. If the long-run player strictly prefers to provide no effort at $z = 1$ we show how to construct a $1 > \hat{\alpha}_2 > \alpha_2(1)$ for which the long-run player is indifferent at $z = 1$ and strictly prefers to provide effort at $z = N$. We show that $1 - c \geq V(\alpha_2(N)) \geq V(\hat{\alpha}_2)$ and use this to show that at $\hat{\alpha}_2$ we must have $\underline{\Delta}(\hat{\alpha}_2) = \overline{\Delta}(\hat{\alpha}_2)$ for $\lambda = 1$. Applying Lemma 15.7.17 then yields the desired condition.

Turning to the details, from Lemmas 15.7.13, 15.7.14, and 15.7.15 we know that $\alpha_1(0) = \alpha_2(0) = 0$, $\alpha_1(N) > 0$, $\alpha_2(N) \geq \alpha_2(1)$, $\alpha_1(1) < 1$, and $0 < \alpha_2(1) < 1$.

If the long-run player strictly prefers not to provide effort at $z = 1$ then $\alpha_1(1) = 0$. Moreover we must have $\alpha_2(N) > \alpha_2(1)$ since if the two are equal and effort is weakly preferred at $\alpha_2(N)$ it would be at $\alpha_1(1)$ as well. For $V(\alpha_2(0))$ we solve the Bellman system to find

$$V_0 = \frac{\delta(1-\pi)}{1-\delta\pi}V_N$$

and for $V(\alpha_2(N))$ we solve

$$V(\alpha_2(N)) =$$
$$\frac{1-\delta}{1-\delta(1-\alpha_2(N))(1-\pi)}\left[\alpha_2(N) - c\right] + \frac{\delta\left(\alpha_2(N) + (1-\alpha_2(N))\pi\right)}{1-\delta(1-\alpha_2(N))(1-\pi)}V(\alpha_2(1)).$$

Hence if hold fixed $\alpha_2(N)$ and take $\lambda = \alpha_2(N)$ and

$$\gamma = \frac{\delta\left(\alpha_2(N) + (1-\alpha_2(N))\pi\right)}{1-\delta(1-\alpha_2(N))(1-\pi)}$$

the Bellman system corresponds to the auxiliary system, so $V(\alpha_2(1)) - V(\alpha_2(0)) = \underline{\Delta}(\alpha_2(1))$. From Lemma 15.7.17 this is strictly increasing.

As earlier we may define the gain to providing no effort at z as

$$G(\alpha_2(z), \alpha_2(1)) = (1-\delta)c - \delta\left(\pi + \alpha_2(z)(1-\pi)\right)\left[V(\alpha_2(1)) - V(\alpha_2(0))\right]$$

and it follows that G is strictly decreasing in both arguments. Hence as we increase $\alpha_2(1)$ to α_2 the gain $G(\alpha_2, \alpha_2)$ to providing no effort at $z = 1$ and the gain $G(\alpha_2(N), \alpha_2)$ to no effort at $z = N$ both strictly decline. Initially at $z = N$ the long-run player weakly preferred to provide effort, hence as we increase α_2 the long-run player strictly prefers to provide effort. At $z = 1$ the long-run player strictly preferred no effortbut when α_2 reaches $\alpha_2(N)$ effort is strictly preferred, and as

G is continuous, this implies for some $\hat{\alpha}_2 < 1$ the long-run player is indifferent at $z = 1$.

To summarize: in all cases with the original value of $\alpha_2(0), \alpha_2(N)$ and the short-run player using $\hat{\alpha}_2 \leq \alpha_2(N)$ in state $z = 1$ the strategy for the long-run player of providing no effort in states $z = 0, 1$, providing effort in state N is optimal and the long-run player is indifferent in state $z = 1$. We next show that with respect to this (possibly modified) strategy by player 2 the long-run player has $1 - c \geq V(\alpha_2(N)) \geq V(\hat{\alpha}_2)$ and use this to show that at $\hat{\alpha}_2$ we must have $\underline{\Delta}(\hat{\alpha}_2) = \overline{\Delta}(\hat{\alpha}_2)$ for $\lambda = 1$. Applying Lemma 15.7.17 then yields the desired condition.

Since it is also optimal for the long-run payer to provide effort, the Bellman system may be written as

$$V(\hat{\alpha}_2) = (1-\delta)\left[\hat{\alpha}_2 - c\right] + \delta\left[\left(\hat{\alpha}_2 + (1-\hat{\alpha}_2)\pi\right)V_1(\hat{\alpha}_2) + (1-\hat{\alpha}_2)(1-\pi)V(\alpha_2(N))\right]$$

$$V(\alpha_2(N)) = (1-\delta)\left[\alpha_2(N) - c\right]$$

$$+\delta\left[\left(\alpha_2(N) + (1-\alpha_2(N))\pi\right)V(\hat{\alpha}_2) + (1-\alpha_2(N))(1-\pi)V(\alpha_2(N)).\right]$$

This implies that for some $0 \leq \lambda \leq 1$ we have $V(\hat{\alpha}_2) = (1-\lambda)\left[\hat{\alpha}_2 - c\right] + \lambda\left[\alpha_2(N) - c\right]$ that is, a weighted average of the period payoffs in the two states. Yet $V(\hat{\alpha}_2) > V(\alpha_2(N))$ gives $V(\alpha_2(N)) > (1-\delta)\left[\alpha_2(N) - c\right] + \delta V(\alpha_2(N))$ so $V(\alpha_2(N)) > \alpha_2(N) - c \geq \hat{\alpha}_2 - c$. which implies $V(\alpha_2(N)) > V(\hat{\alpha}_2)$ a contradiction. Hence $V(\alpha_2(N)) \geq V(\hat{\alpha}_2)$.

For some $0 \leq \lambda' \leq 1$ we also have $V(\alpha_2(N)) = (1-\lambda')\left[\hat{\alpha}_2 - c\right] + \lambda'\left[\alpha_2(N) - c\right] \leq 1 - c$. This establishes the target $1 - c \geq V(\alpha_2(N)) \geq V(\hat{\alpha}_2)$.

From $1 - c \geq V(\alpha_2(N)) \geq V(\hat{\alpha}_2)$ we see that for some $0 \leq \gamma \leq 1$ we have $V(\alpha_2(N)) = \gamma(1-c) + (1-\gamma)V(\hat{\alpha}_2)$. It follows from this and indifference of the long-run player at $\hat{\alpha}_2$that for this value of γ and $\lambda = 1$ that at $\hat{\alpha}_2$ we must have $\underline{\Delta}(\hat{\alpha}_2) = \overline{\Delta}(\hat{\alpha}_2)$. From Lemma 15.7.17 this means that

$$c < \delta\frac{\left(1 - \delta\pi - \delta(1-\pi)\left[\gamma + \lambda(1-\gamma)\right]\right)}{1 - \delta\pi - \delta^2(1-\pi)}$$

$$= \delta\frac{1 - \delta\pi - \delta(1-\pi)}{1 - \delta\pi - \delta^2(1-\pi)}$$

$$= \delta\frac{1-\delta}{(1-\delta)(1+\delta(1-\pi))}$$

the desired result. □

15.7.7. Role of Types. We turn now to a converse of Proposition 15.7.18: that is when

$$c < \delta\frac{1}{1+\delta(1-\pi)}$$

are there equilibria that are not pure? Intuitively this cannot be the case for all Q. If there are very few normal types then basically the short-run player ignores them and plays a best response to the behavioral types - which is to say the pure strategy of staying out on a bad or no signal and entering on a good signal. This we know leads the normal type to best-respond with a pure strategy as given in Proposition 15.7.18.

Our first result is precise result: it shows if there are enough good types there is necessarily a pure strategy equilibrium.

PROPOSITION 15.7.19. *For any Q with*

$$\mu_g > \frac{\overline{B}}{\overline{B} + (1-\overline{B})(\pi/2)}$$

if $\epsilon \leq \underline{\mu}\overline{A}/32$ then all equilibria are pure.

PROOF. From Bayes Law

$$\mu^1(1) \geq \mu_{g|1} = \frac{\mu_{1|g}\mu_g}{\mu_{1|g}\mu_g + \sum_{\tau \neq g} \mu_{1|\tau}\mu_\tau} \geq \frac{1}{1 + (1-\mu_g)/(\mu_{1|g}\mu_g)}.$$

From Lemma 15.7.3

$$\mu_{1|g} \geq$$
$$[\alpha_2(1) + \pi(1-\alpha_2(1))]\,\mu_{1|g} + [\alpha_2(N) + \pi(1-\alpha_2(N))]\,\mu_{N|g} - 2\epsilon.$$

The same Lemma implies $\mu_{0|g} \leq 2\epsilon$, so

$$\mu_{1|g} \geq [\alpha_2(1) + \pi(1-\alpha_2(1)) - \pi]\,\mu_{1|g} + \pi - 4\epsilon \geq \pi - 4\epsilon.$$

Combining the two

$$\mu^1(1) \geq \frac{1}{1 + (1-\mu_g)/((\pi - 4\epsilon)\mu_g)}.$$

By Lemma 15.7.5 if

$$\frac{1}{1 + (1-\mu_g)/((\pi - 4\epsilon)\mu_g)} > \overline{B}$$

or equivalently

$$\mu_g > \frac{\overline{B}}{\overline{B} + (1-\overline{B})(\pi - 4\epsilon)}$$

then $\alpha_2(1) = 1$ so the result follows from Lemma 15.7.15 and the assumption that $\epsilon < \underline{\mu}\overline{A}/32 \leq \pi/2$. □

This is not terribly interesting in itself: the case of interest is when they are many normal types, but it does show that there is no converse to Proposition 15.7.18 without an assumption on Q. Hence we investigate the interesting case of many normal types.

In addition to showing that there are mixed equilibria, we can say what they look like. There are two types, *single mixing* and *double mixing*. In both types of equilibrium in the bad state $z = 0$ there is no effort and the short-run player stays out: $\alpha_1(0) = 0$, $\alpha_2(0) = 0$. In the good state $z = 1$ both players strictly mix: $0 < \alpha_1(1) < 1$, $0 < \alpha_2(1) < 1$. In the single mixing equilibrium this is the only mixing: in the state $z = N$ the normal type provides effort and the short-run player enters $\alpha_1(N) = 1, \alpha_2(N) = 1$. In the double mixing case equilibrium mixing takes place also at $z = N$: the short-run player mixes exactly as in the state $z = 1$,that is $\alpha_2(N) = \alpha_2(1)$, while normal type provides effort with a positive probability $\alpha_1(N) > 0$.

To state a precise result and also be clear about the order of limits, it is useful to define the notion of a *fundamental bound*. This is a number that may depend on the fundamentals of the game π, V, δ, c but not on the type dynamics Q, ϵ. Recall that $\overline{B}$ is the probability of effort that makes the short-run player indifferent to entry.

LEMMA 15.7.20. *There exists a fundamental bound $\overline{\mu} < 1$ such that for any Q with $\mu_n \geq \overline{\mu}$ if for $\epsilon \leq \underline{\mu}\overline{A}/32$ a non-pure equilibrium is either a single- or double-mixing profile.*

PROOF. The only things not covered in Lemmas 15.7.13, 15.7.14, and 15.7.15 are $\alpha_1(1) \neq 0$ and the result that $\alpha_2(N) > \alpha_2(1)$ implies $\alpha_1(N) = 1, \alpha_2(N) = 1$.

For the first result, the idea is since μ_n is large there must be many more normal types at N than good types. Since since $\alpha_2(N) > 0$ this means that $\alpha_1(N)$ cannot be too small, and this in turn implies that even though $\alpha_1(1) = 0$ there must be many more normal types at 1 then good types. If they provide no effort then the short-run player should stay out contradicting the fact that we already know $\alpha_2(1) > 0$.

For the second result we leverage the first to see that we must have $\alpha_1(N) = 1$. Moreover, since $\alpha_1(1) < 1$ there must be many normal types at $z = 0$, and so at $z = N$. As these are all providing effort, it is optimal for the short-run player to enter.

Turning to the details, suppose in fact $\alpha_1(1) = 0$. Since $\alpha_2(N) > 0$ we must have $(1 - \mu_n)V + \alpha_1(N)\mu_{N|n}\mu_n V \geq (1 - \alpha_1(N))\mu_{N|n}\mu_n$. We may rewrite this as

$$\alpha_1(N)\mu_{N|n} \geq \frac{1}{1+V}\left(\mu_{N|n} - \frac{(1-\mu_n)}{\mu_n}V\right).$$

From Lemma 15.7.3 we have $\mu_{1|n} \geq \pi\alpha_1(N)\mu_{N|n} - 2\epsilon$ so

$$\mu_{1|n} \geq \frac{\pi}{1+V}\left(\mu_{N|n} - \frac{1-\mu_n}{\mu_n}V\right) - 2\epsilon.$$

Also from Lemma 15.7.3 we have $\mu_{N|n} \geq (1-\pi)(1 - \mu_{N|n} - \mu_{1|n}) - 2\epsilon$ so that $\mu_{N|n} \geq (1 - \mu_{1|n})(1-\pi)/\pi - 2\epsilon/\pi$ implying

$$\mu_{1|n} \geq \frac{1}{1+V}\left(\left[(1-\mu_{1|n})(1-\pi) - 2\epsilon\right] - \pi\frac{1-\mu_n}{\mu_n}V\right) - 2\epsilon$$

or

$$\mu_{1|n} \geq \frac{\left([(1-\pi) - 2\epsilon] - \pi\frac{1-\mu_n}{\mu_n}V\right) - 2(1+V)\epsilon}{2+V-\pi}.$$

Since $\alpha_2(1) > 0$ we must have $(1-\mu_n)V \geq \mu_{1|n}\mu_n$, so

$$\frac{1-\mu_n}{\mu_n}V \geq \frac{\left([(1-\pi) - 2\epsilon] - \pi\frac{1-\mu_n}{\mu_n}V\right) - 2(1+V)\epsilon}{2+V-\pi}$$

or

$$\frac{1-\mu_n}{\mu_n}V \geq \frac{1-\pi}{2+V} - 2\epsilon.$$

Our assumption implies $\epsilon < (1-\pi)\overline{B}/4$ which means that $\epsilon \leq (1-\pi)/(4+2V)$. Hence

$$\frac{1-\mu_n}{\mu_n} \geq \frac{1-\pi}{V(4+2V)}$$

or

$$\mu_n \leq \frac{V(4+2V)}{V(4+2V) + 1 - \pi} < 1.$$

Hence $\alpha_1(1) = 0$ is ruled out by large μ_n.

Next suppose that $\alpha_2(N) > \alpha_2(1)$. Since $\alpha_1(1) > 0$ Lemma 15.7.1 implies $\alpha_1(N) = 1$. It remains to show that this in turn forces $\alpha_2(N) = 1$.

From Lemma 15.7.2

$$\mu_{N|n} \geq (1-\pi)\left((1-\alpha_2(1))\mu_{1|n} + \mu_{0|n} - 2\epsilon\right).$$

Suppose that $\mu_{N|n} \leq 1/3$. Then either $\mu_{0|n} \geq 1/3$ or $\mu_{1|n} \geq 1/3$.

In the former case we have $\mu_{N|n} \geq (1-\pi)/3 - 2\epsilon$.

In the latter case strict mixing by the short-run player at $z = 1$ implies from Lemma 15.7.6 that

$$1 - \overline{B} = 1 - \mu^1(1) = \frac{[1-\alpha_1(1)]\mu_{1|n}\mu_n + \mu_{1|b}\mu_b}{\mu_1} \leq \frac{[1-\alpha_1(1)]\mu_{1|n}\mu_n + \mu_{1|b}\mu_b}{\mu_{1|n}\mu_n + \mu_{1|b}\mu_b}$$
$$= \gamma(1-\alpha_1(1)) + (1-\gamma).$$

If $1 - \alpha_1(1) < (1-\overline{B})/2$ then

$$1 - \overline{B} \leq \gamma(1-\overline{B})/2 + (1-\gamma)$$

or

$$\gamma \leq \frac{2\overline{B}}{1+\overline{B}}.$$

Lemma 15.7.3 gives $\mu_{1|b} \leq 2\epsilon$ so since $\mu_{1|n} \geq 1/3$

$$\gamma = \frac{\mu_{1|n}\mu_n}{\mu_{1|n}\mu_n + \mu_{1|b}\mu_b} \geq \frac{\mu_n}{\mu_n + 6\epsilon(1-\mu_n)}.$$

Since $\epsilon \leq 1/6$ it follows that $\gamma \geq \mu_n$ so if

$$\mu_n > \frac{2\overline{B}}{1+\overline{B}}$$

we have a contradiction, so $1 - \alpha_1(1) \geq (1-\overline{B})/2$. Hence by Lemma 15.7.3 $\mu_{0|n} \geq \pi(1-\overline{B})/2 - 2\epsilon$ from which $\mu_{N|n} \geq \pi(1-\pi)(1-\overline{B})/2 - 4\epsilon$.

In all cases then $\mu_{N|n} \geq \pi(1-\pi)(1-\overline{B})/2 - 4\epsilon$. As

$$\epsilon \leq \pi(1-\pi)(1-\overline{B})/16$$

this is $\mu_{N|n} \geq \pi(1-\pi)(1-\overline{B})/4$.

The short-run player must enter if

$$\mu_{N|n}\mu_n V > \mu_N - \mu_{N|n}\mu_n$$

while $\mu_N \leq \mu_{N|n}\mu_n + (1-\mu_n)$ so entry must occur if

$$\mu_{N|n}\mu_n V > 1 - \mu_n$$

or

$$\mu_n > \frac{1}{1+V\mu_{N|n}} \geq \frac{4}{4+\pi(1-\pi)(1-\overline{B})},$$

□

Lemma 15.7.21. *In any single- or double-mixing profile if $\mu_n \geq 1/2$ and $\epsilon < (1-a_2(1))(1-\pi)/12$ then*

$$\mu_{n|N} \geq 1 - \frac{1-\mu_n}{(1-\alpha_2(1))(1-\pi)/12}.$$

If in addition $\epsilon < \alpha_1(N)\pi(1-a_2(1))(1-\pi)/24$ then

$$\mu_{n|1} \geq 1 - \frac{1-\mu_n}{\alpha_1(N)\pi(1-\alpha_2(1))(1-\pi)/24}.$$

PROOF. The first result says that if $\alpha_2(1)$ is less than 1 and if there are many normal types there must be many normal types at $z = N$, as they are flowing there from both $z = 0$ and $z = 1$. The second result leverages this to say that if there are many normal types at $z = N$ and $\alpha_1(N)$ is large then there must be many normal types at $z = 1$.

Turning to the details, we start with an inequality that follows from Bayes law:

$$\begin{aligned}
\mu_{n|z} = \frac{\mu_{z|n}\mu_n}{\mu_z} &\geq \frac{\mu_{z|n}\mu_n}{\mu_{z|n}\mu_n + (1-\mu_n)} \\
&= 1 - \left(1 - \frac{\mu_{z|n}\mu_n}{\mu_{z|n}\mu_n + (1-\mu_n)}\right) \\
&= 1 - \left(\frac{1-\mu_n}{\mu_{z|n}\mu_n + (1-\mu_n)}\right) \\
&\geq 1 - \frac{1-\mu_n}{\mu_{z|n}\mu_n}.
\end{aligned}$$

Since $\mu_n \geq 1/2$ this implies

$$\mu_{n|z} \geq 1 - \frac{1-\mu_n}{\mu_{z|n}/2}.$$

To get the required bounds, it then suffices to get a lower bound on $\mu_{z|n}$. Take first $z = N$. Suppose that $\mu_{N|n} \leq 1/3$. Then either $\mu_{0|n} \geq 1/3$ or $\mu_{1|n} \geq 1/3$. If $\mu_{0|n} \geq 1/3$ then by Lemma 15.7.3 $\mu_{N|n} \geq (1-\pi)/3 - 2\epsilon$. If $\mu_{1|n} \geq 1/3$ then by the same Lemma $\mu_{N|n} \geq (1-a_2(1))(1-\pi)/3 - 2\epsilon$. For $\epsilon < (1-a_2(1))(1-\pi)/12$ this is $\mu_{N|n} \geq (1-a_2(1))(1-\pi)/6$ giving the first bound.

From Lemma 15.7.3$\mu_{1|n} \geq \alpha_1(N)\pi\mu_{N!n} - 2\epsilon$ so that the first bound implies $\mu_{1|n} \geq \alpha_1(N)\pi(1-a_2(1))(1-\pi)/6 - 2\epsilon$. For $\epsilon < \alpha_1(N)\pi(1-a_2(1))(1-\pi)/24$ this is $\mu_{1|n} \geq \alpha_1(N)\pi(1-a_2(1))(1-\pi)/12$ giving the second bound. □

The next Lemma is simply an observation:

LEMMA 15.7.22. *A single mixing equilibrium corresponds to the auxiliary system with $\lambda = 1$ and $\gamma = \delta$ and a double mixing equilibrium corresponds to the auxiliary system with $\lambda = 1$ and $\gamma = 1$. In particular in a single mixing equilibrium*

$$V(\alpha_2(1)) = \frac{(1-\delta\pi)\alpha_2(1)}{1+\delta(1-\pi)\alpha_2(1)}$$

which is increasing in $\alpha_2(1)$.

PROOF. In the single mixing case this is just the Bellman equation. In the double mixing case we use the fact that $V(\alpha_2(N)) = V(\alpha_2(1))$. The value $V(\tilde{\alpha}_2)$ follows from plugging into the expression for V_1 in Lemma 15.7.16; that Lemma gives the result that it is increasing. □

PROPOSITION 15.7.23. *There exists a fundamental bound $\overline{\mu} < 1$ such that for any Q with $\mu_n \geq \overline{\mu}$ if $\epsilon \leq \underline{\mu}\overline{A}/32$ and*

$$c < \delta\frac{1}{1+\delta(1-\pi)}$$

there is at least one single-mixing and one double-mixing equilibrium and no other type of mixed equilibrium. In both cases the equilibrium value of $\alpha_2(1)$ is the unique solution of $\underline{\Delta}(\alpha_2(1)) = \overline{\Delta}(\alpha_2(1))$ where $\lambda = 1$ and in the single-mixing case $\gamma = \delta$

and in the double-mixing case $\gamma = 1$. *Moreover, the equilibrium value of* $\alpha_1(z)$ *satisfies*

$$|\alpha_1(z) - \overline{B}| \leq \frac{1-\mu_n}{1-\overline{\mu}}$$

for $z = 1$ *in the single mixing case and* $z \in \{N, 1\}$ *in the double-mixing case.*

Proof. From Lemma 15.7.20 we know there can be no other kind of equilibrium. From Lemma 15.7.17 we know that

$$c < \delta \frac{(1 - \delta\pi - \delta(1-\pi)\left[\gamma + \lambda(1-\gamma)\right])}{1 - \delta\pi - \delta^2(1-\pi)}$$

and from Lemma 15.7.22 with $\lambda = 1$ and $\gamma = \delta$ is a necessary condition for the existence of single-mixing equilibrium and with $\lambda = 1$ and $\gamma = 1$ for the existence of a double-mixing equilibrium. When $\lambda = 1$ the RHS is independent of γ and given as the expression in the Theorem.This gives us a unique solution $0 < \tilde{\alpha}_2 < 1$ for the equilibrium value of $\alpha_2(1)$. The crucial fact is that $\tilde{a}_2$ arising from the optimization problem of the normal type is itself a fundamental bound.

We must now show the existence of an $\alpha_1(1)$ so that the short-run player is indifferent when $z = 1$ and weakly prefers to enter when $z = N$, and in the double mixing case the existence of $\alpha_1(1), \alpha_1(N)$ so that the short-run player is indifferent in both $z = N, 1$, and that any such strategic components satisfy the required bound.

Recall that $\mu^1(z)$ are the beliefs of the short-run player about the probability the long-run player will provide effort. This is given as $\mu^1(z) = \mu_{g|z} + \mu_{n|z}\alpha_1(z)$. Define $\tilde{A}(z, \alpha_1(z)) = \mu_1(z) - \overline{B}$. Hence the equilibrium requirement is that $\tilde{A}(1, \alpha_1(1)) = 0$ and that in the single mixing case $\tilde{A}(N, \alpha_1(N)) = 0$ and in the double-mixing case $\tilde{A}(N, 1) \geq 0$. The complication is that $\mu_{g|z}$ and $\mu_{n|z}$ for $z \in \{N, 1\}$both depend upon $\alpha_1(1)$ and $\alpha_1(N)$. As by the ergodic theorem the ergodic distribution is continuous in $\alpha_1(1)$ and $\alpha_1(N)$ so are $\tilde{A}(z, \alpha_1(z))$ and we will be able to apply fixed point argument.

Write $\tilde{A}(z, \alpha_1) = \mu_{g|z} - (1-\mu_{n|z})\alpha_1 + \alpha_1 - \overline{B}$ and observe that $\mu_{g|z} \leq (1-\mu_{n|z})$. Hence $\tilde{A}(z, \alpha_1) = \alpha_1 - \overline{B} + \tilde{A}_1(1-\mu_{n|z})$ with $|\tilde{A}_1| \leq 2$.

We now apply the first bound from Lemma 15.7.21. We know that $\alpha_2(1) = \tilde{\alpha}_2$ a fundamental bound so we have $\tilde{A}(N, \alpha_1) = \alpha_1 - \overline{B} + \tilde{A}_2(1-\mu_n)$ where $|\tilde{A}_2| \leq A_2$ and A_2 is a fundamental bound. Hence for $\alpha_1 - \overline{B} \leq -A_2(1-\mu_n)$ we have $\tilde{A}(N, \alpha_1) < 0$. Taking $A_2(1-\mu_n) \leq \overline{B}/2$ for $\alpha_1 \leq \overline{B}/2$ we also have $\tilde{A}(N, \alpha_1) < 0$. We may restrict attention then to the region where $\alpha_1(N) \geq \overline{B}/2$ since there can be no equilibrium outside this region.

In the region $\alpha_1(N) \geq \overline{B}/2$ we may now apply the second bound from Lemma 15.7.21 and find that $\tilde{A}(1, \alpha_1) = \alpha_1 - \overline{B} + \tilde{A}_3(1-\mu_n)$ where $|\tilde{A}_3| \leq A_3$ and A_3 is a fundamental bound.

Take first the single-mixing case. Here if we take $A_2(1-\mu_n) \leq (1-\overline{B})/2$ we have $\tilde{A}(N, 1) > 0$ and we have $\tilde{A}(1, \alpha_1)$ negative for $\alpha_1 - \overline{B} < -A_3(1-\mu_n)$ and positive for $\alpha_1 - \overline{B} > A_3(1-\mu_n)$ implying at least one solution $\tilde{A}(1, \alpha_1) = 0$ in the interval $|\alpha_1 - \overline{B}| \leq A_3(1-\mu_n)$ and none elsewhere. That is the first required result.

In the double mixing case we take the rectangle $|\alpha_1(1) - \overline{B}| \leq A_3(1-\mu_n)$ and $|\alpha_1(N) - \overline{B}| \leq A_2(1-\mu_n)$ and observe that $\tilde{A}(1, \alpha_1), \tilde{A}(N, \alpha_1)$ are not both zero outside this region. Moreover, the vector field $(\tilde{A}(1, \alpha_1(1)), \tilde{A}(N, \alpha_1(N))$ points outwards on the boundary of the rectangle. By the continuous vector field version of the Brouwer fixed point theorem there is at least one point inside the rectangle where they both vanish. □

Note that we do not guarantee a unique equilibrium of each type, but show that if there are enough normal types then all equilibria of a given type are similar and the mixing by the long-run normal type is approximately the value that makes the short-run player indifferent. The reason this is only approximate is because the short-run player also faces an endogenous number of good and bad types who are either providing effort or not.

How do the mixed equilibria differ from the pure equilibrium? Roughly speaking we can describe the pure equilibria as having three properties: the signal is informative for the short-run player, reputation is valuable, and the normal type of long-run player remains stuck in either a good or bad situation. The mixed equilibria are quite different: the signal is uninformative for the short-run player, reputation is not valuable, and the normal type of long-run player transitions back and forth between all the states.

Specifically, with the mixed equilibrium we have the following situation. In every state the short-run player is facing mostly normal types. The normal type, starting in state $z = 0$ will eventually have some luck, the short-run player will not observe the long-run player, and the state will move to N. Here the normal type provides effort with positive probability and the short-run player observes this with positive probability so there is a chance of getting to the state $z = 1$. Once there both players are mixing, so there is a chance of moving to either state $z = 0$ or state $z = N$. Indeed, the only transitions that are not seen are moving directly from $z = 0$ to $z = 1$ and in the single mixing case moving directly from $z = N$ to $z = 0$. The normal type transitions back and forth between all the states. Because of this mixing the behavioral types play no role in the inferences of the short-run player. This is similar to the cheap talk literature:[7] the mixing of the long-run player effectively jams the signal of the behavioral types, and reputation plays no role in equilibrium. These equilibria also have the property that $\alpha_2(N) \geq \alpha_2(1)$: the short-run player is no more likely to enter when there is a favorable signal than when there is no signal. This represents a precise sense in which the "signal is jammed."

Finally, we emphasize that for very low c there are always signal jamming equilibria: low c does not guarantee a good equilibrium.

15.7.8. Welfare. Is a mixed equilibrium good or bad for the long-run player? This is irrelevant in the bad equilibrium case where $c > \delta$ as there is no mixed equilibrium there. If $\pi < (1 - \delta)/\delta$ and

$$\delta \frac{\pi}{1 - \delta + \delta\pi} < c < \delta \frac{1}{1 + \delta(1 - \pi)}$$

then there is both a trap equilibrium and mixed equilibrium. The mixed equilibrium is clearly good for a long-run normal type who is trapped with no reputation - that type gets 0 while receives a positive payoff in the mixed equilibria. In this sense signal jamming is potentially good because it can alleviate a reputation trap.

[7] See, for example, Crawford and Sobel (1982).

On the other hand, a long-run normal type with a good reputation gets $1 - c$. The next result shows that in this case a double-mixing equilibrium is unambiguously bad: expected average present value starting in the good state is strictly less.

PROPOSITION 15.7.24. *In a double mixing equilibrium*

$$V(\alpha_2(1)) < \frac{1 - \delta\pi}{1 + \delta(1 - \pi)} \leq 1 - c.$$

PROOF. From Lemma 15.7.16

$$V(\alpha_2(1)) = \frac{(1 - \delta\pi)\alpha_2(1)}{1 + \delta(1 - \pi)\alpha_2(1)}$$

which is strictly increasing in $\alpha_2(1)$, so the first bound follows from $\alpha_2(1) < 1$. The final inequality is a restatement of the condition for the existence of a double mixing equilibrium from Proposition 15.7.23. □

This has the following additional consequence. As $\delta \to 1$ regardless of initial condition utility in the good equilibrium approaches $1 - c$. On the other hand the Theorem shows that $\limsup V(\alpha_2(1))$ is bounded above by $(1 - \pi)/(2 - \pi)$ which does not depend upon c. Hence for small enough c starting in the good state the normal long-run player does strictly worse in the double-mixing equilibrium than in the always provide effort equilibrium even as $\delta \to 1$. This result appears quite different than the long memory case analyzed in Fudenberg and Levine (1989) and Ekmekci et al. (2012).

To understand why this is, observe that with sufficiently long memory by the short-run player the long-run player can foil a signal jamming equilibrium: if the long-run player persists in effort provision Fudenberg and Levine (1992) show that when there is a good type the short-run player must come to believe that the long-run player will provide effort. To understand how the conflict between the conclusions for $\delta \to 1$ arises, observe that for any fixed length of time the Fudenberg and Levine (1992) bound requires the prior probability of the good type to be sufficiently high. Here the length of time is indeed fixed - the long-run player has only one period to convince the short-run player that there will be effort. Hence, as Proposition 15.7.13 shows, and as the Fudenberg and Levine (1992) result suggests, signal jamming is ruled out if the prior probability of the good type is sufficiently high. Hence the result here that equilibrium payoffs remain bounded away from the Stackelberg payoff of $1 - c$ when the probability of the good type is too low is an example confirming that the Fudenberg and Levine (1992) bound must depend on the strength of prior belief in the good type.

CHAPTER 16

Backwards Compatibility

Political economists and political scientists have long recognized the importance of group behavior and have employed a variety of models to analyze contests and other games between groups. This research has been fruitful and yielded many conceptual and empirical insights in a broad range of issues in political economy. While the new normal in behavioral economics seems to be to introduce a new model to explain every new observation and all previous models and observations be damned, we do not subscribe to this as a useful approach to science. As we are introducing an alternative to existing models we think it crucial to examine the extent to which is is consistent with existing models - does it advance the art, or do we throw out the baby with the bathwater and unexplain many facts that already have satisfactory and empirically valid explanations? We conclude this book by examining some important existing models of group behavior and to what extent they are compatible with the theory of social mechanisms. Not to keep the reader in suspense: social mechanisms are backwards compatible, they are consistent with most existing research, and in the cases where that is not true the evidence favors social mechanism theory.

16.1. The Group as Individual

The single most widely used model of group behavior is that of a group acting as a single individual choosing in the best interest of the individual members. This is the case for many models of conflict between elites and masses used to study issues ranging from revolution and extension of the voting franchise (Acemoglu and Robinson (2001)) to trade protectionism (Galiani and Torrens (2014)). No doubt due the catastrophic and genocidal behavior of Lenin and his successors the work of Karl Marx has fallen into disrepute - yet if we were to replace the contemporary terminology of elites and masses with the words bourgeoisie and proletariat and describe our game theoretic models of conflict as class warfare only the names would be changed to protect the innocent.

There is, in fact, a profound skepticism in economics of viewing a group as an individual: the idea of individuals acting in the best interest of the group is part of the utopian ideal of Marxism that has proven profoundly unworkable. Few people today, whether or not they are economists, expect that bread will be delivered from the benevolence of the baker. Hence in studying markets economists have rightly focused on the behavior of individuals. Even cooperative game theory has fallen into a degree of disrepute with its emphasis on common goals. As a result effort has been made to more closely examine the individual moral calculations that might underlie benevolent individual behavior. This has come to be known as *rule consequentialism.* In the context of social norms, each individual asks what would be in the best interest of the group, that is, what social norm φ is most advantageous for the group? Then having determined this (hopefully unique) social norm each member "does their part" by implementing $\phi^i = \varphi$. Conceptually this is supposed to capture the idea that it is unethical to free ride. These models have been studied by Harsanyi (1977), Roemer (2010), Hooker (2011) among others. In the voting literature it is also known as the ethical voter model and has been studied by Feddersen and Sandroni (2006), Coate and Conlin (2004b), Li and Majumdar (2010) and Ali and Lin (2013) among others. The idea is that rule consequentialism is decentralized so that each group member independently calculates what they are supposed to contribute. Roemer (2010) points out this requires that members know each others utility functions and proposes an alternative Kantian notion of equilibrium which can be more easily decentralized.

There are two points to be made about this. First, the literature on rule consequentialism is not naive: it recognizes that, for example, in the provision of private goods individuals do not produce out of their moral calculation of the group interest. We point here particularly to Feddersen and Sandroni (2006) who hypothesize that individuals receive a benefit from doing their moral duty - but if the cost of fulfilling their moral duty should exceed the benefit then they will behavior selfishly. Hence when costs are low as in elections we might expect moral calculations to play a heavy role, but in a setting such production where costs are high we would not expect this. Indeed, our analysis in section 8.3.5 of the internalization of social norms was derived from Feddersen and Sandroni (2006). We will examine the idea further when we discuss public goods below.

Second, the social mechanism model we have proposed also presumes that groups act in their own self-interest. It adds, however, that individuals in the group, perhaps after calculating the group optimal mechanism, or perhaps by agreement with others, in the privacy of their own decision making will only act according to the agreement if it is in their self-interest to do so - hence adds incentive constraints. Certainly social mechanism theory is completely backwards compatible with the group acting as an individual: if there is perfect monitoring then there are no on equilibrium costs of punishment and the conclusions of social mechanism theory are identical to the theory that members of a group all act in the best interest of the group. We should note, however, that there is plentiful evidence that there is punishment on the equilibrium path. To take a rather extreme example, on April 21, 2015 USA Today ran a news story about a women who ran her husband over with an automobile because he failed to vote. This does not happen in the ethical voter model - while some voters might not be ethical, no voter ever punishes another.

16.2. Coordination and Altruism

A model conceptually similar to the group as individual model is a the model of pure altruism. When individuals are completely altruistic they care about group utility, and care about their own utility only to the extent that it contributes to the social whole. When there is a unique equilibrium this is identical to the theory of group as individual or rule consequentialism. However, when equilibria are not unique it allows for the possibility of coordination failure.

Consider as an example a class public goods games called *minimum games* in which the aggregate contribution is equal to the minimum of the individual contributions. In these games it is assumed that group welfare is increasing in the aggregate contribution so that the social optimum is for everyone to contribute the maximum amount. It is assumed as well that the cost savings to an individual from reducing their contribution is less than their loss. A simple example of such a game is the payoff matrix in which there are two players' whose contributions range from 0 to 2:

	0	1	2
0	0,0	0,-1	0,-2
1	-1,0	1,1	1,0
2	-2,0	0,1	2,2

In any minimum game - regardless of how altruistic players might be - if everyone contributes the same amount it is an equilibrium even if contributions are inefficiently low. The reason is easy to see: reducing a contribution lowers cost but also lowers the minimum by the same amount so reduces individual utility. Increasing a contribution does not change the minimum, it merely incurs higher cost. In the 3×3 example above it is an equilibrium for there to be no contributions as well as the efficient equilibrium in which all contribute 2 (and the third pure strategy equilibrium in which they both contribute 1).

By contrast social mechanisms and the group as individual assume that group members can coordinate their effort to pick the most preferred alternative: that all contribute the maximum because they all agree this is the best. As we discussed in section 9.1 there is substantial laboratory evidence that without communication coordination failure is common in games such as the minimum game. However, the environments we are interested have ample opportunity for communication and discussion - and in that case, as we indicated, laboratory evidence points at success in coordination.

While minimum games are rather contrived, coordination problem with Pareto inferior equilibria are not. They often arise, for example, in studying political demonstrations against repressive governments. If there is no demonstration nothing happens and the status quo is maintained. If everyone participates the government is overthrown and everyone benefits. On the other hand if there is only one demonstrator the government remains in power and imprisons the demonstrator. Hence it is an equilibrium even for very altruistic people not to demonstrate: the only benefit of demonstrating when everyone else stays home is to go to prison. This is a classic coordination problem.

The coordination idea has been used to analyze the demonstrations that brought an end to the East German government.[1] Although demonstrations could have ended the government at any time, coordination was not achieved until 1989. In fact this is a good case study for the idea of collective decision making. Prior to Gorbachev's non-intervention policy introduced in 1988 the consequence of the fall of the East German government would have been invasion and conquest by the Soviet military as happened in Hungary in 1956 and Czechoslovakia in 1968. Hence, despite the immense unpopularity of the government, it did not make sense (collectively) to bring it down until after the 1988 change in policy - and indeed once that happened not much time elapsed until it was brought down.

These observations should not be taken as a necessary problem with models of altruism: we may add the assumption that when there are multiple equilibria and

[1] See for example Chwe (2000).

ample opportunity for communication and discussion the most efficient equilibrium is always selected.

16.3. A Simple Public Goods Game

To focus thinking we will examine a simple public goods game. There is a single group with N members. The group faces a simple public goods provision problem. Group members i choose effort levels $\phi^i \in [0,1]$ at an individual cost of $D(\phi^i)$, where D is smooth and convex. This effort yields a per capita group benefit of $W(\sum_{i=1}^N \phi^i/N)$ where W is a smooth and strictly concave utility function. Hence per capita social welfare in this economy is

$$W(\sum_{i=1}^N \phi^i/N) - \sum_{i=1}^N D(\phi^i)/N.$$

How does the group induce members to choose ϕ^i?

16.3.1. The Group as Individual. The group behaves as a single individual making the choice that maximizes group utility. Since the objective function is smooth and strictly concave this is the unique solution to the first order conditions. If we let $\varphi = \sum_{i=1}^N \phi^i/N$, we can write this as $W'(\varphi) = D'(\varphi)$ and $\phi^i = \varphi$.

16.3.2. Pure Altruism. An altruist cares only about the impact of his actions on the entire group. That is, each individual i maximizes the social objective with respect to ϕ^i. If we let $\varphi^{-i} = \sum_{j\neq i} \phi^i/N$ each individual chooses a reaction function $\hat{\phi}^i(\varphi^{-i})$ that is a best response to this aggregate. Since utility is smooth and strictly concave with respect to each ϕ^i this is given by $W'(\varphi^{-i} + \hat{\phi}^i/N) = D'(\hat{\phi}^i)$. As these are the same equations as the first order conditions for central control, and since those equations have a unique solution, the equilibrium within a group of altruists maximizes welfare.

16.3.3. Partial Altruism. The model of pure altruism is quite extreme as most people care more about their own utility than that of others. In a model of partial altruism, a weight $1-\lambda$ is given to own utility and λ to group utility. This model also has been widely used in studying voting: for example Schram and Sonnemans (1996), Fowler (2006), Fowler and Ka (2007), Edlin et al. (2007), Faravelli and Walsh (2011), Ozgur (2012), and Jankowski (2007).

Formally a partial altruist has a utility function

$$(1-\lambda)\left[W(\sum_{i=1}^N \phi^i/N) - D(\phi^i)\right] + \lambda\left[NW(\sum_{i=1}^N \phi^i/N) - \sum_{i=1}^N D(\phi^j)\right].$$

To understand this more clearly, observe (again letting $\varphi^{-i} = \sum_{j\neq i} \phi^i/N$) that $(1-\lambda+\lambda N)\,W(\varphi^{-i}) - \lambda\sum_{j\neq i} D(\phi^j)$ does not depend on ϕ^i so from member i's perspective we may subtract it and the objective function becomes

$$(1-\lambda+\lambda N)\left[W(\varphi^{-i}+\phi^i/N) - W(\varphi^{-i})\right] - D(\phi^i).$$

As we are interested in large groups let $N\to\infty$ and observe that this converges to

$$\lambda W'(\varphi)\phi^i - D(\phi^i). \tag{16.3.1}$$

The Nash equilibrium of the resulting game between group members is then the unique solution of the first order conditions. As this is symmetric, for large N we may write it as

$$\lambda W'(\varphi) = D'(\varphi)$$

with $\phi^i = \varphi$.

Obviously when $\lambda = 1$ we have the case of pure altruism. More generally as λ decreases this is equivalent to a model of pure altruism in which the public good is worth less, that is, worth $\lambda W(\sum_{i=1}^N \phi^i)$. Naturally as λ decreases the amount of public good produced by the group declines.

We should point out one undesirable feature of the model of partial altruism. Consider the problem of an innovator who can raise through a clever invention the utility of the entire group. Such an innovator would face an objective function $(1-\lambda)\left(W(\phi^i) - D(\phi^i)\right) + \lambda\left[NW(\phi^i) - D(\phi^i)\right]$ so that the optimal effort level is given by

$$W'(\phi^i) = \frac{1}{1+(N-1)\lambda} D'(\phi^i).$$

This has the unfortunate implication that as $N \to \infty$ the innovator cares only about the public good and not at all about their own utility, and so provides as much effort as feasible. In other words, if altruism is great enough that an individual who can only make an insignificant contribution to the public good is never-the-less motivated to do so, then an individual who can make a significant contribution to the public good must devote themselves entirely to that. We wonder in fact how many people are this altruistic - and find scarce support for this idea in the history of famous innovators.

From an empirical point of view it is doubtful that λ is sufficiently large to matter. First, while altruism towards people we know well may be relatively strong, in the anonymous setting of the laboratory the forces of altruism are relatively weak[2] and, for example, the altruism of farmers towards millions of other farmers who live in other states does not seem likely to be terribly strong.

16.3.4. Social Mechanisms. Suppose now that that violators of the social norm $\phi^i \neq \varphi$ are punished with probability π_1, while non-violators $\phi^i = \varphi$ are punished with probability π. Hence monitoring costs are given by $\theta D(\varphi)$ (where $\theta = \pi/\pi_1$) and the group objective is

$$W(\varphi) - D(\varphi) - \theta D(\varphi).$$

Hence the optimal social norm is given as the unique solution of

$$W'(\varphi) = (1+\theta)D'(\varphi)$$

This is quite similar to partial altruism: indeed partial altruism is equivalent to a peer punishment model in which the monitoring difficulty $\theta = \lambda^{-1} - 1$. When $\theta = 0$ we of course get the same outcome as rule consequentialism or pure altruism. Indeed, social norms may be internalized - in the sense of people punishing themselves for violating the social norm. In this case the signal is perfect since we have no trouble observing ourselves and this is an alternative interpretation of rule consequentialism - a model of monitoring where because monitoring is internal it has no error and hence no cost. Indeed, some of those who have used the partial

[2]Fudenberg and Levine (1997) show, for example, that incomplete learning is considerably more important in the laboratory than altruism.

altruism model have viewed it as kind of a reduced form of a model arising from underlying peer pressure: to quote Esteban and Ray (2011)

> An equivalent (but somewhat looser) view is that α [λ in our notation] is some reduced-form measure of the extent to which within-group monitoring, along with promises and threats, manages to overcome the free-rider problem of individual contribution.

Here we see in a formal sense that this is true in the context of pure public good provision. Hence if we were to redo Esteban et al. (2012) replacing their model of altruism with a model of peer enforcement, we would not learn anything new because we would get the same result. In other contexts, however, peer enforcement is not equivalent to partial altruism: in the model with types, for example, cost must be convex with partial altruism, while we have seen that it can be concave with peer enforcement.

16.3.5. Heterogeneity and Rule Consequentialism. As we have indicated with homogeneous group members rule consequentialism is the same as the group as individual, but unlike that model it allows the possibility that the group is heterogeneous in the sense of some group members not being rule consequentialist. Here we follow Feddersen and Sandroni (2006) in assuming that the group is made up a fraction λ of rule consequentialists and a fraction of $1 - \lambda$ of free-riders (or selfish individuals). Following literally Feddersen and Sandroni (2006) the rule consequentialists all choose the efficient contribution $W'(\hat{\phi}) = D'(\hat{\phi})$ and the free-riders contribute nothing, so that aggregate contributions are $\lambda\hat{\phi}$. In other words, rule consequentialism, like partial altruism and peer enforcement allows the possibility of aggregate contributions that are intermediate in level. (Of course heterogeneity can be introduced into the other models as well.)

In fact there is evidence[3] that the free-riding reduces the contributions of those who do not free-ride. If we follow Coate and Conlin (2004a) we may instead assume that the rule consequentialists care only about other rule consequentialists and not the free riders, in which case they would choose $W'(\hat{\phi}) = (1/\lambda)D'(\hat{\phi})$, which increases as there are more ethical members.

A similar model has been used by Grossman and Helpman (1992). Here some industries organize lobbies and act as single individuals, while other industries are not represented in the political process. As the solution concept is that of the menu auction the only reason the outcome is not efficient is because some groups are not represented. In our world it is as if some groups have $\theta = 0$ and other $\theta = \infty$. This is an extreme but useful assumption.

16.4. Ethnicity and Conflict

An important idea is that for many applications social mechanism theory is backwards compatible with partial altruism. As an illustration of this we re-examine Esteban et al. (2012)'s model of ethnic conflict. Their empirical analysis sheds important light on the role of ethnicity and polarization on the intensity of conflict as well as useful knowledge about the relative importance of fungible and non-rival prizes. Here we show that these conclusions are robust to replacing partial altruism theory with social mechanism theory.

[3] See Fischbacher and Gachter (2010) for a relevant experiment as well as a good overview of this literature.

We start by summarizing their model. There are K groups of size $\eta_k > 0$ with $\sum_{k=1}^{K} \eta_k = 1$. There is a group prize V received by the winning group and a non-rival prize $vu_{k\ell}$ received by group k if ℓ wins with u_{kk} normalized to one and $1 \geq u_{k\ell} \geq 0$. So the winning group will grab all the group prize V, but the decisions they make about rights will generally benefit similar groups. In the empirical analysis it is assumed that $u_{kk} - u_{k\ell} = u_{\ell\ell} - u_{\ell k}$ is the linguistic distance between the two ethnic groups k, ℓ. The conflict resolution function is the simple Tullock function: the probability of group k winning is $p_k(b_1, b_2, \ldots, b_K) = b_k/(\sum_{\ell=1}^{K} b_\ell)$ where b_k is group k's bid. This uncertainty makes sense in the context of civil war and ethnic conflict.

We adopt the social mechanism which as we know in this context is equivalent to the partial altruism model used by Esteban et al. (2012). The objective function of group k is then given by

$$p_k V + \eta_k \sum_{\ell=1}^{K} p_\ell v u_{k\ell} - \eta_k (1+\theta) C(b_k/\eta_k).$$

This is concave in the bid, so group optima are given by first order conditions. Computing these conditions and aggregating over groups with weights η_k and re-arranging we get a single equation that is necessary for an equilibrium:

$$\begin{aligned}\frac{(C'(b_k/\eta_k)}{V+v} = \frac{1}{1+\theta} \sum_{k=1}^{K} \Bigg[& \left(\frac{1}{\sum_{\ell=1}^{K} b_\ell} - \frac{b_k}{(\sum_{\ell=1}^{K} b_\ell)^2} \right) \eta_k \frac{V}{V+v} \\ & + \frac{\eta_k^2}{\sum_{\ell=1}^{K} b_\ell} v u_{kk} - \eta_k^2 \sum_{\ell=1}^{K} \frac{b_\ell}{(\sum_{\ell=1}^{K} b_\ell)^2} \frac{v}{V+v} u_{k\ell} \Bigg].\end{aligned}$$

In Esteban and Ray (2011) computations are given that show that under reasonable conditions the equilibrium is approximately the same as would be the case if bids are proportional to group size, that is, $b_k = \beta\eta_k$. Plugging this in we get

$$\frac{\beta C'(\beta)}{V+v} = \frac{1}{1+\theta} \left(\left[\sum_{k=1}^{K} \eta_k (1-\eta_k) \right] \frac{V}{V+v} + \left[\sum_{k=1}^{K} \sum_{\ell=1}^{K} \eta_k^2 \eta_\ell (u_{kk} - u_{k\ell}) \right] \frac{v}{V+v} \right).$$

Esteban et al. (2012) offer the following interpretation of this condition:

a.$\beta C'(\beta)/(V+v)$ is the cost incurred by each group relative to the stakes; it is a measure the *intensity of conflict*

b. $v/(V+v)$ is the importance of the non-rival prize; as the non-rival prize has a public component, this is called the *publicness* of the prize

c. $\mathcal{F} = \sum_{k=1}^{K} \eta_k (1-\eta_k)$ measures *ethnic fractionalization*; if all groups are the same size this increases with K

d. $\mathcal{P} = \sum_{k=1}^{K} \sum_{\ell=1}^{K} \eta_k^2 \eta_\ell (u_{kk} - u_{k\ell})$ is the *polarization index*; greater distance between groups objectives as measured by $u_{kk} - u_{k\ell}$ increases polarization.

With this setup Esteban et al. (2012) gather data on 800 groups in 138 countries from 1960 to 2008 over five year sub-periods. They assume that θ and the publicness of the prize are constant across countries. They measure the intensity of conflict with a $0, 1$ variable based on battle deaths and use this as the dependent variable to estimate the equation above together with some controls for economic, geographic and political conditions. They find (standard error in parentheses) the coefficient on fractionalization $\mathcal{F}$ to be 6.5 (0.004) and that on polarization $\mathcal{P}$ to be 1.25 (0.020). This indicates that both prizes are important but that group prize is more important: indeed V/v is given by the ratio of the coefficients, so is estimated to be equal to 5.2, meaning that the group prize is about five times as important as the non-rival prize.

From other evidence Esteban et al. (2012) argue that θ is close to zero. In their setup of partial altruism this is the same as pure altruism - about which we are skeptical - but it is plausible that for military service from poor villages monitoring is not terribly difficult. We do not expect that hiring doctors to report bones spurs is likely to excuse an individual from military service as it might in a more developed setting.

Natural resources entered as a control do not do well in their estimation, but they argue that higher importance of the group prize (the coefficient on $\mathcal{F}$) may proxy for natural resources and they do further empirical analysis of a subsample where they have additional data about oil reserves that supports this conclusion. The bottom line of their study:

> Finally, we point to the quantitative importance of polarization and fractionalization in conflict... Our estimated coefficients imply that if we move from the 20th percentile of polarization to the 80th percentile, holding all other variables at their means, the probability of conflict rises from approximately 13 percent to 29 percent. Performing the same exercise for $\mathcal{F}$ takes us from 12 percent to 25 percent. These are similar (and strong) effects.

16.5. Cartels

The case of cartels leads to a strong distinction between the social mechanism model and the partial altruism model. While equivalent in the case of public goods, they are not equivalent here. Indeed the social mechanism model correctly predicts that unions will both lobby and cartelize, while farmers will only lobby; the partial altruism model predicts that any industry that lobbies will also cartelize.

To see this we work out the implications of the partial altruism model in the cartel problem. Recall from section 2.2 that $\mu(\varphi)$ is the price cost margin and per capita profits are $\mu(\varphi)\varphi$, with derivative $\mu'(\varphi)\varphi+\mu(\varphi)$ with φ denoting average output. Recall from 16.3.3 equation 16.3.1 that with partial altruism and N large we can write the objective function as $\lambda W'(\varphi)\phi^i - D(\phi^i)$. In the present case we take benefits net of production costs so that $W(\varphi) = \mu(\varphi)\varphi$. The "cost" of individual action $D(\phi^i)$ is then the negative of profit from individual output (here the individual action is denoted by x^i rather than ϕ^i), that is to say $-\mu(\varphi)x^i$. Hence the objective function can be written as $\lambda\left[\mu'(\varphi)\varphi + \mu(\varphi)\right]x^i + \mu(\varphi)x^i$. Since this is linear in x^i the equilibrium requires that firms be indifferent, that is, $\lambda\left[\mu'(\varphi)\varphi + \mu(\varphi)\right] + \mu(\varphi) = 0$ which can be written as

$$-\frac{\mu(\varphi)}{\mu'(\varphi)\varphi} = \frac{\lambda}{1+\lambda}.$$

Since $\mu'(\varphi_k)\varphi_k$ is by assumption strictly negative, if $\lambda = 0$ (pure selfishness) then $\mu(\varphi) = 0$ in which case the equilibrium is the competitive one and the cartel does not form. From the elasticity assumption $-\mu(\varphi)/\mu'(\varphi)$ is decreasing, so the same is true of the LHS. Hence we have $\varphi(\lambda)$ is strictly decreasing, meaning that the cartel forms if there is any altruism at all, and the strength of the quota is entirely determined by the level of altruism. In particular the capacity constraint X plays no role, and we are left with a rather vacuous theory in which unions cartelize because union members are altruistic while farmers do not because they are not altruistic. Worse, since altruism also determines public good contributions, we are then left with the counterfactual conclusion that farmers do not lobby!

The point here is that the social mechanism model captures the fact that greater capacity leads to a greater temptation to cheat on the cartel hence increases enforcement costs. Partial altruism does not account for the temptation to cheat and misses this effect.

CHAPTER 17

Conclusion

We have studied exogenous *ex ante* homogeneous groups with perfect communication. As can be seen this is a subject that is by no means trivial. We have presented evidence that this theory is correct and useful. In our conclusion we briefly review what that evidence is, then wrap up by discussing the many things that still need to be done.

17.1. Where Are We?

The fact that our theory is backwards compatible with existing theories is important because it means that evidence for those theories is evidence for ours as well. Hence empirical studies such as Coate and Conlin (2004b) on voting in liquor referenda or Esteban et al. (2012) on conflict in Africa support our theory, and indeed our theory gives a common framework from which the theories examined in those studies emerge as special cases. Similarly, a key element of our theory is to explain why small groups are able to overcome large groups in lobbying: there is a large empirical literature on this, especially the work of Mancur (1965) and Olson (1982). Indeed, as our theory of lobbying predicts, we gave evidence that farmers are more effective at lobbying the fewer of them there are.

We should emphasize as well that the underlying details of the theory of social mechanisms - monitoring and peer punishment - are well supported, in particular by the work of Ostrom (1990), but also empirical studies of voting, such as Della Vigna et al. (2014) and many other studies we discuss in our analysis of voting.

The point of a theory, however, is not merely to integrate existing work into a common framework, but to develop new insights. Our theory provides nuances missed by the simple observation that when it comes to lobbying "small is beautiful." We argued that small groups can be successful at lobbying over fungible stakes, such as financial favors from government. We showed as well that they will fail if they are too greedy or if they lobby over broad issues. We saw why it is that small groups are not effective at lobbying over issues such as rights, but must instead operate by persuasion to convince a majority to support them. We provided evidence that indeed minorities only succeed when they convince majority opinion, and gave examples of lobbying that failed because of excessive greed or broad issues.

We studied the connection between lobbying - subverting the political system - and cartelization - subverting the market. We gave evidence that monitoring cost can explain why groups such as service workers are ineffective at both lobbying and cartelization due to the lack of ability to effectively monitor, while groups such as farmers or manufacturers are effective at lobbying but fail at cartelization due to the ability of the individual firms to ramp up production to take advantage of positive-price cost margins. Finally, we argued that groups with effective monitoring and inelastic supply, as is the case with many trade unions, should be effective at both lobbying and cartelization. We provided supporting evidence for these conclusions.

By understanding the role of monitoring cost we understand which types groups are successful in different types of political contests - voting as well as lobbying. We argued that large groups are generally successful at voting - but also explained why groups such as school boards may win elections despite being small. We explained as well why voter suppression is most popular when it is least needed to win elections - and that the pattern of voter suppression laws supports this. And we showed why it is that the outcome of elections is necessarily uncertain, while that of lobbying need not be.

Our theory of punishment explains why fines are better than ostracism - and so why it is that cartels use fines rather than price wars to enforce quotas and price limits. When we add adjustment costs to the mix we find out why it is that negative demand shocks can increase cartel output - as happened recently with OPEC, as well as why NGOs often reduce rather than increase the output of public goods, or why fining parents for being late to day-care may cause them to arrive even later rather than earlier, and why lowering fines on rural highways led to a near revolution in France. A diverse array of real life events and phenomenon are seen through the lenses of our theory to have a common explanation.

We examined also why social change is sometimes exceptionally rapid and other times painfully slow - and relate this to the "Olympic effect" where positive publicity drives real institutional reform.

For the most part these are qualitative empirics. Yet it is important not just that a theory be useful quantitatively - as it is in the studies of Coate and Conlin (2004b) and Esteban et al. (2012) - but that it explain a broad range of facts. Scientific papers by their nature are narrow so do not provide a forum for giving broad ranging evidence. It is our hope that by putting together in a book the broad range of facts explained by the theory of social mechanisms the power of the theory emerges.

17.2. Where Are We Going?

As indicated, we have studied exogenous *ex ante* homogeneous groups with perfect communication. Under our assumptions it is reasonable to assume as we do that a consensus is reached on the optimal mechanism. For many groups our assumptions are plausible: for example farmers become farmers because they have a relative advantage in farming and they want to farm. Their social ties and networks come about because of common work and because they live in the same towns and they socialize with each other. For the most part farmers do not choose to farm because they want to lobby for farm subsidies. Living together communications are good and also being brought together by common interest in farming the group may be thought of to a good approximation as exogenously formed and homogeneous. The same can be said of many other special interest groups.

Never-the-less groups are heterogeneous, membership is endogenous, communication is costly and imperfect, and consensus is not necessarily quick or easy, or even successful. We do not currently know how to extend our model to allow for imperfect or failed consensus, but we believe that heterogeneity and endogeneity are related in an important way. We do not think it controversial to suggest that greater heterogeneity makes consensus more difficult. Heterogeneous groups may then splinter: to understand how this works we need also to understand how group membership is determined endogenously.

This brings us to the issue of communication. There is a large literature about the communications of facts: what are the consequences of a particular policy? How is information distorted and what are the consequences? There are however three other important reasons for communication: coordination, conformity, and communicating norms. Coordination seems obvious and is perhaps not so hard to model: if we go on strike we should all do so at the same time, and we need to communicate that "today is the day."

Communication about norms is important since a social norm is meaningless unless people know what it is. Moreover, communication of social norms might change preferences. Much has been written about conformity without perhaps a clear understanding of why conformity is important to a group. If heterogeneity leads to failed consensus then a group is stronger when members' preferences can be brought into greater alignment. If we can agree that abortion is evil or that free speech is good it will be much easier to reach a consensus over political action. Communication that is widely viewed as distorted - fake news, myths, virtue signaling and the like - can play an important role both in communicating social norms and in creating pressure for conformity.

Education is a case in point. While not studied in the literature on distorted communication of facts, history as taught in mandatory schools throughout the world contains important elements of myth, and is selective in its coverage. Why do Greeks speak today of Alexander the Great, but he has little place in history books of other countries? How many countries have invented the airplane? And not just history: what purpose does teaching Latin in schools in Italy serve other than to remind Italians that they are all descendants of the Roman Empire, however factually wrong and irrelevant that might be? That education is also indoctrination in social norms - nationalism and patriotism in particular, but not solely - and in "right ways" of thinking may not be central to current economic research, but it is very much on the minds of school boards fighting over how to present US history. Is it the story of noble pioneers opening a new continent? Or of evil slavers and the genocide of the native Americans? It very much on the minds of environmentalists who have had great success: what Western school teacher does not indoctrinate their students in the benefits of being "green?"

When we post memes, whether they be about the evils of abortion or the right to govern our own bodies, we are communicating both social norms and creating pressure for conformist preferences. We hope that it will be possible to build on the exogenous *ex ante* homogeneous theory developed here to incorporate endogeneity, heterogeneity, and communication so that we can begin to address these issues.

Bibliography

Abeler, J., D. Nosenzo, and C. Raymond (2019): "Preferences for truth-telling," *Econometrica*, 87, 1115–1153.

Abreu, D., D. Pearce, and E. Stacchetti (1990): "Toward a theory of discounted repeated games with imperfect monitoring," *Econometrica*, 1041–1063.

Acemoglu, D. and J. A. Robinson (2001): "A theory of political transitions," *American Economic Review*, 938–963.

Acemoglu, D. and A. Wolitzky (2015): "Sustaining cooperation: Community enforcement vs. specialized enforcement," *National Bureau of Economic Research*, 21457.

Akerlof, G. and R. E. Kranton (2005): "Identity and the Economics of Organizations," *Journal of Economic Perspectives*.

Alcalde, J. and M. Dahm (2007): "Tullock and Hirshleifer: a meeting of the minds," *Review of Economic Design*, 11, 101–124.

Ali, S. N. and C. Lin (2013): "Why People Vote: Ethical Motives and Social Incentives," *American Economic Journal: Macroeconomics*, 5, 73–98.

Amat, F., C. Boix, J. Munoz, and T. Rodon (2018): "From Political Mobilization to Electoral Participation: Turnout in Barcelona in the 1930s," Mimeo.

Amegashie, J. A. (2006): "A contest success function with a tractable noise parameter," *Public Choice*, 126, 135–144.

Ashworth, S. and E. B. D. Mesquita (2009): "Elections with platform and valence competition," *Games and Economic Behavior*, 67, 191–216.

Bano, M. (1973): *Breakdown in Pakistan : how aid is eroding institutions for collective action*, Stanford University Press.

Bar-Adon, A. (1972): "Sy Agnon and the Revival of Modern Hebrew," *Texas Studies in Literature and Language*, 14, 147–175.

Baye, M. R., D. Kovenock, and C. G. C. G. De Vries (1993): "Rigging the Lobbying Process: an Application of the All-pay Auction," *The American Economic Review*, 83, 289–294.

Becker, G. S. (1983): "A theory of competition among pressure groups for political influence," *Quarterly Journal of Economics*, 98, 371–400.

Belloc, M., F. Drago, and R. Galbiati (2016): "Earthquakes, religion, and transition to self-government in Italian cities," *The Quarterly Journal of Economics*, 131, 1875–1926.

Benabou, R. and J. Tirole (2006): """ Incentives and prosocial behavior,"," *The American Economic Review*, 96, 1652–1678.

Bigoni, M., S. Bortolotti, M. Casari, D. Gambetta, and F. Pancotto (2013): *Cooperation Hidden Frontiers: The Behavioral Foundations of the Italian North-South Divide*, University of Bologna.

Blume, A. and A. Ortmann (2007): "The effects of costless pre-play communication: Experimental evidence from games with Pareto-ranked equilibria," *Journal of Economic Theory*, 132, 274–290.

Boldrin, M. and D. K. Levine (2005): "The economics of ideas and intellectual property," *Proceedings of the National Academy of Sciences of the United States of America*, 102, 1252–1256.

Boyer, P. C., T. Delemotte, G. Gauthier, V. Rollet, and B. Schmutz (2019): "Les dÃ©terminants de la mobilisation des 'gilets jaunes'," Tech. rep., Working Papers 2019-06, Center for Research in Economics and Statistics.

Brandts, J., D. J. Cooper, and R. A. Weber (2015): "Legitimacy, communication, and leadership in the turnaround game," *Management Science*, 61, 2627–2645.

Brinker, P. (1985): "Violence by U.S. labor unions," *Journal of Labor Research*, 6, 417–427.

Calvo, G. A. (1983): "Staggered prices in a utility-maximizing framework," *Journal of monetary Economics*, 12, 383–398.

Chamberlain, G. and M. Rotschild (1981): "A Note on the Probability of Casting a Decisive Vote," *Journal of Economic Theory*, 25, 152–162.

Chari, V. V. and L. E. Jones (2000): "A reconsideration of the problem of social cost: Free riders and monopolists," *Economic Theory*, 16, 1–22.

Charness, G. M. D. (2006): "Promises and partnership," *Econometrica*, 74, 1579–1601.

Che, Y.-K. and I. Gale (2000): "Difference-form contests and the robustness of all-pay auctions," *Games and Economic Behavior*, 3, 22–43.

Chwe, M. S. Y. (2000): "Communication and coordination in social networks," *The Review of Economic Studies*, 67, 1–16.

Coase, R. H. (1960): "The Problem of Social Cost," *Journal of Law and Economics*, 3, 1–44.

Coate, S. and M. Conlin (2004a): "A group rule-utilitarian approach to voter turnout: theory and evidence," *American Economic Review*, 94, 1476–1504.

——— (2004b): "A Group Rule-Utilitarian Approach to Voter Turnout: Theory and Evidence," *American Economic Review*, 94, 1476–1504.

Coate, S., M. Conlin, and A. Moro (2008): "The Performance of Pivotal-Voter Model in Small-Scale Elections: Evidence from Texas Liquor Referenda," *Journal of Public Economics*, 92, 582–596.

Coleman, J. S. (1988): "Social capital in the creation of human capital," *American Journal of Sociology*, S95–S120.

Cooper, R., D. V. DeJong, R. Forsythe, and T. W. Ross (1992): "Communication in coordination games," *Quarterly Journal of Economics*, 107, 739–771.

Crawford, V. (1998): "A survey of experiments on communication via cheap talk," *Journal of Economic Theory*, 78, 286–298.

Crawford, V. P. and J. Sobel (1982): "Strategic information transmission," *Econometrica*, 1431–1451.

Dasgupta, P. and E. Maskin (1986): "The existence of equilibrium in discontinuous economic games, I: Theory," *Review of Economic Studies*, 53, 1–26.

d'Aspremont, C., J. Cremer, and L.-A. Gerard-Varet (2003): "Correlation, independence, and Bayesian incentives," *Social Choice and Welfare*, 21, 281–310.

Dell, M., N. Lane, and P. Querubin (2018): "The historical state, local collective action, and economic development in Vietnam," *Econometrica*, 86, 2083–2121.

Della Vigna, S., J. A. List, U. Malmendier, and G. Rao (2014): "Voting to Tell Others," NBER Working Paper.

Devetag, G. and A. Ortmann (2007): "When and why? A critical survey on coordination failure in the laboratory," *Experimental Economics*, 10, 331–344.

Docherty, J. C. and S. van der Velden (2012): *Historical Dictionary of Organized Labor*, Scarecrow Press.

Donnelly, S. and T. Inglis (2010): "The media and the Catholic Church in Ireland: Reporting clerical child sex abuse," *Journal of Contemporary Religion*, 25, 1–19.

Dutta, R., D. K. Levine, and S. Modica (2018): "Collusion Constrained Equilibrium," *Theoretical Economics*, 13, 307–340.

——— (2021): "Interventions with Sticky Social Norms: A Critique," *Journal of the European Economic Association*, forthcoming.

——— (2022): "Failing to Provide Public Goods: Why the Afghan Army did not Fight," *Federal Reserve Bank of St. Louis Review*, forthcoming.

Dye, R. A. (1985): "Costly contract contingencies," *International Economic Review*, 26, 233–250.

Edlin, A., A. Gelman, and N. Kaplan (2007): "Voting as a Ra tional Choice: Why and How People Vote to Improve the Well-being of Others," *Rationality and Society*, 19, 293–314.

Ekmekci, M., O. Gossner, and A. Wilson (2012): "Impermanent types and permanent reputations," *Journal of Economic Theory*, 147, 162–178.

Ellison, G. (1994): "Cooperation in the prisoner's dilemma with anonymous random matching," *The Review of Economic Studies*, 61, 567–588.

Ergin, H. and T. Sarver (2010): "A unique costly contemplation representation," *Econometrica*, 78, 1285–1339.

Esteban, J., L. Mayoral, and D. Ray (2012): *Ethnicity and conflict: An empirical study*, American Economic Review.

Esteban, J. and D. Ray (2001): "Collective action and the group size paradox," *American Political Science Association*, 3, 663–672.

——— (2011): "Linking Conflict to Inequality and Polarization," *American Economic Review*, 101, 1345–1374.

Ewerhart, C. (2015): "Mixed equilibria in Tullock contests," *Economic Theory*, 60, 59–71.

——— (2017): "Revenue ranking of optimally biased contests: The case of two players," *Economics Letters*, 157, 167–170.

Ewerhart, C. and G. Z. Sun (2018): "Equilibrium in the symmetric two-player Hirshleifer contest: Uniqueness and characterization," *Economics Letters*, 169, 51–54.

Ewerhart, C. and K. Valkanova (2020): "Fictitious play in networks," *Games and Economic Behavior*, 123, 182–206.

Faravelli, M. and R. Walsh (2011): "Smooth Politicians and Paternalistic voters: a theory of large elections," *National Bureau of Economic Research*, 1739.

Feddersen, T. and A. Sandroni (2006): "A theory of participation in elections," *American Economic Review*, 96, 1271–1282.

FENG, X. AND J. LU (2017): "Uniqueness of equilibrium in two-player asymmetric Tullock contests with intermediate discriminatory power," *Economics Letters*, 159, 61–64.

FINK, N., P. SCHMIDT-DENGLER, K. STAHL, AND C. ZULEHNER (2017): "Registered artels in Austria overview and governance structure," *European Journal of Law and Economics*, 44, 385–422.

FISCHBACHER, U. AND S. GACHTER (2010): "Social preferences, beliefs, and the dynamics of free riding in public goods experiments," *American Economic Review*, 100, 541–56.

FOWLER, J. H. (2006): "Altruism and Turnout," *Journal of Politics*, 68, 674–683.

FOWLER, J. H. AND C. D. KA (2007): "Beyond the self: Social identity, altruism, and political participation," *Journal of Politics*, 69, 813–827.

FRASS, A. AND D. GREER (1977): "Market structure and price Collusion: an empirical analysis," *Journal of Industrial Economics*, 26, 21–44.

FU, T., Y. JI, K. KAMEI, AND L. PUTTERMAN (2017): "Punishment can support cooperation even when punishable," *Economics Letters*, 154, 84–87.

FUDENBERG, D. AND D. K. LEVINE (1989): "Reputation and Equilibrium Selection in Games with a Single Long-Run Player," *Econometrica*, 57.

——— (1992): "Maintaining a Reputation when Strategies are Imperfectly Observed," *Review of Economic Studies*, 59, 561–579.

——— (1997): "Measuring Subject's Losses in Experimental Games," *Quarterly Journal of Economics*, 112, 508–536.

——— (2006): "Superstition and Rational Learning," *American Economic Review*, 96, 630–651.

——— (2007): "Continuous Time Limits of Repeated Games with Imperfect Public Monitoring," *Review of Economic Dynamics*, 10, 173–192.

FUDENBERG, D., D. K. LEVINE, AND E. MASKIN (1995): "Balanced-budget mechanisms with incomplete information," UCLA Department of Economics, DK Levine's Working Papers Archive.

FUDENBERG, D., D. K. LEVINE, AND W. PESENDORFER (1998): "When are nonanonymous players negligible?" *Journal of Economic Theory*, 79, 46–71.

FUDENBERG, D. AND E. MASKIN (1994): "The Folk Theorem with Imperfect Public Information," *Econometrica*, 62, 997–1039.

GALIANI, S. AND G. TORRENS (2014): "Autocracy, democracy and trade policy," *Journal of International Economics*, 93, 173–193.

GENESOVE, D. AND W. P. MULLIN (2001): "Rules, communication, and collusion: Narrative evidence from the Sugar Institute case," *American Economic Review*, 91, 379–398.

GERBER, A. S., D. P. GREEN, AND C. W. LARIMER (2018): "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment," *American Political Science Review*, 102, 33–48.

GLICKSBERG, I. (1952): "A Further Generalization of the Kakutani Fixed Point Theorem, with Application to Nash Equilibrium Points," *Proceedings of the American Mathematical Society*, 3, 170–174.

GNEEZY, U. AND A. RUSTICHINI (2000): "A fine is a price," *The Journal of Legal Studies*, 29, 1–17.

GRAY, M. AND M. CAUL (2000): "Declining Voter Turnout in Advanced Industrial Democracies, 1950 to 1997," *Comparative Political Studies*, 33, 1091–1122.

Green, E. J. (1980): "Noncooperative Price Taking in Large Dynamic Markets," *Journal of Economic Theory*, 22, 155–181.

Green, E. J. and R. H. Porter (1984): "Noncooperative collusion under imperfect price information," *Econometrica*, 87–100.

Grossman, G. M. and E. Helpman (1992): "Protection for sale," *National Bureau of Economic Research*, 4149.

Group, C. J. R. (2000): Medical Malpractice by the Numbers.".

Hajnal, Z., N. Lajevardi, and L. Nielson (2016): *Voter Identification Laws and the Suppression of Minority Votes*, UCSD.

Harrington, J. E. and A. Skrzypacz (2011): "Private monitoring and communication in cartels: Explaining recent collusive practices," *American Economic Review*, 101, 2425–2449.

Harsanyi, J. C. (1973): "Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points," *International Journal of Game Theory*, 2, 1–23.

——— (1977): "Morality and the theory of rational behavior," *Social Research*, 623–656.

Hart, O. and J. Moore (1988): "Incomplete contracts and renegotiation," *Econometrica*, 56, 755–785.

Haucap, J., U. Heimeshoff, and L. M. Schultz (2010): "Legal and Illegal Cartels in Germany Between 1958 and 2004," Tech. rep., No 08, DICE Discussion Papers, University of Dusseldorf.

Hay, G. and D. Kelley (1974): "An empirical survey of price fixing conspiracies," *Journal of Law and Economics*, 17, 13–38.

Heller, M. (2008): *The Gridlock Economy: How Too Much Ownership Wrecks Markets, Stops Innovation, and Costs Lives*, Basic Books.

Herrera, H., D. K. Levine, and C. Martinelli (2008): "Policy Platforms, Campaign Spending and Voter Participation," *Journal of Public Economics*, 92, 501–513.

Herrera, H., M. Morelli, and S. Nunnari (2015): *Turnout Across Democracies*, American Journal of Political Science.

Hillman, A. L. and J. G. Riley (1989): "Politically contestable rents and transfers," *Economics and Politics*, 1, 17–39.

Hirshleifer, J. (1989): "Conflict and Rent-Seeking Success Functions: Ratio vs. Difference Models of Relative Success," *Public Choice*, 63, 101–112.

Hooker, B. (2011): *Rule Consequentialism*, The Stanford Encyclopedia of Philosophy.

Hyytinen, A., F. Steen, and O. Toivanen (2019): "An anatomy of contracts," *Economic Cartel Journal*, forthcoming.

Ingraham, C. (2016): "New evidence that voter ID laws 'skew democracy' in favor of white Republicans," *Washington Post*, February 14.

Jackson, M. O. (2010): *Social and economic networks*, Princeton: university press.

Jankowski, R. (2007): "Altruism and the Decision to Vote Explaining and Testing High Voter Turnout," *Rationality and Society*, 19, 5–34.

Jansen, M. B. (2002): *The making of modern Japan*, Harvard University Press.

Jia, H., S. Skaperdas, and S. Vaidya (2013): "Contest functions: Theoretical foundations and issues in estimation," *International Journal of Industrial*

Organization, 31, 211–222.

KAKUTANI, S. (1941): "A generalization of Brouwer's fixed point theorem," *Duke Mathematical Journal*, 8, 457–459.

KANDORI, M. (1992): "Social Norms and Community Enforcement," *The Review of Economic Studies*, 59, 63–80.

KINCHEN, K. S., L. A. COOPER, D. LEVINE, N. Y. WANG, AND N. R. POWE (2004): "Referral of patients to specialists: factors affecting choice of specialist by primary care physicians," *The Annals of Family Medicine*, 2, 245–252.

KNACK, S. (1992): "Civic Norms, Social Sanctions, and Vote Turnout," *Rationality and Society*, 4, 133–156.

KRAUT, A. I. (1975): "Prediction of managerial success by peer and training-staff ratings," *Journal of Applied Psychology*, 60, 14.

KREPS, D. M. AND R. WILSON (1982): "Sequential Equilibria," *Econometrica*, 863–894.

LANGTON, L. AND M. DUROSE (2013): *Police Behavior during Traffic and Street Stops*, Bureau of Justice Statistics.

LEVENSTEIN, M. C. AND V. Y. SUSLOW (2006): "What determines cartel success?" *Journal of Economic Literature*, 44, 43–95.

LEVINE, D. AND S. MODICA (2016): "Peer Discipline and Incentives within Groups," *Journal of Economic Behavior and Organization*, 123, 19–30.

LEVINE, D. AND T. PALFREY (2007): "The Paradox of Voter Participation? A Laboratory Study," *American Political Science Review*, 101, 143–158.

LEVINE, D. K. (2012): *Is behavioral economics doomed?: The ordinary versus the extraordinary*, Open Book Publishers.

——— (2021a): "Fine Cartels," *Economic Theory Bulletin*, forthcoming.

——— (2021b): "The Reputation Trap," *Econometrica*, forthcoming.

LEVINE, D. K. AND A. MATTOZZI (2019): *Success in Contests*, EUI.

——— (2020): "Voter Turnout with Peer Punishment," *American Economic Review*, 110, 3298–3314.

LEVINE, D. K., A. MATTOZZI, AND S. MODICA (2020): "Trade Associations: Why Not Cartels?" *International Economic Review*, forthcoming.

LEVINE, D. K. AND S. MODICA (2017): "Size, Fungibility, and the Strength of Organizations," *European Journal of Political Economy*, 49, 71–83.

LEVINE, D. K. AND W. PESENDORFER (1995): "When are agents negligible?" *American Economic Review*.

LEVY, P. (2016): "North Carolina Republicans Make Brazen Bid for Permanent Power After Losing Governor's Race," *Mother Jones*, December 15.

LI, M. AND D. MAJUMDAR (2010): "A psychologically based mo del of voter turnout," *Journal of Public Economic Theory*, 12, 979–1002.

LIJPHART, A. (2012): *Patterns of democracy: Government forms and performance in thirty-six countries*, Yale University Press.

MAILATH, G. J. AND L. SAMUELSON (2001): "Who wants a good reputation?" *The Review of Economic Studies*, 68, 415–441.

MANCUR (1965): *The Logic of collective action: public goods and the theory of groups*, Harvard Economic Studies.

MATHESON, V. (2006): "Mega-Events: The effect of the world" s biggest sporting events on local, regional, and national economies," Mimeo, Holy Cross.

MILGROM, P. AND J. ROBERTS (1982): "Predation, reputation, and entry deterrence," *Journal of Economic Theory*, 27, 280–312.

MODICA, S. AND A. RUSTICHINI (1994): "Awareness and partitional information structures," *Theory and Decision*, 37, 107–124.

MOE, T. M. (2003): *Teacher Unions and School Board Elections*, Hoover Institute.

——— (2006): "Political Control and the Power of the Agent," *Journal of Law, Economics, and Organization*, 22, 1–29.

MULLIGAN, C. B. AND C. G. HUNTER (2003): "The Empirical Frequency of a Pivotal Vote," *Public Choice*, 116, 31–54.

MYERSON, R. B. (1982): "Optimal coordination mechanisms in generalized principal-agent problems," *Journal of Mathematical Economics*, 10, 67–81.

——— (1998): "Population uncertainty and Poisson games," *International Journal of Game Theory*, 27, 375–392.

MYERSON, R. B. AND M. A. SATTERTHWAITE (1983): "Efficient mechanisms for bilateral trading," *Journal of Economic Theory*, 29, 265–281.

NAGOURNEY, A., A. PARKER, J. RUTENBERG, AND J. ZELENYNOV (2012): "How a Race in the Balance Went to Obama," *New York Times*, 7.

NITZAN, S. AND K. UEDA (2011): "Prize sharing in collective contests," *European Economic Review*, 55, 678–687.

NTI, K. O. (1999): "Rent-seeking with asymmetric valuations," *Public Choice*, 98, 415–430.

——— (2004): "Maximum efforts in contests with asymmetric valuations," *European Journal of Political Economy*, 20, 1059–1066.

OLSON, M. (1982): *The Rise and Decline of Nations: Economic Growth, Stagflation, and Social Rigidities*, Yale University Press.

ORDONEZ-DE-HARO, J., R. BORRELL, AND J. L. JIMENEZ (2018): "The European Commission's Fight Against Cartels (1962 - 2014): A Retrospective and Forensic Analysis," *Journal of Common Market Studies*, 56, 1087–1107.

OSTROM, E. (1990): *Governing the commons: The evolution of institutions for collective action*, Cambridge: university press.

OZGUR, E. (2012): "Altruism and Voting: A Large-turnout Result that Does not Rely on Civic Duty or Cooperative Behavior," *Journal of Economic Theory*, 147, 2124–2157.

PALFREY, T. R. AND H. ROSENTHAL (1985): "Voter Participation and Strategic Uncertainty," *American Political Science Review*, 79, 62–78.

PENROSE, L. S. (1946): "The Elementary Statistics of Majority Voting," *Journal of the Royal Statistical Society*, 109, 53–57.

POSNER, R. (1970): "A statistical study of antitrust enforcement," *Journal of Law and Economics*, 13, 365–419.

PROKOPOVYCH, P. AND N. C. YANNELIS (2014): "On the existence of mixed strategy Nash equilibria," *Journal of Mathematical Economics*, 52, 87–97.

RADNER, R. (1980): "Collusive behavior in noncooperative epsilon-equilibria of oligopolies with long but finite lives," *Journal of Economic Theory*, 22, 136–154.

ROEMER, J. (2010): "Kantian Equilibrium," *Scandinavian Journal of Economics*, 112, 1–24.

ROSE, A. K. AND M. M. SPIEGEL (2011): "The Olympic Effect," *Economic Journal*, 121, 652–677.

Rotemberg, J. and G. Saloner (1986): "A supergame-theoretic model of price wars during booms," *American Economic Review*, 76, 390–407.

Sabourian, H. (1990): "Anonymous Repeated Games with a Large Number of Players and Random Outcomes," *Journal of Economic Theory*, 51, 92–110.

Sarsons, H. (2017): "Interpreting Signals: Evidence from Doctor Referrals," Working Paper.

Schram, A. and J. Sonnemans (1996): "Why people vote: Experimental evidence," *Journal of Economic Psychology*, 17, 417–442.

Shachar, R. and B. Nalebuff (1999): "Follow the Leader: Theory and Evidence on Political Participation," *American Economic Review*, 89, 525–547.

Siegel, R. (2014): "Asymmetric Contests with Head Starts and Nonmonotonic Costs," *American Economic Journal: Microeconomics*, 6, 59–105.

Sims, C. A. (2003): "Implications of rational inattention," *Journal of monetary Economics*, 50, 665–690.

Siragusa, M., M. Beretta, and M. Bay (2009): *Competition Law in Italy: The first 20 years of law and practice*, Cleary Gottleib.

Skarbek, D. (2014): *The social order of the underworld: How prison gangs govern the American penal system*, Oxford University Press.

Sugaya, T. and A. Wolitzky (2021): "Repeated Games with Many Players," mimeo, Stanford GSP.

Suslow, V. Y. (2005): "Cartel inter-war international contract duration: empirical evidence from cartels," *Industrial and Corporate Change*, 14, 705–44.

Tirole, J. (2009): "Cognition and incomplete contracts," *American Economic Review*, 99, 265–94.

Tisserand, J. C., F. Cochard, and J. Le Gallo (2015): *Altruistic or Stategic Considerations: A Meta-Analysis on the Ultimatum and Dictator Games*, Besancon: CRESE, Universite de Franche-Comte.

Townsend, R. M. (1994): "Risk and Insurance in Village India," *Econometrica*, 62, 539–591.

Tullock, G. (1967): "The welfare costs of tariffs, monopolies, and theft," *Western Economic Journal*, 5, 224–232.

Van Huyck, J. B., R. C. Battalio, and R. O. Beil (1990): "Tacit coordination games, strategic uncertainty, and coordination failure," *The American Economic Review*, 234–248.

Walker, M. and J. Wooders (2001): "Minimax play at Wimbledon," *American Economic Review*, 91, 1521–1538.

Wang, Z. (2010): "The optimal accuracy level in asymmetric contests," *B.E. Journal of Theoretical Economics*, 10.

Whitney, H. (1934): "Analytic extensions of differentiable functions defined in closed sets," *Transactions of the American Mathematical Society*, 36, 63–89.

Williams, J. C. (2016): *What So Many People Don't Get About the U.S. Working Class*, Harvard Business Review.

Wolitzky, A. (2013): "Cooperation with network monitoring," *Review of Economic Studies*, 80, 395–427.

Yang, M. (2015): "Coordination with flexible information acquisition," *Journal of Economic Theory*, 158, 721–738.

——— (2020): "Optimality of Debt under Flexible Information Acquisition," *Review of Economic Studies*, 87, 487–536.

YORK, B. (2012): "In Boston, stunned Romney supporters struggle to explain defeat," *Washington Examiner*, 7.

ZIEGER, R. H. (1994): *American workers, American unions*, Johns Hopkins Univ Press.