Institutional Design Under Asymmetric Information

Madhav Shrihari Aney

Dissertation submitted for the degree of
Doctor of Philosophy in Economics at
The London School of Economics and Political Science

May 2009

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others. The third chapter draws on work that is part of an ongoing project carried out jointly with Maitreesh Ghatak and Massimo Morelli. However the draft version that is included as part of this thesis is mostly my own work.

Madhav Shrihari Aney

Abstract

This thesis contributes to the literature that seeks to understand institutions. In particular the aim of this thesis is to shed light on how certain institutions arise in society as a result of collective choice, how in turn they shape behaviour of agents, and finally what their welfare properties are. These questions are tackled using the methodology of microeconomic theory where agent preferences, the state of technology, and the informational environment are taken as exogenous. In particular it is argued that the existence of different constraints on the informational environment can give rise to a rich theory of institutions that can explain why inefficient and seemingly inefficient institutions arise in a second best world. The first chapter of this thesis is concerned with the incidence of costly dispute resolution in society. The question of why agents fail to revolve disputes costlessly is tackled. This contributes to the positive theory of individual behaviour given the existence of certain institutions. The second chapter of this thesis tackles the question of why the judiciary is characterised by certain inherently costly attributes. This contributes to the normative theory of institutional choice. The last chapter deals with the positive question of how institutions are chosen. A model is presented where the political alignments in a society are endogenously generated and the effect of varying the informational environment on these alignments is analysed. These three chapters collectively contribute to the incipient theory of institutions that comprises of two elements; first where the existence of institutional structure arises as an equilibrium interplay between individual choices and technological and informational constraints, and second where conversely, individual games are shaped by the structure of existing institutions.

Acknowledgments

I thank Maitreesh Ghatak for his innate goodness and for the wisdom of his approach to supervision. I remain indebted to him for helping me construct my identity as a researcher. I thank my friend Giovanni Ko whose company fueled my interest in pursuing a quiet, contemplative career in the academia, and for his help in shaping my ideas about economics and much else. I thank my friends in STICERD for making my time here enjoyable and fulfilling. In particular I thank Konrad Burchardi for being an admirable juggling partner and Stefan Penczynski for being a noble squash adversary. From time to time they even gave me significant feedback on my research. I thank all the faculty in the Department with whom I have discussed my research over the years. I thank Jayant Deshpande for introducing me to the baroque pleasure of mathematics.

I dedicate this thesis to Aii and Baba for their love, and for supplying an upbringing that enabled me to march to the tune of a drummer of my choice.

Contents

List of Figures

Preface

The aim of this thesis is to contribute to the normative and positive questions surrounding institutional design under asymmetric information. The first chapter asks the question of why agents engage in costly dispute resolution such as litigation and arbitration when costless settlement is available. It has been argued that parties are asymmetrically informed about facts and the law surrounding a dispute. This causes the expected payoff from litigation for an agent to be unobservable to her opponent. This unobservability can lead to the break down of pre-trial bargaining. This approach leaves two fundamental questions unanswered: How does informational asymmetry between parties survive given the incentives for full disclosure of certifiable information? Does efficient settlement ensue if parties communicate in forms richer than bargaining? To address the first question I argue that pre-trial informational asymmetry could arise from inherently non-certifiable information, in particular through private valuation of the subject matter in dispute. The second question is tackled by adopting a mechanism design framework to show conditions under which the only possible equilibrium is one where agents litigate. This result arises when parties at the pre-trial stage, lack the ability to fully contract away their right to litigate. This in effect induces agents to exaggerate their true willingness to litigate in order to increase the settlement their opponents are willing to offer. Consequently the credibility of statements made in pre-trial negotiations is destroyed and costly dispute resolution emerges as an equilibrium phenomenon.

The first chapter supplies a positive analysis of the existence of costly conflict in equilibrium. It shows that if courts do not enforce contracts where agents commit not to go to court, then litigation emerges. Consequently the question of why court do not enforce waivers emerges as a puzzle. The second chapter shows that once the incidence of disputes is endogenised this puzzle disappears. The following trade-off is highlighted: conditional on a dispute arising, it is efficient to enforce commitments not to litigate. However, by choosing not enforce such commitments, courts can increase surplus since the threat of costly litigation deters some disputes from arising in the first place.

The final chapter, which draws on work that is being jointly carried out with Maitreesh Ghatak and Massimo Morelli, explores the issue of how market failure interacts with the choice of institutional reform made by an electorate. This issue is studied in an occupational choice framework, where agents are endowed heterogeneously with wealth and talent. In our model, market failure due to unobservability of talent endogenously creates a class structure that affects voting on institutional reform. We find that the preferences of these classes are often aligned in ways that creates a tension between institutional reforms that are growth maximising and those that are politically feasible. This is in contrast to the world without market failure where the electorate unanimously votes in favour of surplus maximising institutional reform. We conclude that inefficiencies of market failure may be further amplified by political choices made by interest groups created in the inefficient market.

Chapter 1

Why do People Litigate?

1.1 Introduction

Underpinning much of the architecture of neo-classical economics lies the assumption of an omniscient judiciary. This judiciary is so efficient that it deters undesirable behavior by its very existence. From Arrow-Debreu contingent commodities to incentive contracts, agents perform their legal obligations in the knowledge that if they do not, they will be punished. Although invoking the court is costly, this does not lead to an inefficiency since even in the unlikely event of a dispute, there is instantaneous resolution through bargaining as both parties are aware that taking the dispute to court is costly.

This logic creates the paradox of litigation: Why do we observe litigation at all when parties are aware of its costliness and costless settlement is available? This is the question that is addressed here. I argue that parties have private valuation of the subject matter in dispute. This causes the expected payoff from litigation to become unobservable to the opponent. At the pre-trial stage parties attempt to negotiate a costless resolution of the dispute in the presence of this informational asymmetry. The outside option to negotiations is the expected payoff from litigation which is increasing in the agent's valuation of the surplus. This creates an incentive for agents to overstate their valuation since high valuation agents receive greater settlement offers during negotiations. The incentive to exaggerate ones valuation creates a lack of credibility about statements made during pre-trial negotiations and consequently causes parties to litigate.

In environments where agents are limited in their ability to contract away their right to litigate, low valuation agents are wary of revealing their type truthfully. This is because truthful revelation weakens their position as their opponents use the revealed information to credibly threaten litigation, and consequently get larger shares of the surplus. In section 1.3.2 I show how litigation is prevented if agents can fully contract away their right to litigate at the pre-trial stage. In section 1.5.2 I argue that this explanation for the existence of litigation generalises to some other forms conflict as well.

This chapter contributes in two ways to the literature on conflict in general and litigation in particular. Firstly it shows the existence of costly dispute resolution in equilibrium in a mechanism design framework rather than is a more limited bargaining framework, a point discussed in section 1.1.3. Secondly, the use of non-certifiable information to generate informational asymmetry immunises this model to the full disclosure critique that has plagued the literature on litigation. This point is discussed in greater detail in section 1.1.2.

1.1.1 First and Second Generation Literature

The large literature that has arisen in response to the question of why people litigate is now two generations old. The first generation literature started with Landes (1971) who argued that litigation arises when its expected benefit is greater than the expected costs for the parties. Parties do not strategically interact in the pre-trial stage and litigation is avoided when the expected benefit of litigating is lower than the expected cost.

In this literature out of court settlement occurs when parties have similar expectations about the outcome of the trial. It is worth explaining this point. Uncertainty about the outcome of a trial on its own cannot be the cause of litigation. With uncertainty, both parties would form expectations about what would happen in court. If the probabilities both associate with winning add up to one, they would settle outside thereby saving themselves the cost of litigation. Litigation arises for instance if both parties overestimate their chances of winning in court. Though this literature acknowledges the role of such overestimation in generating litigation, it stops short of modelling both how this overestimation arises and more importantly the strategic behaviour of parties when they overestimate the expected payoff from litigation.[1]

In response to this unresolved issue, a second generation literature has arisen starting with P'ng (1983) and Bebchuk (1984). In Bebchuk (1984) the defendant knows the probability of winning whereas the plaintiff only knows the distribution over the probability of winning. The plaintiff makes an offer of settlement which the defendant can accept or reject. If the offer is rejected the case goes to court. Since this bargaining game is played out between the parties in an environment of incomplete information, the inefficiency of

[1]More examples include Gould (1973), Posner (1973), and Priest and Klien (1984). More recently Yildiz (2003) and Yildiz (2004) formalised how diverging expectations about the bargaining process can lead to inefficiencies in settlement.

litigation arises.[2]

This is a reflection of the broader theoretical insight that full efficiency is not guaranteed with bargaining under incomplete information. In the next two subsections the two problems with the second generation literature that this chapter seeks to address are explained.

1.1.2 Litigation and Full Disclosure

The first problem concerns the relationship between private information of parties and the unobservability of opponent's payoff from litigation. The justification given in this literature for private information leading to litigation payoffs being unobservable is that a party to a dispute may be in possession of information that once revealed in court increases its probability of winning.[3] It is quite plausible that parties would have such information. A defendant in a law suit for negligence is likely to have more information than the plaintiff about whether she exercised due care. The plaintiff on the other hand is likely to have private information about the exact amount of damage he has suffered.

However if parties possess information that is assumed to be certifiable in court, parties can choose to reveal it to each other outside court at the pre-trial stage. If parties choose to disclose their private information, they find themselves symmetrically informed and consequently litigation is avoided through bargaining. The question that arises at this point is; do parties have an incentive to reveal their private information before trial? The answer to this question is yes even under very weak conditions.

Grossman (1981) shows that when private information is certifiable, there are very strong incentives to reveal it. The intuition for this is that when an agent has information that is favourable to himself, he would always want to reveal it since this leads to better offers from his opponent. This leads to an unravelling in the sense that the agent who chooses not to reveal his information ends up signalling that he has unfavourable information[4]. The

[2]This result has been generalised in different ways. Schweizer (1989) allows for both parties to be in possession private information. Nalebuff (1987) allows for the informed agent to make the settlement offer, thereby considering the signalling implications of the size of the offer and its rejection. Spier (1992) considers more stages to bargaining. Friedman and Wittman (2006) explore pre-trial settlement when parties employ the Chatterjee and Samuelson (1983) protocol for bargaining. Although, as noted in Daughety and Reinganum (1994), the predictions of these models vary in terms of equilibrium allocations for plaintiff and defendant, a non zero probability of litigation emerges as a robust phenomenon. In fact Spier (1994), using a mechanism design approach, shows that litigation would arise even when parties bargain using the most efficient extensive form. See Cooter and Rubinfeld (1989) and Hay and Spier (1998) for surveys of this literature.

[3]Though the literature has focused on this channel, there are other channels through which private information can generate unobservability of the payoff from litigation. For example even if parties have the same priors but have private valuation of the subject matter in dispute, this is sufficient for bargaining to be inefficient. What is required is that the expected payoff from litigation be private information. Overestimating the probability of winning in court is only one of the ways this can happen.

[4]Okuno-Fujiwara et al. (1990) derive conditions sufficient for this argument to work. Shavell

existence of litigation in equilibrium in this literature disappears as soon as parties can communicate in forms that are richer than bargaining. This is because parties would divulge their certifiable information and then bargain efficiently in the environment of complete information.

I propose a different approach by assuming that the asymmetry between parties is about information that is inherently non-certifiable. In my model the valuation that parties place on the subject matter in dispute is private information. This valuation determines the amount of effort an agent is willing to exert in court, which in turn determines the probability of winning. Hence the diverging expectations that parties have about the payoff from litigation are endogenously generated here. In contrast to private information on evidence which can be certified by the informed agent, declarations of valuation are essentially cheap talk; all types would declare that they have high valuation since this increases the settlement offer they are likely to receive. Unlike evidence that is certifiable, there may not exist an efficient way to credibly display high valuation. High valuation may be credibly revealed only through a costly action, like spending more in a trial, that an agent with lower valuation will not find optimal.

1.1.3 Litigation and Mechanism Design

The second problem with the literature on litigation is its focus on bargaining as a means of resolving disputes outside court. Focusing attention singularly on bargaining implies that parties communicate only through offers and counter offers. This assumption about the nature of pre-trial negotiation is very restrictive since communication between parties is not limited to a sequence of offers and counter offers. Communication between parties can include a sequence of messages exchanged in a rich language that could in principle mitigate the informational asymmetry that exists between parties. Hence by restricting the form of pre-trial negotiation to be of the bargaining variety, it is possible to miss out on equilibria in which parties settle out of court.

The model presented here is the first to attempt the resolution of this problem using a mechanism design approach. The seminal paper by Myerson (1982) shows that an equilibrium of any Bayesian game can be replicated through a direct mechanism. This result is known as the revelation principle. Using this insight, the result presented here will show that litigation may arise even when no restrictions are made about the nature of communication between parties during pre-trial negotiation. Since bargaining under incomplete

(1989) finds that this argument in the setting of litigation leads to certifiable private information washing away before trial through voluntary disclosure. Hay (1995) finds the opposite result while focusing on laws mandating full disclosure. However he does not consider the possibility of signalling through non-disclosure.

information is only one of the ways parties could communicate, this is subsumed in the model presented here.

1.1.4 Alternative Explanation for Litigation

The Bebchuk (1984) framework has been the most widely accepted one for explaining the existence of litigation. However there are other explanations. A possible explanation that has received some attention is one based on the existence of communication costs. If the costs of communication between parties are high, then parties would prefer to simply take the matter to court rather than settle it between themselves. In fact even if the costs are very low, as long as both parties need to pay the costs non-cooperatively to start communication, Anderlini and Felli (2001) show that there would always exists an equilibrium where communication will not take place. This explanation may fit a certain class of litigation. For example it may explain divorce battles between spouses where the prospect of communicating with the opponent is so odious that costly litigation is preferred.

In a similar vein Robson and Skaperdas (2008) construct a model where parties need to pay costs non-coperatively before they can enter the stage of pre-trial settlement. These costs influence the probability of winning the case and hence influence the outcome of pre-trial bargaining. The idea is that parties by committing to litigate may reduce total costs if a substantial part of these costs is paid ex-ante before bargaining takes place. This happens because committing not to settle dampens the incentive to make costly effort ex-ante. However, just like in Anderlini and Felli (2001), once parties are allowed to meet before these costs are incurred the result disappears. Since the existence of ex-ante costs is the crucial ingredient here, it is difficult to extend this explanation to all litigation unless a micro foundation for the existence of these costs is supplied.

The model presented here synthesizes the two main approaches used for analysing litigation. The literature on pre-trial negotiations treats the court process as exogenous.[5] In contrast the literature on conflict treats the failure of pre-trial negotiation as exogenous and models the court process as a complete information contest between two parties where the probability of winning is endogenously determined by the effort exerted by parties.[6] This chapter combines these two approaches by modeling the court process as a contest in an environment of incomplete information, with a mechanism design stage preceding litigation where parties can negotiate to avoid costly litigation.

[5] See for example Bebchuk (1984) and Spier (1994)

[6] See for example, Hirshleifer and Osborne (2001)

1.2 Model

There are two agents who find themselves in a dispute. The subject matter of the dispute is characterised as surplus of size 1 over which agents have competing claims. Both agent have positive valuation of the surplus which is their type. Agent 1's valuation is θ_1 which is observable where as agent 2's valuation is unobservable and can be θ_2^H with probability q^H and θ_2^L with probability $q^L = 1 - q^H$.[7]

if Agents are aware of their own type before the game begins. The model can be easily generalised to the case where the informational asymmetry is two-sided, but uncertainty over the type of one agent is sufficient to generate litigation in equilibrium. I assume that

$$\theta_1 > \theta_2^H > \theta_2^L.$$

The assumption that valuation of a party is unobservable is the key driver of litigation in this model. It is worthwhile to see some examples where litigation can be interpreted as a dispute over surplus. These examples have been chosen to illustrate how the model may apply to a large range of situations. Examples include:

- Dispute over property: A party has private valuation over a piece of property and it is unclear as to who has title over it. The property could be tangible such as land or intangible such as an invention.
- Suits for specific performance: There may be a dispute as to whether an agent has performed its contractual obligation. The plaintiff may have private valuation over the benefit accruing from the action or the defendant may have private valuation over the costs of performing the action.
- Custody battle over children: When a couple separates, the spouses may have private valuations over the custody of their children.

Private valuation of the subject matter in dispute is plausible when the dispute involves something more than just monetary compensation. The model can be extended to cases involving only monetary compensation if the assumption of utility functions being linear in money is relaxed. However the analysis presented here excludes these cases since the mechanism design problem becomes less tractable if the utility of parties is nonlinear in transfers.

[7] An earlier draft allowed both parties to have private information on valuation. The assumption of one sided asymmetric information is preferred since it has two advantages. Firstly it simplifies the model and delivers a clear intuition about the result. Secondly, it demonstrates how the mechanics that drive the result are not the ones subsumed in Myerson and Satterthwaite (1983).

Timeline:

Stage 1: Agent 2 realises his private valuation of the surplus.

Stage 2: A dispute arises between the 2 parties.

Stage 3: Parties to dispute start pre-trial negotiation which is a game that may help them avoid taking the matter to court.

Stage 4: Parties play the game from stage 2 and receive the equilibrium allocation.

Stage 5: Parties non-cooperatively choose between the equilibrium allocation from stage 2 and approaching the court.

Stage 6: If either agent has approached the court, then both non co-operatively choose their effort levels.

Stage 7: Court observes the effort of each agent and makes a final decision.

It is helpful at this point to preview how the result of litigation in equilibrium is established. At stage 3 in the model agents undertake pre-trial negotiations which formally is a game that will help them avoid litigation. This game yields some equilibrium in stage 4. Due to the revelation principle it is possible to characterise the existence of litigation in this equilibrium without specifying the actual game in stage 3. This is because the revelation principle allows for the replication of any equilibrium of a Bayesian through a direct mechanism. I will first prove the non-existence of a separating and semi-separating equilibrium. In particular I will show that when agents cannot contract away their right to litigate at stage 3, an incentive to lie for a low valuation agent 2 arises. This is because he anticipates that if his declare their types truthfully, their opponent would force him to re negotiate the allocation from stage 4 with a credible threat of litigation. This implies that the only possible equilibrium is the pooling equilibrium where all agents declare themselves to have a high valuation. I will show that in this equilibrium litigation arises since agent 1 has a higher expected payoff from litigation than the settlement payoff from stage 4 under certain conditions. In section 1.3.2 I show how litigation disappears in equilibrium when the ability to contract away the right to litigate is introduced.

This model can be solved starting backwards. In the next subsection a stylized model of litigation is presented. This is meant to crudely capture what happens in court. Since parties know what would happen in court, the equilibrium allocations from the game they play in stage 3 must at least make parties indifferent between litigating and not litigating in stage 4. I call this a litigation-proofness constraint. This constraint is derived in section 1.2.2 for the different kinds of equilibria. Going back another step, in section 1.2.3 the incentive-compatibility constraints for different equilibria are derived. Finally in section 1.3 I present the result that shows the existence of litigation in equilibrium.

1.2.1 Litigation

In line with the large literature on the question of why conflict occurs, the court process is modeled as a static contest where the probability of winning is determined by the effort x exerted by parties.[8] Following are the objective functions of the two agents.

$$\theta_1 \, \mathrm{P}(x_1, x_2) - x_1 \qquad \text{and} \qquad \theta_2^j(1 - \mathrm{P}(x_1, x_2)) - x_2 \qquad j \in \{L, H\}$$

where θ_1 is the valuation of agent 1 and θ_2^j is the valuation of agent 2. Note that henceforth j refers to the type of agent 2. This formulation assumes that the payoff of parties is linear in money (effort). It is assumed that the probability of winning is increasing in ones own effort and decreasing in the effort of the opponent.

Effort can be thought of as fees of the lawyer hired by an agent. If lawyers are paid in proportion to their marginal product then it must be the case that the heterogeneity in the fees lawyers charge can be explained by the degree of persuasiveness they have in court. Apart from the sensitivity of the judicial process to the skills of a lawyer, it is also possible to have an interpretation of effort in terms of how much money is paid off to a corrupt judiciary to secure a favourable decision[9].

If the valuation of agent 2 was observable, then this would be a game of complete information where we could compute the Nash equilibrium effort levels of the agents. Here, since his valuation is private information, agent 1 instead plays a Bayesian game where the optimal effort level of agent 1 is:

$$x_1(\theta_1) = \operatorname*{argmax}_{x_1 \geq 0} \theta_1 \left(\sum_{j \in \{H,L\}} q^j \, \mathrm{P}\left(x_1, x_2(\theta_2^j)\right) \right) - x_1. \tag{1.1}$$

Having computed the Bayesian Nash equilibrium effort levels agent 1, we can work out the expected payoff from litigation for agent 1. This is

[8] See Skaperdas (2006) for surveys of this literature.

[9] A contest function is not necessary for the results of this chapter. What is required is that $\lim_{\theta_2^j \to 0} \mathrm{P}(\theta_1, \theta_2^j) \to 1$ and $\lim_{\theta_1 \to 0} \mathrm{P}(\theta_1, \theta_2^j) \to 0$. The contest function specified later delivers this in a reduced form through the equilibrium efforts of the agents. The attraction of using a contest function is that it allows the model to endogenise litigation effort and consequently the inefficiency of litigation.

$$v_1(\theta_1) = \theta_1 \left(\sum_{j\in\{H,L\}} q^j \, \mathrm{P}\,(x_1(\theta_1), x_2(\theta_2^j)) \right) - x_1(\theta_1). \tag{1.2}$$

Since agent 1's valuation is observable, agent 2's equilibrium effort level is

$$x_2(\theta_2^j) = \operatorname*{argmax}_{x_2\geq 0} \theta_2^j \, \mathrm{P}\,(x_1(\theta_1), x_2(\theta_2^j)) - x_2 \qquad j \in \{H, L\}. \tag{1.3}$$

Plugging these effort levels back into the objective function, we get the expected payoff from litigation for agent 2 which is

$$v_2(\theta_2^j) = \theta_2^j q^j \, \mathrm{P}\,(x_1(\theta_1), x_2(\theta_2^j)) - x_2(\theta_1^j) \qquad j \in \{H, L\}. \tag{1.4}$$

Inspecting equation (1.2) we can already see how private information can lead to agent 1 overestimating her probability of winning in court. This could happen when agent 2 has high valuation. But agent 1, taking expectations over the type of agent 2, would believe him to have an average valuation. The value function can be re-written as

$$v_1(\theta_1) = \theta_1 \, \mathrm{E}_{\theta_2^j} \left(\mathrm{P}\,(x_1(\theta_1), x_2(\theta_2^j))\right) - x_1(\theta_1).$$

It can be seen that the sum of the expectation over the probability of winning for agent 1 and the that of a high valuation agent 2 can be greater than 1.

This is not the only feature that makes the payoff of one agent unobservable to the other. In addition to affecting the probability, the type also enter the payoff function directly as the value placed on the surplus, and indirectly through the equilibrium effort of the agent. The inability of agent 1 to observe agent 2's expected payoff from litigation, generates the existence of litigation in equilibrium. If this contest was played in an environment of complete information, then expected payoffs from litigation would be common knowledge. This would allow parties to bargain around costly litigation since the opponent's outside option to bargaining would be observable. The unobservability of types causes the outside option to bargaining being unobservable and consequently, as shown by Schweizer (1989), full efficiency is no longer attainable with bargaining.

At this point one may naturally ask why parties should restrict themselves to bargaining as a pre-trial mechanism to avoid litigating? Why can they not design any other mechanism that will allow agent 2 to reveal his valuation? I address this question in the next subsection.

1.2.2 Litigation-Proofness

Before resorting to costly litigation parties can play any Bayesian game such that the type of agent 2 will stand revealed. This problem of a general game form is tractable using the revelation principle since any equilibrium in a Bayesian game can be replicated by the use of a direct mechanism where the parties reveal their types truthfully to a mediator. To see whether litigation can be avoided, we need to check whether a more efficient allocation that does not require external financing is implementable[10].

The equilibrium allocation that is replicated using a direct mechanism needs to be litigation-proof. This means that the payoff from the equilibrium should be weakly greater than the payoff from litigation. The expected payoff from litigation depends on the nature of equilibrium we try to implement using the direct mechanism. This problem can be tackled by considering different kinds of equilibria separately.

Separating Equilibrium

In a separating equilibrium the valuations of the agent 2 would stand revealed in stage 4. Consequently parties would find themselves in an environment of complete information. At this stage parties should prefer the allocations that have been prescribed by the mechanism to litigation. If parties choose to litigate at this stage, their payoffs would be the Nash equilibrium payoff from litigation under complete information. The Nash equilibrium levels of effort that would be played when an agent 1 meets an agent 2 of type j can be computed. Let these be x_1^j and x_2^j:

$$x_1^j = \operatorname*{argmax}_{x_1 \geq 0} \theta_1 \, \mathrm{P}(x_1, x_2^j) - x_1 \qquad \text{and} \qquad x_2^j = \operatorname*{argmax}_{x_2 \geq 0} \theta_2^j (1 - \mathrm{P}(x_1^j, x_2)) - x_2.$$

These may not be unique.[11] Using these optimal effort levels we can calculate v_1^j, the expected payoff from litigation when agent 1 confronts a type j agent 2.

$$v_1^j = \theta_1 \, \mathrm{P}(x_1^j, x_2^j) - x_1^j \qquad v_2^j = \theta_2^j (1 - \mathrm{P}(x_1^j, x_2^j)) - x_2^j \tag{1.5}$$

Although parties would never actually litigate in an environment of complete information, v_1 and v_2 become credible threat points that parties would

[10] It is reasonable to impose the restriction of no external financing since parties in the real world cannot expect outside subsidies for settlement of private disputes. If budget balance is not imposed then the problem would disappear since a Groves mechanism would always ensure incentive-compatibility. See Groves (1973).

[11] In the results when a functional form for the contest function is specified, conditions that ensure uniqueness are presented.

use to force the renegotiation of allocations ex-post. In other words, bargaining would ensue in states of the world where parties find that they are guaranteed a higher expected payoff by litigating rather than accepting the allocations specified by the mechanism. While being completely agnostic about the extensive form that such bargaining would take, we know that v_1 and v_2 will be the outside options to such bargaining. If parties anticipate that such bargaining will take place ex-post, this destroys existence of a separating equilibrium unless the allocations are designed to ensure that parties cannot credibly threaten litigation in stage 5. Hence to avoid a credible threat of litigation the transfers from the mechanism must satisfy litigation-proofness constraints. These are

$$\begin{aligned} \theta_1 - v_1^j \geq t_1^j \geq v_2^j \\ \theta_2^j - v_2^j \geq t_2^j \geq v_1^j \end{aligned} \tag{1.6}$$

for $j \in \{H, L\}$ where t_1^j is the net transfer paid by agent 1 to type j agent 2 in the event the mechanism allocates the surplus to agent 1. Similarly t_2^j is the net transfer paid by type j agent 2 to agent 1 in the event the mechanism allocates the surplus to agent 2. The constraints state that the payoff from negotiations should be greater than the payoff from litigating. This should be true for both the agents regardless of who receives the surplus, and for all realisations of agent 2's type.

Pooling Equilibrium

In a pooling equilibrium the agent 1 learns nothing about the type of agent 2 at stage 4. Hence their expected payoff from litigation remains the Bayesian Nash equilibrium payoff $v_1(\theta_1)$ and $v_2(\theta_2^j)$ defined in equations (1.2) and (1.4). Let the μ_1 and μ_2 be the transfer made by agents 1 and 2 respectively to their opponent when the surplus is allocated to them. The litigation-proofness constraints are

$$\begin{aligned} \theta_1 - v_1(\theta_1) \geq \mu_1 \geq v_2(\theta_2^j) \\ \theta_2^j - v_2(\theta_2^j) \geq \mu_2 \geq v_1(\theta_1) \end{aligned} \tag{1.7}$$

What defines a pooling equilibrium is that the declaration of agent 2 conveys no information about his type. Consequently there is no change in agent 1's prior about the type of agent 2. Note that in terms of the declarations that agent 2 makes, there are various ways in which a pooling equilibrium can arise. It arises when agent 2 makes the same declaration regardless of his type. More generally, it arises whenever agent 2 has the same probability distribution over declarations regardless of his type. What is important here is that the constraints defined in (1.7) are unaffected by which pooling equilibrium we consider. This is because the payoff from litigation $v_2(\theta_2^j)$, which is the outside option to μ_2, remains constant.

Semi-Separating Equilibrium

Just like in the case of pooling equilibria, there are infinitely many semi-separating equilibria that can arise. A semi-separating equilibrium is defined by the fact that there is some information conveyed to agent 1 through the declaration of agent 2. Hence the payoff from litigation thereafter is modified since agent 1 uses updated probabilities when deciding his optimal effort level in court. Let the optimal effort of agent 1 and 2 be $\tilde{x}_1$ and $\tilde{x}_2$ where

$$\tilde{x}_1 = \underset{x_1 \geq 0}{\operatorname{argmax}}\, \theta_1 \left(\frac{q^H(1-\gamma)}{q^H(1-\gamma) + q^L\gamma} \mathrm{P}(x_1, \tilde{x}_2^H) + \frac{q^L\gamma}{q^H(1-\gamma) + q^L\gamma} \mathrm{P}(x_1, \tilde{x}_2^L) \right) - x_1, \tag{1.8}$$

$$\tilde{x}_2^H = \underset{x_2 \geq 0}{\operatorname{argmax}}\, \theta_2^H(1 - \mathrm{P}(\tilde{x}_1, x_2)) - x_2, \tag{1.9}$$

and

$$\tilde{x}_2^L = \underset{x_2 \geq 0}{\operatorname{argmax}}\, \theta_2^L(1 - \mathrm{P}(\tilde{x}_1, x_2)) - x_2. \tag{1.10}$$

The value of $\gamma \in (0, 1)$ is determined by the posterior probability that agent 1 associates with agent 2 being a low type. Note that γ is a short hand for the amount of information that is revealed to agent 1 by agent 2's declaration. If the declaration reveals nothing then $\gamma = \frac{1}{2}$ and we are back in a pooling equilibrium. When $\gamma = 0$ agent 1 knows the that agent 2 is a high type with certainty and we are in a separating equilibrium. Similarly $\gamma = 1$ implies that we are in a separating equilibrium where agent 2 has been revealed to be a low type. The intermediate value of γ strictly between 0 and 1 but not equal to one half are ones that would arise in a semi-separating equilibrium.

Using the semi-separating equilibrium effort levels calculated above we can back out the litigation value functions for agent 1 and 2. These are $\tilde{v}_1, \tilde{v}_2^H$ and $\tilde{v}_2^L$ respectively.

$$\tilde{v}_1 = \theta_1 \left(\frac{q^H(1-\gamma)}{q^H(1-\gamma) + q^L\gamma} \mathrm{P}(\tilde{x}_1, \tilde{x}_2^H) + \frac{q^L\gamma}{q^H(1-\gamma) + q^L\gamma} \mathrm{P}(\tilde{x}_1, \tilde{x}_2^L) \right) - \tilde{x}_1, \tag{1.11}$$

$$\tilde{v}_2^H = \theta_2^H(1 - \mathrm{P}(\tilde{x}_1, \tilde{x}_2^H)) - x_2^H, \tag{1.12}$$

and

$$\tilde{v}_2^L = \theta_2^L(1 - \mathrm{P}(\tilde{x}_1, \tilde{x}_2^L)) - x_2^L. \tag{1.13}$$

We care now ready to characterise the litigation-proofness constraints in a semi-separating equilibrium. These are

$$\theta_1 - \tilde{v}_1 \geq \tilde{\mu}_1 \geq \tilde{v}_2^j \tag{1.14}$$

$$\theta_2^j - \tilde{v}_2^j \geq \tilde{\mu}_2 \geq v_1$$

where $\tilde{\mu}_1$ and $\tilde{\mu}_2$ are the transfers made to the opponent when the surplus is allocated to agent 1 and 2 respectively.

1.2.3 Incentive-Compatibility Constraints for a Separating Equilibrium

Consider a direct mechanism where agent 2 declares a type j where $j \in \{H, L\}$. If the declarations of agent 2 θ_2^j, then agent 1 is allocated the surplus with probability δ^j and agent 2 is allocated allocated the surplus with probability $1 - \delta^j$. Following are the incentive-compatibility constraints for agent 2 in a separating equilibrium.

$$H_2 : \delta^H t_1^H + (1 - \delta^H)(\theta_2^H - t_2^H) \geq \delta^L t_1^L + (1 - \delta^L)(\theta_2^H - t_2^L) \tag{1.15}$$

$$L_2 : \delta^L t_1^L + (1 - \delta^L)(\theta_2^L - t_2^L) \geq \delta^H t_1^H + (1 - \delta^H)(\theta_2^L - t_2^H)$$

The exercise here is to find an allocation composed of transfers t_1^j, t_2^j along with probability δ^j that satisfies incentive-compatibility and litigation-proofness for all j. If such an allocation exists then parties would reveal their types truthfully knowing that for any possible realisation of types, the allocation guarantees that the opponent cannot credibly threaten litigation ex-post.

Since δ^j can be less than one, there will be states when the surplus is allocated to agent 2 even though agent 1 always has greater valuation of the surplus than agent 2. If this happens both agents could voluntarily renegotiate and allocate the surplus to agent 1 in exchange of a transfer for agent 2. Since agents are aware that this will happen in all states when the surplus is allocated to the agent with the higher valuation, this has to be taken into account while implementing the separating equilibrium allocation.

Lemma 1.1. *The non-existence of transfers satisfying litigation-proofness and incentive-compatibility for $\delta^j = 1$ is sufficient for proving the non existence of transfers for any $\delta^j \in [0, 1]$ if agents are allowed to trade the surplus once it is allocated.*

Proof. In a separating equilibrium, the type of agent 2 corresponds to his declaration. Assume that the surplus is allocated to agent 2. In this state agents would find it profitable to allocate the surplus to agent 1 in exchange of a transfer to agent 2. Let this transfer be $\hat{t}_1^j$. The condition

$$\theta_1 - t_2^j \geq \hat{t}_1^j \geq \theta_2^j - t_2^j \tag{1.16}$$

(1.16) must be satisfied for agents to voluntarily agree on trade of the surplus for transfer $\hat{t}_1^j$. Using (1.16) and the litigation-proofness constraints from (1.6) we have

$$\begin{aligned} \hat{t}_1^j &\geq \theta_2^j - t_2^j &\geq v_2^j. \\ \theta_1 - v_1^j &\geq \theta_1 - t_2^j &\geq \hat{t}_1^j. \end{aligned}$$

This implies that if no $\hat{t}_1^j$ exists that satisfies

$$\theta_1 - v_1^j \geq \hat{t}_1^j \geq v_2^j$$

then there cannot exist a $\hat{t}_1^j$ that satisfies (1.16). Note that this larger range for $\hat{t}_1^j$ is the same as the one imposed on t_1^j by (1.6).

In a state where the declaration of 2 is j, the term in the incentive-compatibility constraint for agent 2 modifies to

$$\delta^j t_1^j + (1 - \delta^j)(\theta_2^j - t_2^j) = \delta^j t_1^j + (1 - \delta^j)\hat{t}_1^j$$

which can be replaced with t_1^j without loss of generality since $\delta_1^j + \delta_2^{ij} = 1$ and the range for both transfers is the same. Therefore the non existence of t_1^{ij} that satisfies incentive-compatibility and litigation-proofness for $\delta_1^{ij} = 1$, is sufficient for proving the non existence of t_1^{ij}, t_2^{ij} for any $\delta_1^{ij} \in [0, 1]$. □

Lemma 1 tells us that given agents will voluntarily trade the surplus when it is allocated to the agent with the lower valuation. We can therefore, without loss of generality, focus our attention on the direct mechanism that always allocates the surplus to the agent with the higher valuation. Applying Lemma 1.16 to Case 1 where the distribution of valuations do not intersect, we have $\delta_1^{HH} = \delta_1^{HL} = \delta_1^{LL} = \delta_1^{LH} = 1$. This simplifies the incentive-cmpatibility constraints of agent 2 to

$$t_1^H = t_1^L \tag{1.17}$$

The intuition for this condition is the following. Since agent 2 knows that finally the surplus will always go to agent 1, he has an incentive to make the declaration that guarantees him the maximum possible transfer. The only way to incentivize him to tell the truth is to make the transfer independent of his declaration. The results will show how this restrictions placed on the range of transfers may be inconsistent with the range for transfers in the litigation-proofness constraints derived in section 1.2.2.

The ability to trade surplus ex-post simplifies the analysis by allowing us to focus on efficient allocations, that is allocations where surplus goes to the agent 1. The introduction of trade however is is not the driver of the results of the model, it merely simplifies the analysis. It will be seen that what creates litigation is the inability of parties to fully contract away their right to litigate. In the extension, I show that even if parties can commit not to trade, litigation may still arise as long as certain conditions are satisfied.

1.3 Results

In this section main result regarding the existence of litigation in equilibrium is proven. In the next two subsections the existence of litigation in two extreme cases of full commitment and complete non-contractability is discussed. Thereafter the implications of partial contractibility on existence of litigation are analysed.

1.3.1 Litigation under Non-Contractability

It is difficult to proceed further without pinning down a functional form for the contest function. This allows for a computation of the value function v_1 and v_2. The contest function that is used here is

$$\mathrm{P}(x_1, x_2) = \frac{\alpha x_1^{\lambda}}{\alpha x_1^{\lambda} + (1-\alpha)x_2^{\lambda}} \qquad \lambda, \alpha \in (0, 1). \tag{1.18}$$

Skaperdas (1996) provides the axiomatic foundations of this contest function for the case of α equal to $\frac{1}{2}$. Clark and Riis (1998) generalise this to the case where α takes any value between zero and one. This contest function is unique in that the winning probability depends on the ratio of equilibrium efforts[12] It differs from the exponential contest function where the winning probability depends on the difference of the efforts exerted by parties. This function is easily parameterised, and allows a closed form characterisation of the value functions for both agents. λ less than 1 implies concavity and ensures the uniqueness of equilibrium.

The primary parameter that characterises the function is λ. This captures how sensitive the probability is to the effort exerted by parties. A higher

[12] Strictly, what is required for the arguments in this chapter to go through, is that $\lim_{\theta_1 \to 0} \frac{\partial \mathrm{P}(\theta_1,\theta_2)}{\partial \theta_1} = 0$ and $\lim_{\theta_2 \to 0} \frac{\partial \mathrm{P}(\theta_1,\theta_2)}{\partial \theta_1} = 1$. This is what the contest function in (1.18) delivers in the reduced form when the equilibrium effort levels as a function of valuations are computed. This is in contrast to the exponential contest function $\frac{\alpha \exp^{\lambda x_1}}{\alpha \exp^{\lambda x_1} + (1-\alpha)\exp^{\lambda x_2}}$ where this property does not hold.

λ implies a greater sensitivity of the judicial process to the persuasiveness of lawyers. A judicial process completely insensitive to the skill of lawyers implies a λ equal to 0. Alternatively a high responsiveness of the probability of winning to effort could simply mean that it is cheap and easy to bribe judges. In this interpretation λ can be thought of as a parameter capturing how corrupt the judiciary is. In this interpretation λ equal to 0 implies an incorruptible judiciary since the decision of the court is insensitive to bribes.

The point of departure from Skaperdas (1996) for this contest function is the presence of the additional parameter α which captures how strong agent 1's case is ex-ante relative to agent 2. This parameter is introduced to capture the fact that legal disputes may be skewed towards one side. It is rarely the case that both sides to a dispute have equally strong legal positions.

An α equal to 1 implies that agent 1 is certain to win the case; that the case is 'open and shut'. Similarly α equal to 0 implies that agent 2 is certain to win. Note that in these two corner cases the efforts of parties will not play a role as the probability of winning would be insensitive to effort. Note that if α is either 0 or 1 there would never be any litigation since one of the two parties regardless of its valuation would have an observable payoff of 0 from litigation and hence would always settle for 0 outside. For intermediate values of α, the effort of parties would influence the probability of winning. An α equal to one half implies that if both agents were to exert the same effort, the outcome of the case is equiprobable.

α is a reduced form catchall parameter that captures both the legal characteristics and the facts of the dispute. The following examples illustrate this.

- In a custody battle, the laws of most countries usually favour the mother. If this is the case then α would be greater than one half when agent 1 is the mother. The value α would depend on the specific laws on custody of children of the country in which the dispute takes place. The value of α would also depend on the facts of the particular case. If for example, agent 1 is known to have a history of drug problems then α would be lower.

- In a battle over intellectual property agent 2, the alleged infringer, may or may not have the right to contest the validity of the patent. For example, in the UK a patent can be challenged only once in a court of law. If it is upheld, agent 1 the holder of the patent, is granted a certificate of contested validity which protects it from any further challenges. Hence the question of whether the law allows more than one challenge to the patent, along with the factual position of whether or not the patent in question has been previously challenged, may together affect α.

- In a suit for specific performance if the provisions of the contract make a clear case in favour of the plaintiff then this brings α closer to 1. On the other hand if it can be shown that the contract is in violation of public policy then the contract would be set aside and the defendant would not be called upon to perform his contractual obligations. This would push α closer to zero.

Following the discussion on disclosure in the introduction, even if facts of the case are private information to begin with, as long as they are certifiable, they would be revealed before litigation takes place and would influence α. Hence α like λ is common knowledge. In a world where given a set of facts there is no room for disagreement about the application of the law, and consequently the outcome of the case is certain, α would always either be 0 or 1. However this is never the case since the interpretation of the law is often contentious. Even if the interpretation of the individual laws is clear, there may be ambiguity about which law is to be applied to the case at hand.

In addition to the fact that the application of the law is inherently uncertain due to complexity of the specific case to which it is applied, there could be perverse incentives in judicial systems that increases this uncertainty even further. Levy (2005) argues that career concerns could induce judges to contradict previous decisions consequently creating uncertainty about the law. This creates a role for lawyer's skill since courts are open to persuasion.

Apart from the lawyer's skill in persuading the court on points of law, there is also often room for persuasion on points of fact. What truly happened at a point in the past is often unobservable and exists only as a probability distribution for the court. A lawyer's skill could therefore play a role in influencing what the court believes to be true about an event. All this creates uncertainty about the outcome of the case which can lead to a realisation of α that is not 0 or 1.

Using this contest function it is possible to solve out for the Nash equilibrium effort levels when agent 1 confronts a type j agent 2.

$$x_1^j = \theta_1 \frac{(\theta_2^j\theta_1)^\lambda \alpha(1-\alpha)\lambda}{(\alpha(\theta_1)^\lambda + (1-\alpha)(\theta_2^j)^\lambda)^2}, \qquad x_2^j = \theta_2^j \frac{(\theta_1\theta_2^j)^\lambda \alpha(1-\alpha)\lambda}{(\alpha(\theta_1)^\lambda + (1-\alpha)(\theta_2^j)^\lambda)^2}. \quad (1.19)$$

The corresponding value functions are:

$$v_1^j = \theta_1 \left(\frac{(\theta_1)^\lambda}{(\theta_1)^\lambda + (\theta_2^j)^\lambda} - \frac{(\theta_2^j\theta_1)^\lambda \alpha(1-\alpha)\lambda}{(\alpha(\theta_1)^\lambda + (1-\alpha)(\theta_2^j)^\lambda)^2} \right), \quad (1.20)$$

and

$$v_2^j = \theta_2^j \left(\frac{(\theta_2^j)^\lambda}{(\theta_1)^\lambda + (\theta_2^j)^\lambda} - \frac{(\theta_1 \theta_2^j)^\lambda \alpha(1-\alpha)\lambda}{(\alpha(\theta_1)^\lambda + (1-\alpha)(\theta_2^j)^\lambda)^2} \right).$$

Inspecting these value functions confirms certain desirable properties.

$$\frac{\partial x_2^j}{\partial \theta_2^j} > 0 \quad \text{and} \quad \frac{\partial x_1^j}{\partial \theta_2^j} < 0. \tag{1.21}$$

The effort an agent exerts in court is increasing in her own type. This is intuitive since a greater valuation makes an agent more active in court since there's more at stake for her. Conversely the payoff of an agent is decreasing in the opponent's valuation.

$$\frac{\partial v_2^j}{\partial \theta_2^j} > 0 \quad \text{and} \quad \frac{\partial v_1^j}{\partial \theta_2^j} < 0. \tag{1.22}$$

The expected payoff from litigation is monotonically increasing in own valuation and decreasing in the valuation of the opponent. This is a consequence of the property in (1.21).

$$\lim_{\theta_2^j \to 0} v_2^j = 0 \quad \text{and} \quad \lim_{\theta_2^j \to 0} v_1^j = \theta_1. \tag{1.23}$$

As the agent's valuation goes to 0, so does her payoff from litigation. On the other hand an agent's litigation payoff goes to her valuation as the valuation of the opponent goes to zero. This is the case since facing an opponent with low valuation implies that even a small effort is sufficient for securing a high probability of winning.

Non-Existence of a Separating Equilibrium

Using the value functions defined above it is now possible to prove the non existence of a separating equilibrium. Litigation-proofness constraints impose restrictions on the transfer of the mechanism that parties design for pre-trial settlement of dispute. The following result shows that these restrictions may be inconsistent with the restrictions imposed by incentive-compatibility.

Proposition 1.1. *Assuming that* $\mathrm{P}(x_1, x_2)$ *in equation* (1.18) *is the contest function that characterises litigation; a separating equilibrium does not exist if* $\theta_2^H - \theta_2^L$ *is greater than some threshold* Δ.

Proof. We have

$$\theta_1 - v_1^L \geq t_1^L$$

from equation (1.6). Using this together with (1.23) we get

$$\lim_{\theta_2^L \to 0} t_1^L = 0.$$

From (1.17) we have $t_1^L = t_1^H$. And from (1.6) we have $t_1^H > v_2^H$. Since $v_2^H > 0$ from equation (1.20), when $\theta_2^H > 0$. Given the monotonicity property in equation (1.22), there must exist a threshold Δ such that for $\theta_2^H - \theta_2^L > \Delta$, $t_1^L = t_1^H$ cannot be satisfied. Hence a separating equilibrium does not exist. □

The intuition for this result is straightforward. If the minimum transfer that a high type agent 2 receives from agent 1 is large enough, then the incentive for a low type agent 2 to tell the truth is destroyed. With the contest function specified in (1.18), this happens as θ_2^L goes to zero. This is because the effort levels depend on the ratio of the valuations. As θ_2^L goes to zero the payoff from litigation for a low type agent 2 also goes to zero and the likelihood that agent 1 wins in court goes to 1. This in turn restricts the transfers a low type agent 2 can expect from the mechanism. Once θ_2^L is sufficiently far apart from θ_2^H, that is when the difference between the two valuations is large enough, it becomes more attractive for a low type agent 2 to declare himself to be a high type.

The Existence of Litigation in the Pooling Equilibrium

Proposition 1.1 shows that a separating equilibrium cannot exist when $\theta_2^H - \theta_2^L > \Delta$. This leaves open the possibility of the existence of a pooling and a semi-separating equilibrium. In this sub-section, it will be shown that a litigation free pooling equilibrium cannot exist. For litigation to exist, agent 1 should prefer litigating when offered the alternative of allowing agent 2 to pool across his types. The next result shows that this is indeed the case when q^L the probability of agent 2 being a low type is greater than some threshold.

Proposition 1.2. *There exists a threshold q^{L*} such that for any $q_2^{L*} < q_2^L < 1$ agent 1 chooses to litigate rather than accept the payoff from a pooling equilibrium.*

Proof. In a pooling equilibrium, the optimal effort for agent 1 when she litigates is:

$$x_1(\theta_1) = \operatorname*{argmax}_{x_1 \geq 0} \theta_1 \left(\sum_{j \in \{H,L\}} q^j \frac{\alpha x_1^\lambda}{\alpha x_1^\lambda + (1-\alpha) x_2(\theta_2^j)^\lambda} \right) - x_1.$$

Note that the objective function is concave in x_1 since it is a sum of two concave functions. Hence the first order condition yields the optimum. Plugging in the optimal effort levels we get the expected payoff from litigation for agent 1:

$$v_1(\theta_1) = \theta_1 \left(\sum_{j \in \{H,L\}} q_2^j \frac{\alpha x_1(\theta_1)^\lambda}{\alpha x_1(\theta_1)^\lambda + (1-\alpha) x_2(\theta_2^j)^\lambda} \right) - x_1(\theta_1).$$

Now note that:

$$\lim_{\theta_2^L \to 0} x_2(\theta_2^L) = 0.$$

This is true because if the limit of $x_2(\theta_2^L)$ is a positive constant then the expected payoff from litigation for the agent would be negative. This is a contradiction of $x_2(\theta_2^L)$ being an optimum since the agent could reduce x_2 and increase his payoff to 0.

$$\lim_{q^L \to 1, \theta_2^L \to 0} v_1(\theta_1) = \theta_1$$

The minimum transfer that an agent 2 with a high declaration is guaranteed to accept from an agent 1 as settlement outside court is $v_2(\theta_2^H)$. Since agent 2 would always declare himself to be a high type, the payoff for agent 1 from playing the mechanism and accepting its allocation is $\theta_1 - v_2(\theta_2^H) < \theta_1$.

Note that the first order condition for the Bayesian game converges to the first order condition of the game with complete information where agent 2 is a low type:

$$\lim_{q^L \to 1, \theta_2^L \to 0} \sum_{j \in \{H,L\}} q^j \frac{\alpha(1-\alpha) x_1^{\lambda-1} x_2(\theta_2^j)^\lambda}{(\alpha x_1^\lambda + (1-\alpha) x_2(\theta_2^j)^\lambda)^2} - \frac{1}{\lambda \theta_1} = \frac{\alpha(1-\alpha) x_1^{\lambda-1} (x_2^L)^\lambda}{(\alpha x_1^\lambda + (1-\alpha)(x_2^L)^\lambda)^2} - \frac{1}{\lambda \theta_1}.$$

This implies that the optimal effort in the game with incomplete information converges to the optimal effort in the game with complete information as $q^L \to 1$. Since the expected payoff from litigation goes to θ_1 as $q^L \to 1$ and $\theta_2^L \to 0$ there must exist a threshold for q^{L*} such that:

$$\theta_1 \left(\sum_{j \in \{H,L\}} q^{j*} \frac{\alpha x_1(\theta_1)^\lambda}{\alpha x_1(\theta_1)^\lambda + (1-\alpha) x_2(\theta_2^j)^\lambda} \right) - x_1(\theta_1) = \theta_1 - v_2^H.$$

This defines q_2^{L*}. Note that for any $1 > q_2^L > q_2^{L*}$ litigation has a higher expected payoff for agent 1 when agent 2 always declares himself to be a high type. □

Given that a low type agent 2 is sure to declare himself to be a high type, agent 1 has two options; she can either accept the pooling equilibrium allocation from the negotiations or she can litigate. If the likelihood that the opponent she faces is a low type is high enough, she will always prefer to litigate as her payoff from litigation goes to θ_1 whereas her payoff from settling outside court through negotiations is at most $\theta_1 - v_2(\theta_2^H)$.

The Existence of Litigation in a Semi-Seperating Equilibrium

I will now show that the logic that creates the existence of litigation in a pooling equilibrium generalises to create the existence of litigation in a semi-separating equilibrium.

Proposition 1.3. *When the conditions of proposition 1.1 are satisfied, a semi-separating equilibrium cannot exist.*

Proof. Note that in a semi-separating equilibrium the transfer to agent 2 must be independent of his declaration. If not the agent would have an incentive make the declaration that gets him the higher transfer. This transfer is $\tilde{\mu}_1$ and it must satisfy the litigation-proofness constraints defined in (1.14). This implies

$$\theta_1 - \tilde{\mu}_1 \geq \tilde{v}_1$$

and

$$\tilde{\mu}_1 \geq \tilde{v}_2^H \geq \tilde{v}_2^L.$$

Let $\frac{q_H(1-\gamma)}{q_H(1-\gamma)+q_L\gamma} = 1 - \tilde{\gamma}$. In the limit as $\theta_2^L \to 0$ this yields the following objective function for agent 1:

$$\theta_1 \left((1-\tilde{\gamma}) \frac{\alpha x_1^\lambda}{\alpha x_1^\lambda + (1-\alpha)\bar{x}_2^{H\lambda}} + \tilde{\gamma} \right) - x_1$$

since

$$\lim_{\theta_2^L \to 0} \bar{x}_2^L = 0.$$

The corresponding value function for a high type agent 2 is

$$\theta_2^H \frac{(1-\alpha)x_2^\lambda}{\alpha \bar{x}_1^\lambda + (1-\alpha)x_2^\lambda} - x_2$$

for a high type agent 2. Solving for the optimal effort levels and plugging them back into the objective function of agent 2 we get

$$\tilde{v}_2^H = \theta_2^H \left(\frac{(\theta_2^H)^\lambda}{((1-\tilde{\gamma})\theta_1)^\lambda + (\theta_2^H)^\lambda} - \frac{((1-\tilde{\gamma})\theta_1\theta_2^H)^\lambda \alpha(1-\alpha)\lambda}{(\alpha((1-\tilde{\gamma})\theta_1)^\lambda + (1-\alpha)(\theta_2^H)^\lambda)^2} \right)$$

Note that $\tilde{v}_2^H = v_2^H$ for $\gamma = 0$, where v_2^H is the separating equilibrium litigation payoff for a high type agent 2 defined in (1.20). Furthermore $\tilde{v}_2^H > v_2^H$ since $\frac{\partial \bar{x}_1}{\partial \gamma} < 0$.

Keeping $\tilde{\mu}_1 \geq \tilde{v}_2^H \geq v_2^H > 0$ there exists a threshold $\bar{q}_2^{*L}$ such that for $q_2^L > \bar{q}_2^{*L}$ agent 1 would prefer to litigate than to settle with a transfer of $\tilde{\mu}_1$ to agent 2. This is true since $\tilde{v}_1 \to \theta_1$ as $q_2^L \to 1$ Hence the maximum transfer that agent 1 is willing to make to avoid litigation is lower than the minimum needed to satisfy the litigation-proofness constraint of a high type agent 2 and litigation arises in equilibrium. □

The intuition for this result is similar to the previous result. In a semi-separating equilibrium agent 2 must receive a transfer independent of his declaration. This transfer must be greater than the minimum transfer required for keeping a high type agent 2 indifferent between litigation and settlement. However, as the likelihood of agent 2 being a low type increases, agent 1 prefers to litigate and 'take his chances' rather than pay out a high settlement.

Propositions 1.1, 1.3, and 1.2 taken together establish the existence of litigation in equilibrium. In a nutshell Proposition 1.1 shows that under certain conditions agent 2 would always lie about his type in any negotiation for pre-trial settlement. Hence it would not be possible for agents to resolve their informational asymmetry. Propositions 1.2 and 1.3 show that since agent 1 knows about this, she would prefer litigation to settling outside court. The results hold when the difference between θ_2^H and θ_2^L is large and the value of q_2^L is high enough. If these conditions are satisfied, agents have no option but to litigate.

1.3.2 Full Waiver of the Right to Litigate

As we would expect, if parties can contract away their right to litigate then this turns out to be sufficient to avoid litigation.

Proposition 1.4. *There always exists an unsubsidised and incentive compatible allocation that Pareto dominates the equilibrium allocation under litigation.*

Proof. Litigation is a Bayesian game, where the allocation is composed of the probabilities of the surplus being transferred to the two agents for the two possible types of agent 2 and the corresponding transfers. Using the revelation principle, any equilibrium allocation under litigation can be replicated by a direct mechanism that specifies probabilities of acquiring the surplus β_1 for agent 1 and β_2^j for a type j agent 2 and transfers x_1 and x_2^j. The incentive-compatibility constraints for agent 2 are

$$\theta_2^H(\beta_2^H - \beta_2^L) > x_2^H - x_2^L > \theta_2^L(\beta_2^H - \beta_2^L)$$

where

$$x_2^j > 0, \qquad j \in \{H, L\},$$

since litigation is costly. Similarly x_1 is the cost of litigation for agent 1. Consider an allocation where the probabilities β_2^j are preserved and litigation costs are replaced by the following transfers from agent 2 to 1

$$t_2^H = \theta_2^L(\beta_2^H - \beta_2^L) - x_1 \qquad t_2^L = -x_1$$

By construction we have

$$\theta_2^H \beta_2^H - t_2^H > \theta_2^L \beta_2^L - x_2^L.$$

This allocation is unsubsidised by a third party, incentive compatible, and preferred by both agents over litigation since the expected transfers they make are strictly lower than the costs of litigation. □

Note that the phrase 'Pareto dominance' is used here in the interim sense. Since litigation is simply a Bayesian game, applying the revelation principle, the equilibrium allocation of litigation can be replicated using a direct mechanism. Proposition 1.4 states that in fact, using a direct mechanism, a superior allocation can be implemented without having to subsidize the implemetation externally. Given litigation is costly for all parties, it is easy to see why this result obtains. The probabilities with which agents expect to win in court can be replicated in a direct mechanism. Compared to litigation this allocation reduces the amount of resources that are burnt for separation of types.

Proposition 1.4 implies that under full contractability litigation would never occur since it would be individually rational for agents to contract on a mechanism that guarantees a better allocation. In a world with full commitment, agents could write a contract wherein they commit to sticking with the allocation that the mechanism specifies. In such a world it would not be possible for agents to credibly threaten the other agent with litigation ex-post to force the renegotiation of the allocation. Hence these separating equilibrium allocations need not satisfy the additional constraint of litigation-proofness. This proposition is obvious when seen in the light of the well understood theoretical insight that the possibility of renegotiation ex-post creates incentive problems ex-ante[13].

1.3.3 Litigation Under Partial Waiver

The discussion in the preceding section raises the question of whether litigation would arise if a limited ability, to contract away their right to litigate, was available to agents. The degree of commitment available to parties can be thought of as a point in a continuum that is bounded by full contractibility on one end and complete non-contractibility on the other. A natural way to capture the partial commitment in the contest function specified in (1.18) is through α. Once agents sign a contract to stick to the allocations specified

[13]See for example Weitzman (1976).

by the mechanism it affects α when the case reaches court ex-post. In the world with complete contractibility, when agent 1 considers approaching the court ex-post, she would find that α equals 0. This means that agents would know that approaching the court in violation of the commitment to stay out of court would invite a certain ruling in favour of the opponent. The world with imperfect commitment, would be one where the value of α would change but the change would still not be sufficient to bring about complete certainty about the outcome of the case, that is, α ex-post would still be between 0 and 1. The result in Proposition 1.1 shows that with a low enough valuation for the low type, as long as α is strictly between 0 and 1, it would not be possible to satisfy incentive-compatibility. Hence as long as complete commitment is not available, it is possible to still apply propositions 1.1 and 1.2, and consequently justify the existence of litigation.

The area of law that governs the right of parties to contract away their rights, in this case the right to judicial remedy, is called waiver. Whether a waiver is valid is itself a contentious issue in law. Among other things, the court would verify whether "functional equivalence", that is some other form of judicial process was available to the agents. If the mechanism for resolving disputes looks fairly close to a judicial process, then court would be more likely to uphold the allocations. For example arbitral awards are open to appeal on very limited grounds. The problem with arbitration however is that in terms of the technology of decision making it is identical to the court. Therefore designing a settlement mechanism comes with the following tradeoff; the allocation it specifies is more likely to be upheld the more the mechanism resembles a court but this makes the mechanism costly in itself. This model does not explain when parties would choose arbitration or litigation but provides an explanation for why dispute resolution can be inherently costly.

The court would also look into the bargaining power between parties when it decides whether to uphold the mechanism designed by parties.[14] Since bargaining power is not verifiable, the decision of the court on the validity of waiver itself is subject to the same technology of decision making. This would mean that parties would have to take into account litigation-proofness constraints even when they add clauses waiving their right to litigate.

There could be several reasons why courts do not always enforce what contracting parties agree on ex-ante. There could be behavioural reasons for not allowing agents to tie themselves into contracts that are detrimental to them in the future. For example it is easy to see why court would void contracts where an agent sells himself in slavery to another.

There could also be efficiency based reasons. Anderlini et al. (2006b)

[14] An exposition of factors that courts usually take into account in the US while deciding on the legitimacy of waivers is discussed in Yale Law (1978) and Rubin (1980-1981)

argue that by committing to void certain contracts the court increases ex-ante efficiency. It is possible that similar considerations induce judges to void contracts where agents contract away their right to litigate. For example, consider a stage 0 that occurs in the timeline before the dispute arises. In this stage one of the two parties can take an action that is privately costly but which stops the dispute from arising. Now assume that a dispute comes with some inherent costs for both parties once it arises even if it is resolved efficiently. In the case of property disputes these can be thought of as the opportunity cost of sitting down to negotiate with the other agent. In case of a custody battle, one can think of this as the impact on the child of a dispute between parents. If the costs of the dispute are borne by both parties but the cost of the action that prevents the dispute are borne by just one agent, then there would be a tendency for too many disputes to arise. This would be mitigated if parties anticipated an inefficient settlement of disputes since that would increase the private costs of the dispute, and thereby increase the incentives for dispute prevention ex-ante. If these actions that prevent a dispute from arising are non verifiable, then by committing to be inefficient ex-post, courts increase efficiency ex-ante.

1.4 Extension: Committing Not to Trade Ex-Post

The results in the chapter have relied on the incentive-compatibility constraints that are restricted by trades of surplus that agents would voluntarily make once the allocations are assigned. Although the assumption that agents would exploit gains from trade ex-post is natural, it is not required for litigation to arise in equilibrium. This extension shows that a different assumption about the distribution of valuations along with the parameters of the contest function also creates litigation in equilibrium.

Proposition 1.5. *$\delta^j = 1 \; \forall j$ if $v_1^H + v_2^H > \theta_2^H$.*

Proof. Consider a state where the surplus is allocated to agent 2. In such a state the transfer from agent 2 to agent 1 must satisfy the following litigation-proofness constraints from (1.6):

$$\theta_2^j - v_2^j \geq t_2^j \geq v_1^j. \tag{1.24}$$

However if

$$v_1^j > \theta_2^j - v_2^j, \qquad j \in \{H, L\}; \tag{1.25}$$

then it is not possible to have transfers that are litigation-proof. This implies that the surplus must always be allocated to agent 1: $\delta^j = 1$. Equation (1.25), is composed of two constraints, one for each possible agent 2 type. Using the value functions defined in (1.20) we can check that the constraints for the state where agent 2 has low valuation is subsumed by the state where his valuation is high. That is $v_1^H > \theta_2^H - v_2^H$ implies $v_1^L > \theta_2^L - v_2^L$. In fact this condition turns out to be sufficient to ensure that $\delta^j = 1$ even when we consider pooling and semi-separating equilibria. To see this note that the corresponding constraint for these is

$$\tilde{v}_1 > \theta_2^j - \tilde{v}_2^j.$$

Since $\tilde{v}_1 + \tilde{v}_2^H \geq v_1^H + v_2^H > \theta_2^H$, condition (1.24) is sufficient to ensure $\delta^j = 1$ for all equilibria since the pooling equilibrium in subsumed in the treatment of semi-separating equilibrium for the case $\gamma = \frac{1}{2}$.

Equation (1.24) reduces to

$$(\theta_1 - \theta_2^H)\,\mathrm{P}(x_1^H, x_2^H) > x_1^H + x_2^H$$

The final constraint we get on the parameter space is

$$(\theta_1 - \theta_2^H)\frac{\theta_1^\lambda}{\theta_1^\lambda + \theta_2^{H\lambda}} - \frac{\theta_2^{H\lambda}\theta_1^{(1+\lambda)}\alpha(1-\alpha)\lambda}{(\alpha\theta_1^\lambda + (1-\alpha)\theta_2^{H\lambda})^2} - \frac{\theta_1^\lambda\theta_2^{H(1+\lambda)}\alpha(1-\alpha)\lambda}{(\alpha\theta_1^\lambda + (1-\alpha)\theta_2^{H\lambda})^2} > 0 \quad (1.26)$$

□

This result proves that under certain conditions it will not be possible to have an allocation where the probabilities of the surplus being transferred to agent 2 are positive. This is because in the event the surplus is allocated to agent 2, agents would find that the transfer to agent 1 does not satisfy litigation-proofness constraints. Therefore ex-ante agents would only contract on a mechanism that always allocates the surplus to agent 1. However if this happens, then $\delta_1^{ij} = 1$ and we are back in the world where proposition 1.1 applies. This result demonstrates that the assumption that agent are capable of trading the surplus ex-post is not crucial for litigation to arise.

An interesting testable implication about the incidence of litigation arises from this assumption. Equation (1.26) is more easily satisfied when the case is biased in favour of one of the two parties, that is, the value of α is close to zero or one. This is because equilibrium efforts are lower when α is close to zero or one. The intuition for this is that when the case is biased, parties spend less in court because the marginal impact of effort on the probability of winning is lower. This makes litigation less inefficient and consequently more likely.

1.5 Applications

The existence conflict has always been a puzzle. Rational explanations of conflict are based on the existence of informational asymmetry between agents. This informational asymmetry is preserved by restricting communication between parties in some way. The model presented here sheds some light on this issue by showing that regardless of how parties communicate, conflict may arise when parties are limited in their ability to commit.

The argument formalised in the model is that informational asymmetry between agents persists when agents are unable to contract away the possibility of renegotiating once they reveal their information. The impossibility of committing not renegotiate allocations once information is revealed, affects the incentives for truthfully revealing information. This insight is common to many types of conflict other than litigation.

In this section the application of the model to different kinds of conflict is discussed. The model sheds some light on the forces at work that prevent agents from effectively avoiding conflict. I also review some evidence that seems to be consistent with the predictions of the model.

1.5.1 Patent Litigation

In this model litigation arises due to unobservability of valuations. The model therefore predicts that the incidence of litigation should be negatively correlated with the degree of observability of valuations. This implies that less litigation should be observed in sectors where disputes are about objects over which agents are unlikely to have private valuation.

The model predicts that litigation over intellectual property would be expected in industries where a firm is likely to have private information on how much expected profits would arise if it succeeds in securing the patent in court. Conversely in an industry where the profitability of a patent is observable, litigation would be rare.

A related prediction regarding the incidence of litigation is the rate of litigation should be positively correlated with the variance of the distribution of valuation. In the model we saw how litigation arises only when the two values θ_2 take are sufficiently apart. In the limit as the variance goes to zero we are back in the world where valuations are observable. Depending on the use of the patent, firms are likely to have different valuations of the patent. Under the assumption that the variance of valuations increases with the possible uses a patent has, we should expect a positive correlation between the breadth of a patent and the incidence of litigation.

Lerner (1994) uses a data set where an index for the scope of a patent is constructed. Lanjouw and Schankerman (2004) studies the determinants of patent suits using data from US patent office, the federal courts and industry sources. In their data set they have measures for the market value of the patent. Together these data sets could be used to test the theory presented if. If the theory is correct, we would expect to find a positive correlation between the scope of a patent and the incidence of litigation even after controlling for things such as the market value of the patent.

1.5.2 War

Fearon (1995) argues that miscalculation of the opponent's willingness to fight is one of the causes of war. While discussing the incentives of states to reveal their true willingness to fight he states:

"While states have an incentive to avoid the costs of war, they also wish to obtain a favourable resolution of the issues. This latter desire can give them an incentive to exaggerate their true willingness or capability to fight, ... if they are concerned that revelation would make them militarily (and hence politically) vulnerable..."

The model presented here supplies the micro-foundations for this idea. Here the willingness to fight is determined by the valuation parties place on the subject matter in dispute. A low valuation agent takes into account the ex-post incentive of the opponent to threaten litigation once she finds out that he has low valuation. This vulnerability created by truthful revelation destroys the incentives for truthfully declaring ones valuation.

A historical example that seems to fit the argument formalised in this model is the Russo-Japanese conflict of 1904-05 over Korea and Manchuria. The primary reason for the conflict was the desire for exclusive economic control over Korea and Manchuria. In particular, both Russians and the Japanese had made significant investments in transport infrastructure in these regions. Their competing interest in securing exclusive control over these regions was a large factor in generating the conflict.

For instance, in early 1903 the Russians started lobbying in Korea for rights to construct a railway line between Seoul and Uiju. The Japanese were opposed to this since they wanted exclusive control over railway in Korea, being in the process of constructing a line between Seoul and Fusan. In Manchuria, Russia wanted exclusive control to protect the large investments in the Chinese-Eastern railway that was to facilitate transit of goods from ports on the Pacific Ocean into Russia. Furthermore the Russians were planning to build a port in Dalny for getting access to sea for the Chinese-Eastern Railway. The Japanese who controlled the port of Niuchuang were worried about the loss of trade resulting from the construction of a rival port.

There were several negotiations between the two countries in the time leading to the conflict. The first communication happened in 1901 in the aftermath of the boxer rebellion which presented the Russians with an opportunity to increase their influence over Manchuria. In early 1901 the Russians entered into an agreement with China that consolidated their power in Manchuria. The Japanese were strongly opposed to this agreement but the Russians never took this opposition too seriously, believing that the Japanese would never go to war against a strong western power.

In late 1901 Ito Hirobumi, a Japanese minister, travelled to Russia. There are accounts of his negotiations with the Russians that indicate how he attempted to convey to the Russians the Japanese desire for exclusive control over Korea. The Russians however were only willing to make concessions to the extent of sharing control over Korea. This position was continued in the final negotiations in December of 1903 when the Russians refused to accede to the Japanese demand for a neutral zone on the banks of the Yalu river in Korea. Furthermore the Russians refused to discuss the issue of Manchuria and maintained their stand that the Manchurian issue was not on the table.[15]

These accounts indicate that this instance of conflict has many of the ingredients that this model highlights. Both the Russians and the Japanese valued the control rights over Manchuria and Korea (see White (1964)). Furthermore, the Russians were unwilling to believe that Japanese sabre-rattling before the war was anything more than cheap talk and believed that Japan would be in a weak position in the event of a war. This example illustrates how the incentives of parties to always overstate their willingness to fight creates an informational asymmetry that leads to conflict. The opponent disbelieves any declaration about the willingness to fight and consequently agents are left with no option but to fight.

1.6 Conclusion

This chapter has attempted to solve two longstanding problems in the literature on why people litigate. The first problem is microfounding the presence of litigation through the existence of private information in a way that is consistent with full discolure theorems. The model proposed here tackles this issue by allowing all certifiable information to be disclosed at the pre-trial stage. Private information that creates informational asymmetries between parties is purely the non-certifiable component, which is the valuation that parties place on the subject matter in dispute. This influences the amount spent in court which consequently influences the expected payoff from litigation thereby making it unobservable.

[15]See Nish (1985) for a rich account of the negotiations between Russia and Japan preceding conflict.

The second problem that this chapter tackles is the restriction that the literature has placed on the pre-trial interaction between parties. The literature so far has assumed that parties can only interact in a bargaining framework where they communicate through offers and counteroffers. By studying settlement in the framework of mechanism design, this chapter allows for richer communication between parties.

The main insight supplied here is that if the possibility of committing to alternative mechanisms for dispute resolution is limited, then this dilutes the incentives for truth telling. In further work it would be interesting to develop a normative theory of the judiciary using this model where the possibility of inefficient litigation ex-post may create incentives for efficient behaviour ex-ante. This ties back to the conception of courts in neo classical economics with a slight twist: courts by their very existence deter undesirable behaviour that leads to disputes by ensuring that parties cannot efficiently negotiate themselves out of disputes once they arise.

Chapter 2

Should Courts Enforce Waiver of Remedial Rights?

2.1 Introduction

Economic theory predicts that disputes between agents are off the equilibrium path and should therefore not happen. There are strong forces backing this prediction. Agents are aware that resolving a dispute takes up resources that often have a considerable opportunity cost. This makes agents behave in ways that pre-empt the creation of disputes. Hence the terms of contracts are always followed and property rights are never infringed.

However in the real world disputes between economic agents arise with an alarming frequency. This seems to be inefficient since scarce resources are diverted away from productive activities into dispute resolution. This is especially the case when disputes end up in court. In the previous chapter of this thesis I have argued that disputes are litigated when parties lack the ability to contract away their right to litigate at the pre-trial stage. Given that disputes are costly and litigating them is even costlier, the question that naturally arises here is why don't courts enforce such contracts when doing so would reduce the costs of resolving a dispute once it arises? This chapter attempts to answer this question.

The first ingredient in the argument presented here is that disputes are inefficient because an agent only takes into account her own costs of resolving the dispute and not the costs that her opponent would have to bear. This leads to too many disputes. Secondly, the agent who started the dispute cannot be identified since courts are constrained in their ability to observe what actually happened in the past. Thirdly a dispute is typically unobservable unless it ends up in court. Building on these elements, the model presented here shows that a benevolent social planner may choose not to enforce the waiver of remedial rights to increase ex-ante welfare. This result arises because ensuring that disputes end up in court makes agents internalise the costs of a dispute and this avoids the creation of disputes in the first place. Once a dispute arises it would be optimal for courts to never step in since they are costly. However the real possibility of costly courts, created by the non enforcement of waiver, raises ex-ante welfare by ensuring that some fraction of disputes don't arise to start with.

This chapter is related to the literature on optimal incidence of litigation. This literature seeks to uncover mechanisms that determine the efficient level of litigation in society. Shavell (1997), one of the first papers in this literature, pointed out that the amount of litigation may not equal the efficient amount. In that paper litigation can exceed the efficient level since parties do not take into account their opponent's cost from litigation. This can lead to too much litigation. Similarly parties fail to take into account the social benefits such as the value of the precedent created through litigation, and the social costs of litigation such as the cost of maintaining a judicial system. Again these factors can drive a wedge between the observed level and the efficient level of litigation.

Another paper in this literature on the question of social costs and benefits of litigation that are not internalised by parties is Hua and Spier (2005). The authors argue that the information revealed during trial about the liability of the defendant has positive externalities on potential defendants in the future since they can fine tune their level of care based on the information revealed during the trial. Since parties only care about their private benefits and costs, they do not internalise this effect.

It is well understood that when parties interact in an environment of asymmetric information inefficiencies arise. However it is unclear whether that generates the role for an interventionist court when the court does not have access to private information of the parties. This chapter is related to a small literature that explores this issue. The paper by Anderlini et al. (2006b) shows that courts can increase ex-ante welfare by voiding some contracts. In their paper parties contract under asymmetric information and inefficient pooling equilibria can obtain. When courts void certain kinds of contracts with positive probability, the pooling equilibria are weeded out and parties are forced to separate. This result is generalised further in Anderlini et al. (2006a). These papers make the general point that there is a role for an interventionist court when parties contract in an environment of asymmetric information. The mechanism that generates the result their paper is different than this chapter. In Anderlini et al. (2006b), what generates the inefficiency is the fact that courts maximise ex-ante surplus but parties only contract at the interim stage where their own type is already known to them. In contrast what is emphasised here is the externality that is imposed on the opponent when an agent decides to raise a dispute.

The following is the outline for the chapter. Firstly the model of how disputes arise is presented in section 2.2. The equilibrium of the game between the two agents is presented in section 2.3, and it is shown how this equilibrium varies with enforcement and non enforcement of waivers. In section 2.4 the efficiency properties of the equilibrium presented in 2.3 are analysed. The main results are presented in section 2.5. In section 2.6 the role of various informational assumptions is discussed. In this section the different types of inefficiencies that arise in this model are also discussed. Finally section 2.7 concludes.

2.2 Model

The setup is similar to the previous chapter of this thesis. The innovation is that the arrival of a dispute is no longer exogenous. Instead a dispute arises if at least one party starts it. The dispute is over a surplus that agent 1 values at θ_1 and agent 2 values at θ_2^H with probability q and θ_2^L with probability $(1-q)$. Agent 2's valuation is only known to him.

Initially the surplus is either with agent 1 with probability $(1-\phi)$ and agent 2 with probability ϕ. The initial allocation of possession implies that in the case the dispute does not arise, the agent with the initial possession continues to enjoy the surplus. The property rights over this surplus are fuzzy. This simply means that if the dispute goes to court, there is a non-zero probability that either of the two parties will win. Therefore if a dispute arises between the two agents, they will negotiate over the allocation of the surplus in exchange of some transfers. Just like the previous chapter of this thesis, the payoff from litigation is the outside option to negotiations.

An agent chooses between the status quo allocation or starting a dispute. The private cost of the dispute to an agent is c that is either c^H or c^L with probability δ and $(1-\delta)$. If a dispute is successfully negotiated, one party is allocated the surplus in exchange for a transfer to the other party. An unsuccessful negotiation leads to the dispute going to court.

Timeline:

Stage 1: The status quo allocation of surplus is decided.

Stage 2: Agents realise c_1 and c_2 and non-cooperatively decide whether to begin a dispute.

Stage 3: If dispute arises, parties start pre-trial negotiation which is a game that may help them avoid taking the matter to court.

Stage 4: Parties play the game from stage 2 and receive the equilibrium allocation.

Stage 5: Parties non-cooperatively choose between the equilibrium allocation from stage 2 and approaching the court.

Stage 6: If either agent has approached the court, then both non co-operatively choose their effort levels.

Stage 7: Court makes a final decision on the allocation of the surplus.

This model applies to all the disputes that are mentioned in the previous chapter. At this point it is useful to sketch a particular example to fix ideas. Consider a case where there are two firms; firm 1 and firm 2. Firm 1 owns an intellectual property right over widget 1. Firm 2 begins production widget 2 which is similar to widget 1. It is unclear whether the two widgets are sufficiently similar for the production of widget 2 to be a violation of the property right of firm 1. In response to firm 2's production of widget 2, firm 1 has two options. It can either choose to do nothing or it can choose to send a notice to firm 2 about the potential violation of firm 1's right. If it chooses the latter course, and firm 2 is unwilling to stop production of widget 2, then a dispute arises between the two firms.

2.2.1 Litigation

Just like the previous chapter of this thesis, litigation is modeled as a contest between the two agents. Following are the objective functions of the two agents.

$$\theta_1 \mathrm{P}(x_1, x_2) - x_1 \qquad \text{and} \qquad \theta_2^j(1 - \mathrm{P}(x_1, x_2)) - x_2 \qquad j \in \{L, H\}$$

where θ_1 is the valuation of agent 1 and θ_2^j is the valuation of agent 2. Note that henceforth j refers to the type of agent 2. It is assumed that the probability of winning is increasing in ones own effort and decreasing in the effort of the opponent.

If the valuation of agent 2 was observable, then this would be a game of complete information where we could compute the Nash equilibrium effort levels of the agents. Here, since his valuation is private information, agent 1 instead plays a Bayesian game where the optimal effort level of agent 1 is:

$$x_1(\theta_1) = \operatorname*{argmax}_{x_1 \geq 0} \theta_1 \left(\sum_{j \in \{H,L\}} q^j \, \mathrm{P}\left(x_1, x_2(\theta_2^j)\right) \right) - x_1. \tag{2.1}$$

Similarly the optimal effort level of type j agent 2 is

$$x_2(\theta_2^j) = \operatorname*{argmax}_{x_2 \geq 0} \theta_2^j \, \mathrm{P}\,(x_1(\theta_1), x_2) - x_2. \tag{2.2}$$

Having computed the Bayesian Nash equilibrium effort levels for agent 1 and 2, we can work out the expected payoff from litigation for the agents. These are

$$v_1(\theta_1) = \theta_1 \left(\sum_{j \in \{H,L\}} q^j \, \mathrm{P}\,(x_1(\theta_1), x_2(\theta_2^j)) \right) - x_1(\theta_1). \tag{2.3}$$

and

$$v_2(\theta_2^j) = \theta_2^j q^j \, \mathrm{P}\,(x_1(\theta_1), x_2(\theta_2^j)) - x_2(\theta_1^j) \qquad j \in \{H, L\}. \tag{2.4}$$

2.2.2 Negotiations

Once disputes arise agents are inclined to resolve them as efficiently as possible. To this end parties would negotiate. It is impossible to predict what game form the negotiation takes. Agents could bargain over the surplus. Such bargaining may be stretched out over multiple periods or one of the two agents may have the bargaining power enabling her to make a take it or leave it offer to her opponent. Moreover, in addition to bargaining, there may be cheap talk involved with agents making threats of dubious credibility. Rather than imposing a game form on the negotiations, I use a mechanism design approach where equilibrium allocations of any game parties play can be replicated using a direct mechanism. This insight is known as the revelation principle.

The success or failure of negotiations depends on whether courts enforce waiver. This section analyses the outcome of negotiations in the two cases of enforcement and non enforcement of waiver. The direct mechanism that captures the equilibrium of negotiations is composed of probability of being allocated the surplus and the corresponding transfer that the agent makes. Let β_i^j be the probability with which the surplus is allocated to agent i of type j and t_i^j be the corresponding transfer that the agent makes to her opponent when she is allocated the surplus.

Negotiations with Waiver

If agents are allowed to commit not to litigate, they will do so as long as their expected payoff from negotiations is greater than the expected payoff from litigation. Let the expected payoffs from negotiations be $\mu_1(\theta_1)$, $\mu_2(\theta_2^H)$ and $\mu_2(\theta_2^L)$. For negotiations to be successful it must be the case that

$$\mu_i(\theta_i^j) \geq v_i(\theta_i^j) \qquad \forall i, j \qquad i \in \{1, 2\} \qquad j \in \{\Phi, H, L\}.$$

In words this simply means that negotiations will be successful if the expected payoff for both agents is from negotiation is greater than the expected payoff from litigation. In proposition 1.4 it was shown that as long as agents can commit not to litigate, such $\mu_i(\theta_i^j)$ always exist. This implies that the participation constraint for agents to choose negotiations would always be satisfied.

The exact allocation that comes about as a result of negotiations depends on the relative bargaining power, or more precisely negotiation power, of the two agents. However it is possible to characterize the most efficient allocation that can arise as an equilibrium through negotiations. The logic of focusing attention on the most efficient equilibrium allocation is explained in section 2.4.1.

Note that the most efficient allocation is one where the surplus is allocated to agent 1 in return for some transfer to agent 2. However this allocation cannot be an equilibrium in the space of parameter values captured in the previous chapter. Recall that β_2^H is the probability with which the direct mechanism allocates the surplus to a high type agent 2. The equilibrium allocation must satisfy $\beta_2^H > 0$. This is shown in the next lemma.

Lemma 2.1. *$\beta_2^H > 0$ in any negotiation equilibrium when proposition 1.2 holds.*

Proof. Firstly note that a pooling equilibrium cannot exists when proposition 1.2 holds. Hence we can focus on a separating equilibrium. Since $\beta_2^H \geq \beta_2^L$, and we want to show that $\beta_2^H > 0$, we can set $\beta_2^L = 0$ to find the lower bound of β_2^H. For separation to be possible IC constraints for agent 2 must be satisfied. These reduce to

$$\mu_1(\theta_1) = \big(q(1-\beta_2^H) + (1-q)\big)\theta_1 + t_2^H - \beta_2^H\theta_2^L,$$

$$\mu_2(\theta_2^H) = \beta_2^H\theta_2^H - t_2^H, \qquad \mu_2(\theta_2^L) = \beta_2^H\theta_2^L - t_2^H.$$

Furthermore, these must satisfy the participation constraints

$$\mu_i(\theta_i^j) \geq \nu_i(\theta_i^j) \qquad \forall i, j \qquad i \in \{1, 2\} \qquad j \in \{\Phi, H, L\}.$$

for agents to prefer negotiations over litigation. If $\beta_2^H = 0$ then the transfer made to agent 2 $-t_2^H = \nu_2(\theta_2^H)$. However as proposition 1.2 shows, $\nu_1(\theta_1) > \theta_1 - \nu_2(\theta_2^H)$. Hence if $\beta_2^H = 0$ then the participation constraint of agent 1 cannot be satisfied. □

If surplus is always allocated to agent 1, then the transfer to agent 2 must be independent of his declaration. However as this is not possible since the transfer that then needs to be made by agent 1 to agent 2 is too large for agent 1 to consent to negotiations. Hence some inefficiency, in the form of the surplus being allocated to agent 2, must ensue. This implies $\mu_1(\theta_1) + q^H\mu_2(\theta_2^H) + q^L\mu(\theta_2^L) < \theta_1$, that there is some loss of efficiency inherent in the negotiation process.

Negotiations without Waiver

As shown in the previous chapter of this thesis, when waivers are not enforced by courts, negotiations break down. This is because agents anticipate that the statements of their opponents are not credible. In the case where waivers are enforced, the participation constraint is an ex ante constraint. This means that agents choose negotiations as long as the expected payoff from negotiations is larger than litigation before the type of the opponent is revealed. Now however, the constraint also becomes an ex post constraint. Agent should prefer his payoff to the litigation payoff for any declaration of agent 2 type. Since this is a much stronger requirement, pre-trial negotiations break down and litigation occurs.

2.2.3 Dispute

This section models the decision of an agent to start a dispute. A dispute arises when two agents find themselves in a situation where both are laying claim on the same surplus.

In the status quo agent 2 is endowed with the surplus with probability ϕ and agent 1 with probability $(1-\phi)$. Once the surplus is allocated, agents realise their costs of starting a dispute. With probability δ the costs are c_H and with probability $(1-\delta)$ the costs are c_L with $c_H > c_L$. The costs of the two agents are independently drawn.[1] Whenever the surplus is allocated to an agent it is assumed that her opponent prefers to start a dispute as long as the dispute yields the negotiation payoff. Mathematically this implies

$$\mu_i(\theta_i^j) - c_k > 0 \qquad \forall i, j, k \qquad i \in \{1, 2\} \qquad j \in \{\Phi, L, H\} \qquad k \in \{L, H\}. \tag{2.5}$$

Given this assumption, a dispute always arises whenever waivers are enforced. This is because the payoff from a successful negotiation is assumed to be higher than the costs of the dispute for the agent that is not endowed with the surplus. Agent i of type j knows that pre-trial negotiations will be successful and will result in the negotiation payoff $\mu_i(\theta_i^j)$. On the other hand when instead the agent faces the prospect of costly litigation, the a dispossessed agent wants

[1]The results are qualitatively similar when costs of the two agents are correlated. The results are strengthened (weakened) when the correlation between the costs is positive (negative).

to start a dispute only when the cost is low. This assumption implies

$$v_i(\theta_i^j) - c_L > 0 > v_i(\theta_i^j) - c_H \qquad \forall i,j \qquad i \in \{1,2\} \qquad j \in \{\Phi, L, H\}. \quad (2.6)$$

2.3 Equilibrium

It is now possible to characterise the equilibrium of this game.

Lemma 2.2. *When waivers are enforced, disputes always arise. When waivers are not enforced, disputes only arise when the costs of the dispossessed agent are low.*

Proof. Follows trivially from (2.5) and (2.6). □

Given the assumption on costs of the agents it directly follows that in the full waiver regime disputes always arise. This is because the negotiation payoff $\mu_i(\theta_i^j)$ is always larger than the costs for the dispossessed agent. On the other hand, if waivers are not enforced, disputes only arise when the costs of the dispossessed agent are low. This happens because the agent anticipates that once disputes arise they are resolved in a court and courts are costly.

2.4 Social Surplus

The equilibrium of the game between agents was described in the previous section. In this section the welfare properties of that equilibrium are analysed. The analysis will be limited to evaluation of the total surplus under the two possible court policies, that is, when waivers are enforced and when they are not. Note that the surplus will be computed from an ex-ante stage where the costs of the agents are yet to be realised. This is the correct position from which to calculate the total surplus if the welfare consequences of the waiver policy on incidence of disputes is to be evaluated.

2.4.1 Full Waiver

When courts enforce waivers, disputes arise regardless of the costs of the agents. Hence the expected total surplus is

$$\mu_1(\theta_1) + q^H \mu_2(\theta_2^H) + q^L \mu(\theta_2^L) - 2\,\mathrm{E}(c) \qquad (2.7)$$

From lemma 2.1 we know that the expected surplus from negotiations is less than θ_1. Hence we can rewrite equation (2.7) as

$$(1-\tau)\theta_1 - 2\,\mathrm{E}(c) \qquad \text{where} \quad 0 < \tau < 1.$$

Assume that the surplus when disputes don't arise is greater than the surplus when disputes arise. This implies

$$\phi\,\mathrm{E}(\theta_2) + (1-\phi)\theta_1 > (1-\tau)\theta_1 - 2\,\mathrm{E}(c) \tag{2.8}$$

At this point, a clarification for the use of the most efficient negotiation equilibrium is required. What is assumed here is that even the best possible negotiation allocations are inferior to status quo since agents do not internalise the costs of the dispute that are incurred by their opponents. It is well understood that with complete information bargaining leads to efficient allocations regardless of how relative bargaining power is distributed between the agents. In an environment of incomplete information however, the level of inefficiency may depend on the particular game form and consequently on the bargaining power of agents.[2] Hence it is possible that the actual equilibrium that ensues as a result of negotiations is less efficient. However the argument that is made in this chapter is that since negotiations may be too efficient, agents may raise too many disputes and this is undesirable. If negotiations are inefficient in themselves then this automatically dampens the private incentives to create disputes thereby reducing the number of disputes. In such cases the policy of not enforcing waivers will have no effect since negotiations themselves would be inefficient enough to deter disputes. However non enforcement of waivers will make a difference if the equilibrium that arises through negotiations is 'too efficient'.

2.4.2 No Waiver

When waivers are not enforced by courts, disputes only arise when costs of the dispossessed agent are low. This is shown in lemma 2.2. This implies that the expected surplus without waivers is

$$\delta\,(\phi\,\mathrm{E}(\theta_2) + (1-\phi)\theta_1) + (1-\delta)\left(v_1(\theta_1) + \mathrm{E}(v_2(\theta_2^j)) - c_L - \mathrm{E}(c)\right)$$

With probability δ the dispossessed agent has high costs. When this happens no dispute arises and we get the first part of the expressions. On the other hand, if the agent has low costs, a dispute arises and we get the second part of the expression. This happens with probability $(1-\delta)$.

2.5 Result

If the court acts as a social planner that maximises ex-ante social surplus, we have the following proposition.

[2] See Ausubel et al. (2002) for discussion of this issue and a survey of the related literature.

Proposition 2.1. *There always exists a $\delta^* < 1$ such that for a $\delta > \delta^*$, it is optimal for court not to enforce the waiver of remedial rights.*

Proof. Given the assumption in equation 2.8, the surplus is always higher when disputes don't arise. Recall that the dispossessed agent only raises a dispute when costs are low when waivers are not enforced. Hence there exists a δ^* such that

$$\delta^* \left(\phi \, \mathrm{E}(\theta_2) + (1-\phi)\theta_1\right) + (1-\delta^*)\left(v_1(\theta_1) + \mathrm{E}(v_2(\theta_2^j)) - c_L - \mathrm{E}(c)\right)$$

$$= (1-\tau)\theta_1 - 2\,\mathrm{E}(c).$$

And for any $\delta > \delta^*$, the first expression must be strictly larger. □

The intuition for this result is as follows. Surplus is always greater when disputes don't arise. The enforcement of waiver maximises the surplus conditional on disputes arising. When waivers are not enforced, agents are faced with costly litgation. The threat of costly litigation however implies that disputes, under the non enforcement of waiver, only arise if costs are low. Recall that δ is the probability with which an agent has high costs. Hence when δ is high, disputes arise less frequently.

The optimal policy that would replicate the first best would be if courts could set a fine that makes the agent that starts the dispute internalise the externality she imposes on her opponent. This however is not possible for two reasons. Firstly the agent who bears the responsibility for starting a dispute cannot be identified. Secondly it is not possible for courts to police all disputes. Courts are limited in the exercise of their judgment to the proportion of disputes that actually end up in court. The presence of these handicaps create a second best world where courts find it optimal to use the non-enforcement of waiver of remedial rights as a way to optimise the number of disputes that arise and hence to maximise social surplus.

Figure 2.1: Social Planner's Problem

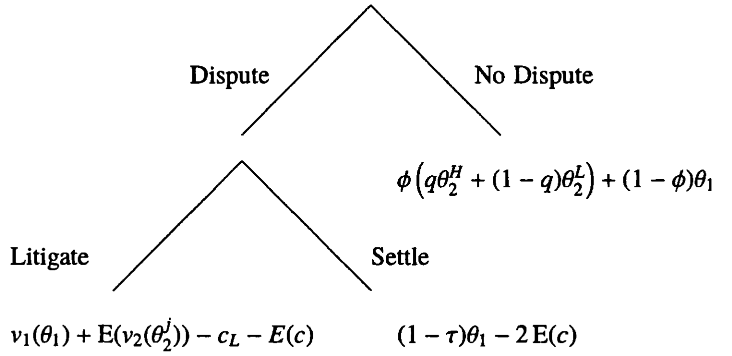


Figure 2.5 shows what the problem looks like from the point of view of a social planner. It is not the extensive form of the game. The moves in the figure cannot be directly controlled by the planner. The social planner can influence v_1 and v_2 by choosing λ to affect the equilibrium effort level. If the effort levels are high, then litigation is inefficient. Furthermore the social planner in the form of a court can choose not to enforce contracts where agents have waived their right to litigate. It will be shown in the results that by doing so courts, under certain conditions, rule out the possibility of successful negotiations. Since parties anticipate ex-ante that a dispute will not be effectively negotiated and is therefore likely to end up in court, they prefer to avoid actions that would create a dispute. This leads to an increased social surplus ex-ante.

The primary implication of non-enforcement of waiver is to strengthen the position of the agent that is endowed with the surplus is status quo. Since dispute resolution is made costly as a result of this policy, the opponent's incentive to create a dispute is dampened. In this model it is assumed that this is always a good thing since the total costs of the dispute outweigh the benefits of the dispute that arise through the potential reallocation of the surplus to the party with the higher valuation.

In addition to the ingredients of the previous result if we assume further that

$$v_1(\theta_1) + \mathrm{E}(v_2(\theta_2^j)) - \mathrm{E}(c) > \phi\,\mathrm{E}(\theta_2) + (1-\phi)\theta_1 \tag{2.9}$$

then the following proposition emerges

Proposition 2.2. *There exists c_L^* such that for $c_L < c_L^*$, keeping* $\mathrm{E}(c)$ *constant, non enforcement of waiver maximises total surplus.*

Proof. When waivers are not enforced, disputes only arise when costs of the dispossessed agent are low. Hence the surplus is

$$v_1(\theta_1) + \mathrm{E}(v_2(\theta_2^j)) - c_L - \mathrm{E}(c).$$

There must exist a c_L^* such that the two sides of equation (2.9) are equalised, and below which non enforcement of waiver dominates. □

Unlike proposition 2.1, this proposition demonstrates the trade-off between the two types of inefficiencies in the model. In the status quo the surplus may be inefficiently allocated to the agent with the lower valuation (ϕ is high). Hence it may be optimal to allow some disputes. However there is an inefficiency to disputes in form of costs that are not internalised by agents. This inefficiency dominates when the costs are high. Hence it is socially optimal to allow disputes when the costs of one of the agents is low. The non enforcement of waiver allows the courts to exactly implement this outcome. By making dispute resolution inefficient enough, courts ensure that disputes only arise when the costs to the dispossessed agent are low.

2.6 Discussion

2.6.1 Role of Informational Environment

In this section I discuss the role that the informational environment plays in generating this result.

The Identity of the Agent that Started the Dispute is Non-Verifiable

This assumption is needed because if the social planner could observe the identity of the agent who begins the dispute, then she could simply tax that party. This would be the efficient way of stopping disputes from arising and waivers would no longer be needed. The example of the two firms engaged in a dispute over intellectual property rights clarifies how both the agents could be responsible for starting the dispute. Firm 1 can stop the dispute from arising by deciding to allow the firm 2 to continue its production of widget 2. Similarly firm 2 can voluntarily stop the production of widget 2 and thereby avoid the dispute. Any other interaction between the 2 firms, such as transfer from firm 1 to firm 2 for stopping the production of widget 2, is ruled out here till stage 3 commences. It is only in stage 3 that the two firms can come together and negotiate. However at this stage the dispute costs are already sunk since coming to the negotiation table is costly for both firms.

Incompleteness in the Law

It should be noted that the act of starting the dispute may be completely independent of the legal position in the case. Dispute can only arise when precise predictions about the outcome of the case cannot be made. Going back to the example, if the intellectual property rights were clearly defined, say in favour of firm 1, then firm 2 would know any dispute that arises would be settled in favour of party 1. This would deter disputes. It is the incompleteness in the law that causes disputes to go to court. It is natural however to assume that the law is incomplete. Firstly is impossible for law makers to envisage all possible disputes that can arise. Furthermore the existence of a law that completely specifies the allocation of surplus in every conceivable state seems implausible since the state space can be extremely rich. And lastly, even if such a complete law could be written, it would be of limited use unless courts could verify the true state to implement the corresponding allocation.

2.6.2 Inefficiencies

The inefficiencies generated by the model can be classified in four groups. In this section I discuss how these arise, why they are a necessary ingredient for the result, and how they can be mapped to inefficiencies that we observe in the real world.

Inefficient Status Quo Allocation of the Surplus

The initial allocation of the surplus is made to agent 2 with probability ϕ and agent 1 with probability $(1 - \phi)$. Since $\phi > 1$ there is an inefficiency in the initial allocation since the surplus stays with agent 2 with some probability even though agent 1 always values it more. Inefficiencies of this sort seem quite common place in reality unless one believes that existing allocations are pareto efficient.

Externality of the Costs of Dispute

The main inefficiency that drives the result in this model is fact that agents do not take into account their opponents costs when they decide to start a dispute. This implies that when a dispute begins there is an externality on the opponent. The costs that are envisaged here are the costs of meeting the opponent. The managers of the two firms that engage in a dispute need to take time out of productive activities and devote it to set up meeting with each other. In addition to the time of the managers, the firms also need to divert resources away from productive activities into dispute resolution. Firms typically set up in house specialists that deal with dispute resolution such as a legal department. To the extent that the expenditure on these is proportional to the number of disputes that are expected to arise in the future, these can all add up to costs that are externalities imposed by one firm on another.

Inefficiencies in Negotiation

As noted in lemma 2.1, there is an inefficiency in negotiations. This arises because in any equilibrium that arises in negotiations, must entail a non zero probability with which the surplus goes to a high type agent 2. This is inefficient since agent 1 always has higher valuation of the surplus. This inefficiency is not a necessary ingredient for the result of this model to arise but nonetheless, it emerges as a feature of any pre-trial negotiation.

Litigation Costs

The costs of litigation arise as a result of the contest like nature of litigation and as such is a general feature of all contests. This inefficiency plays a crucial role in generating the result. It is the threat of costly litigation, generated by the non-enforcement of waiver, that deters disputes from arising.[3] This inefficiency is traded off with the inefficiency of the costs of the dispute and consequently agents are made to partly internalise the effect of their decision to start a dispute on their opponent.

2.7 Conclusion

The story that is captured here is very simple. Once disputes arise, it is always efficient to allow parties to settle as efficiently as possible. This happens when courts enforce waiver of remedial rights. However disputes are costly over and above the cost of litigation. When parties decide to raise a dispute they do not internalise the costs of dispute that their opponents have to bear. Hence if waivers are not enforced, then the anticipation of costly dispute resolution can increase the surplus as this deters some disputes from arising.

[3] In addition to the waiver policy, the social planner decides on the elasticity of the contest function with respect to effort to optimize how much inefficiency is generated in equilibrium in litigation. She can do so by varying the value of λ. This has been left unmodeled.

Chapter 3

Can Market Failure Cause Political Failure?

3.1 Introduction

It is well known that market failures abound in the real world. A key insight in the institutional approach to development economics is that capital market failures prevent individuals and economies from reaching their full potential and can lead to poverty traps (see Banerjee and Newman (1993) and Galor and Zeira (1993)). In this literature institutional frictions are taken as exogenous.[1]

It is also well known that even fully accountable governments can fail to implement growth maximising policies when they lack sufficient instruments for compensating losers. Furthermore, the political economy approach to development has emphasized how concentration of political power in the hands of an elite, may lead to distortion of the market by the elites for maximising their own payoffs.[2] This strand within the political economy literature makes the argument that the distribution of political power may be sufficiently skewed so as to allow the elites to distort the market outcome in their favour, and this typically leads to inefficiencies.

In this chapter we highlight the reverse link, namely that market failure may create a political failure even when political power is uniformly distributed. We think of political failure as the failure of the electorate to pick the surplus maximising reform.[3] In our model, in the first best world with well functioning markets, the electorate unanimously chooses institutions that maximise total surplus. However once a market imperfection in the form of

[1] See Banerjee (2001) for a survey of this literature.

[2] This is most obvious when elites lobby for barriers to entry (Djankov et al. (2002)). Acemoglu (2003) makes the argument that concentration of political power may lead to distortion of the market through manipulation of factor prices in ways that benefit the political elites.

[3] For a discussion on somewhat different notions of political failure see Besley (2006).

unobservability of entrepreneurial talent is introduced, things change dramatically. The competitive market responds to this imperfection by screening agents based on their wealth. This leads to creation of a class structure in the economy with preferences that are aligned in ways that defeat surplus maximising reforms. In a nutshell, the motivation for this chapter is to uncover the political implications of market failure.

There is an important distinction between our approach and the existing literature on political economy. Instead of taking political classes or interest groups as exogenous and studying the impact of their alignment on markets, we derive them from economic fundamentals, namely, the nature of technology, and the informational environment in the economy. In this regard, the mechanism that our chapter identifies fits into a theme present in both Marxist and Neo-Classical theories of institutions that use economic forces as the base over which the political superstructure is built.[4]

We argue that in addition to the well known impact of market failure articulated in the literature on poverty traps, there may also be a political impact. The latter problem could turn out to be more persistent since unlike the solutions to poverty traps that are easier to characterise[5], the solutions to political failure that are politically feasible may not exist. A more general message emerging from our model is that market and political failures complement each other in terms of generating economic inefficiencies.

This chapter is related to the growing literature on micro political economy. This literature looks at failure of alternative institutions and asks two questions:

1. Which institutions make an economy more productive?

2. Which institutions are more likely to be chosen given a certain distribution of political power?

We now present a review of papers that ask similar questions. Boyer and Laffont (1999) present a model where a monopolist produces a socially valuable good and some amount of pollution as a byproduct. The regulator has a choice of several instruments that can be used to make transfers to the monopolist. The incentives of the electorate may not be aligned with those of a total surplus maximising regulator since the electorate is composed of voters of whom a certain proportion are also shareholders in the monopoly. This can lead to non surplus maximising policies being chosen, regardless of information asymmetries.

Perotti and Volpin (2004) have a model where agents are endowed with wealth and are either consumers or entrepreneurs. There is a non convexity in the production function and entrepreneurs with wealth lower than a certain

[4] See chapter 1 in Bardhan (1989) for a review of the common themes in these literatures concerning the theory of institutions.

[5] Micro-lending has been a big theme in this literature. See for example Ghatak and Guinnane (1999).

threshold are financed by equity. Project returns are subject to the ex post moral hazard problem and investor protection, the institution that they study, can mitigate the problem. Elites that have wealth over the threshold required to start an enterprise, lobby for lower investor protection so as to face a lower competition in the product market. The political economy process is modelled as a social planner that maxmises the weighted sum of the total surplus and bribes from lobbies. As the weight on the bribes increases, investor protection goes down.

Rajan and Zingales (2006) study a model evaluating the incentives of the educated and non-educated class to pass educational and pro market reform. Educational reforms allow the uneducated to become educated and increase their wages through an increase in their productivity. Pro market reforms allow educated workers to setup their own firms. An agent's preference for any reform is driven by which group the agent belongs to.

Biais and Mariotti (2003) address the question of optimal bankruptcy laws. They have a model of occupational choice where agents can be entrepreneurs or workers. Credit market is imperfect because entrepreneurial effort is unobservable. The mechanism through which bankruptcy law affects total surplus is the following: a tough bankruptcy law implies a strong threat of liquidation ex-ante. This induces high effort which increases surplus. However liquidation is ex-post inefficient since some surplus is lost when a company is harvested for its assets at liquidation. In terms of the political economy aspects, the rich want soft laws to induce lower wages. The poor want the opposite. The agents with intermediate wealth align with rich if they are entrepreneurs and align with poor otherwise. This paper is similar to ours in the sense that here too a market failure generates the need for institutions. The paper differs from this chapter in terms of the result they find on the choices an electorate make. In their model soft laws which are often chosen by the electorate are often efficient due to inefficiency of liquidation ex post. In contrast, our results indicate that their exists an inherent tension between politically feasible and surplus maximising reforms.

Another paper that is related to this chapter is Caselli and Gennaioli (2008). In their model agents differ in two discrete dimensions; talent and license. There is an exogenous mismatch between talent to run an enterprise and the endowment of license that is required to run an enterprise. They model how this exogenously conferred incumbency and talent interact to create preferences for deregulation and legal reform. Deregulation lowers the cost of acquiring a new license whereas legal reform makes the trade of licenses between agents easier.

In these models, markets can be complete and perfectly competitive if the best possible institutions are chosen. In absence of such institutions, frictions are created that take the economy away from the growth maximising outcome. The source of problems in these models is purely the exogenous presence of political alignments that undermine the support for best possible institutions. In our model on the other hand these political alignments are endogenised and the fundamental source of inefficiency will be the adverse selection problem created by the unobservability of entrepreneurial talent. Institutions, depending on their quality, would mitigate or worsen this problem.

In our model, even with fully benevolent government and perfectly competitive markets, there are market frictions arising from informational (i.e., adverse selection) and transactional constraints (limited liability). As in the standard neoclassical model, preference and technology differences might have seemingly similar implications: e.g., in the Solow model, low steady state output could result from lower saving propensity or use of less efficient technology. However, the policy implications are dramatically different: preference differences are more intractable than technology differences and this is especially so if we recognize the potential mutual interaction of preferences and technology adoption which, for example, reflects some underlying market failure. Analogously, we argue that with government frictions the policy implications are to be found in the political domain and are relatively easy to characterize which is not to say they are easy to implement: improve political institutions to improve the quality of candidates, improve incentives for incumbents so that inefficient rent-extracting policies are removed. In contrast, with market frictions the policies are far less easy to characterize, and this is especially so if they interact with an otherwise frictionless political system where the distribution of political power is uniform.

3.2 Model

The basic setup extends the model presented in Ghatak et al. (2007).

3.2.1 Technology

There are two technologies in the economy: a subsistence technology that yields $\underline{w}$ with certainty for one unit of labour and a more productive technology y that yields a return R in case of success and 0 in case of failure and requires n workers and 1 entrepreneur to run it.

3.2.2 Preferences

All agents are assumed to be risk neutral with a utility function that is additively separable in effort and money. The net disutility of labour effort relative to entrepreneurial effort is normalised to M. This can also include any perks that entrepreneurs enjoy relative to workers such as the comfort of sitting in an air conditioned office, or the psychological payoff from not having a boss.

3.2.3 Endowments

Agents are endowed with one unit of labour, entrepreneurial talent and illiquid wealth. Talent θ of an agent is the probability of success of the more productive technology if she becomes an entrepreneur. θ is distributed with a cdf $F(\theta)$. Agents are endowed with illiquid wealth a with a distribution $G(a)$. We assume that the distributions of wealth and talent are independent.

3.2.4 Informational and Institutional Frictions

The entrepreneurial ability θ can be either observable or unobservable. In the first best world θ is observable and the first welfare theorem operates ensuring that the competitive equilibrium is Pareto efficient. In contrast when θ is unobservable, a market failure arises. The illiquid wealth a, and output y, are verifiable. M is also verifiable but is not appropriable since it is the psychological net benefit of being an entrepreneur.

The 2 institutional parameters in the model are ϕ and τ. ϕ is the proportion of collateral that is recovered from a borrower when she defaults. This can be thought of as the strength of judicial enforcement of contracts. τ is the probability with which the wealth a is expropriated. The efficiency of both these institutions affect the credit contract that an agent is offered in the second best world as the credit market takes into account the efficiency of the judiciary and the risk of expropriation when accepting the agent's wealth as collateral. We discuss this in greater detail in section 3.5.

In addition to these institutional variables, a limited liability constraint also operates in the economy. This implies that in the event an entrepreneurial project fails, the agent can only be liable upto the illiquid asset a. In other words agents are guaranteed a non negative payoff in all states of the world.

3.2.5 Occupational Choice

Agents choose their occupation. They can either choose to work in the subsistence sector, become workers, or become entrepreneurs. They are paid a wage w at the end of the period if they choose to work for a wage. If they choose entrepreneurship, their payoff is stochastic. The project succeeds with a probability θ which is the unobservable talent of the agent. To set up a firm an entrepreneur needs to hire n workers and pay them a wage w up front. Where $w \geq \underline{w}$ since working with the subsistence technology is an outside option that all agents have.

Our assumption that the productive technology requires n workers and 1 entrepreneur implies that workers and the entrepreneur are perfect complements in the production function. This assumption greatly simplifies our analysis and allows us to get sharp political economy results, though is not central to our analysis.

3.2.6 Markets

We will present a general equilibrium model with two markets; the labour and credit market. The need for credit arises as workers need to be paid up front when an entrepreneurial project is set up and the wealth of agents is illiquid. Both the markets are assumed to be perfectly competitive. The risk free interest rate is assumed to be zero.

3.3 Credit Contracts

Since the wealth of an agent is illiquid, agents need to borrow from the credit market to become entrepreneurs. The credit market is assumed to be perfectly competitive. The supply of credit is assumed to be perfectly elastic at interest rate equal to 1.

3.3.1 First Best

If talent was observable, then credit contract would not be based on collateral due to the presence of contractual friction ϕ that arises when collateral is used. Hence an agent with talent θ would be offered a contract with an interest rate $\frac{1}{\theta}$. Since the mass of entrepreneurs in this economy cannot exceed $\frac{1}{n+1}$, in equilibrium the wage would ensure that agents with talent less than θ^* become workers where

$$\theta^* : \int_{\theta^*}^{1} f(\theta)d\theta = \frac{1}{n+1}.$$

For the labour market to be in equilibrium, an agent with talent lower than θ^* should prefer working for a wage and agents with greater talent should prefer entrepreneurship. This implies that in equilibrium the agent with talent θ^*, who is indifferent between working for a wage and becoming an entrepreneur, has the following occupational choice condition:

$$\theta^* \left(R - \frac{n\overline{w}}{\theta^*} \right) + M + (1 - \tau)a = \overline{w} + (1 - \tau)a.$$

We can rearrange this condition to back out the equilibrium wage $\overline{w}$ at which the labour market clears:

$$\overline{w} = \frac{\theta^* R + M}{n+1}. \tag{3.1}$$

It follows that in the first best world the value of ϕ will not matter since wealth will not be used in the credit contract. In contrast, the value of τ would matter since an increase in τ would lower the expected payoff of agents due to the increased risk of expropriation.

We assume that

$$\overline{w} > M > \underline{w}.$$

The first part of the assumption ensures that the returns from the project when it succeeds are large enough to make interest payments.[6]

The second part of the assumption, $M > \underline{w}$ is necessary for the existence of a credit constraint in this economy.[7]

3.3.2 Second Best

The second best world is characterised by the unobservability of entrepreneurial talent. In all other respects it is identical to the first best world. Since talent is unobservable, the credit market can no longer offer contracts that are indexed by the agent's talent. However agents are endowed with wealth which they can use as collateral to access credit. Hence the credit contract will be defined by a pair (r, a) that is, interest rate and collateral.

We now discuss the possible credit contracts that can be offered to entrepreneurs and we characterise the equilibrium in the credit and labour market. The reader interested in the choice of institutions by the electorate in the first and second best world can see the figure in section 3.4 that captures the characterisation of the equilibrium and skip directly to section 3.5.

Separating Contract

Let us first consider the separating contracts that can be offered to the agents. A separating contract exists if the contract is such that agents have an incentive to reveal their types. Since the probability of success is increasing in type,

[6] Note that he interest rate offered to entrepreneur with talent θ^* is $\frac{1}{\theta^*}$. Backing out the value of θ^* from equation (3.1), we can check that

$$R - \frac{n\overline{w}}{\theta^*} > 0$$

is satisfied when $\overline{w} > M$.

[7] Consider an agent with zero wealth and talent. He would be attracted to entrepreneurship only if $M > \underline{w}$. Hence if this condition is not satisfied, his occupational choice condition in the second best world would be such that he would prefer working for a wage when $R - rnw < 0$ and consequently there may not be a credit constraint in the economy. The existence of a credit constraint introduces interesting results. We discuss this in greater detail in section 3.5.

agents with higher entrepreneurial ability are offered contracts with lower interest rates. This feature of the credit contract creates an incentive to lie for low ability agents. Hence for such contracts to be incentive compatible, agents need to have sufficient wealth that the credit market can use as a screen. The separating contract is defined by the incentive compatible pair $(r_s(\theta, a(\theta)), a(\theta))$ which is the interest rate and the collateral that is offered to an agent with talent θ. The wealth level below which a separating contract is not feasible is determined by the constraint

$$R - r_s nw \geq 0$$

holding with an equality. This is shown by the following lemma.

Lemma 3.1. *No separating contract (r_s, a) can exist if $R < r_s nw$*

Proof. In the appendix. □

The intuition for this result is the following. When $R < r_s nw$, the entire return from the project has to be handed over to the bank when the project succeeds. In addition to R, agents also need to hand over a proportion of their wealth when the project succeeds. This additional requirement makes separation impossible. This happens because the separating contracts that are offered to high types are ones that return a large proportion $\gamma(a)$ of collateral in the success state. However these contracts are attractive to all agents that choose entrepreneurship regardless of their type.

Given Lemma 3.1, we can restrict our attention to the region where $R > r_s nw$. In this region, the zero profit condition for the bank is

$$\tilde{\theta} r_s(a) nw + (1 - \tilde{\theta})(1 - \tau)\phi a = nw \tag{3.2}$$

when lending to an agent of type $\tilde{\theta}$. Similarly, the feasibility condition for the loan is

$$R - r_s(a) nw \geq 0.$$

At the point where this feasibility constraint binds, we can find the talent of the least talent agent that becomes an entrepreneur by plugging in the zero profit condition, the feasibility condition for the loan, and the agent's occupational choice condition to find the lowest level of talent and collateral that is consistent with a separating contract. The occupational choice constraint of an agent indifferent between entrepreneurship and working for a wage is

$$\theta(R - r_s(a) nw) + M - (1 - \tau)(1 - \theta)a = w.$$

When the feasibility constraint of the loan binds, we have $R = r_s(a) nw$. We can now back out the talent of the least talented agent who could become an entrepreneur. This is $\underline{\theta}_s$ such that

$$M - (1 - \tau)(1 - \underline{\theta}_s)a = w. \tag{3.3}$$

Using equations (3.2) and (3.3) along with the feasibility condition for the loan, we can substitute out the equilibrium interest rate to find the expressions for $\underline{\theta}_s$ and $\underline{a}_s$, the lowest level of talent and wealth that are consistent with the existence of a separating contract. These are

$$\underline{\theta}_s = \frac{nw - \phi(M - w)}{R} \tag{3.4}$$

and

$$\underline{a}_s = \frac{(M - w)R}{(1 - \tau)(R - nw + \phi(M - w))}. \tag{3.5}$$

$\underline{a}_s$ is a threshold wealth below which a separating contract is not feasible.

The strategy for deriving the separating contract schedule is the following. Equation (3.2) gives us the expression for the interest rate that is charged to an agent with type $\tilde{\theta}$. An agent with type θ has an incentive to declare his true type if a truthful declaration maximises his payoff from entrepreneurship. Hence if a separating contract can be designed such that a truthful declaration by the agent globally maximises her payoff from entrepreneurship, then we can say that such a separating contract is incentive compatible.

The existence of the separating contract depends on the existence of a type dependent collateral schedule that is implementable. In other words, letting $\tilde{\theta}$ be the type that an agent declares in a direct mechanism, if we can find a schedule of collateral $a(\tilde{\theta})$ such that agents find it optimal to declare their true types $(\tilde{\theta} = \theta)$, then $(r_s(\tilde{\theta}, a(\tilde{\theta})), a(\tilde{\theta}))$ is a separating contract. It is optimal for an agent of type θ with wealth a to declare her type truthfully if:

$$\operatorname*{argmax}_{\tilde{\theta}} v_\theta(\tilde{\theta}) = \theta \tag{3.6}$$

where

$$v_\theta(\tilde{\theta}) = \theta\left(R - r(\tilde{\theta}, a(\tilde{\theta}))nw + (1 - \tau)a(\tilde{\theta})\right) - (1 - \tau)a(\tilde{\theta}) + M. \tag{3.7}$$

The first order condition of this problem yields a differential equation that we can use to solve for the collateral schedule $a(\tilde{\theta})$ such that agents have an incentive to reveal their types truthfully. This is

$$a(\tilde{\theta}) = \frac{nw}{\phi(1 - \tau)}\left(1 - \frac{(R - nw)}{nw}\left(\frac{\underline{\theta}_s}{1 - \underline{\theta}_s}\right)^{\frac{1}{1-\phi}}\left(\frac{1 - \tilde{\theta}}{\tilde{\theta}}\right)^{\frac{\phi}{1-\phi}}\right). \tag{3.8}$$

The steps for the derivation of $a(\tilde{\theta})$ and the expression showing the concavity of the objective function are in the appendix.

The uniqueness of the solution to the differential equation tells us that there exists a unique collateral schedule such that agents find it optimal to declare their types truthfully. Since the zero profit condition for the banks is embedded into the expression for the interest rate in the objective function of the agents. We can recover the interest rate by plugging in the collateral schedule $a(\tilde{\theta})$ into the interest rate $r_s(\tilde{\theta}, a(\tilde{\theta}))$. It is possible to check that $a(\tilde{\theta})$ is monotonically increasing in $\tilde{\theta}$. High types are willing to post higher collateral since the value that an entrepreneur places on the reduction in the interest rate relative to the increase in collateral is increasing in her type. When the type of the agent is the highest possible, that is, $\theta = 1$, the corresponding collateral is $\overline{a}$ and the interest rate charged is 1.

Pooling Contract

In addition to a separating contract, there may also exist pooling contracts in this economy. Unlike the separating contract that is only available when $R \geq r_s(a)nw$, a pooling contract is possible both for the region of wealth that satisfies the corresponding condition, and also for a certain interval of wealth where this constraint is violated.

Let us first consider the region of wealth such that $R \geq r_p(a)nw$. Any pooling contract that could be offered must satisfy the necessary condition of zero profit for competitive banks:

$$r_p(a)\theta_p(a)nw + (1 - \theta_p(a))(1 - \tau)\phi a = nw. \tag{3.9}$$

Like we saw in the case of the separating contract, we can use the occupational choice constraint of the agents to evaluate the talent of the least talented agent that chooses entrepreneurship. This is $\hat{\theta}$ such that

$$\hat{\theta}(R - r_p(a)nw) + M - (1 - \tau)(1 - \hat{\theta})a = w \tag{3.10}$$

Now let us consider the zero profit condition for banks when $R < r_p nw$. In this region, in addition to the project returns R, the banks also need to be pledged a proportion of collateral for them to break even. The zero profit contract is now defined by

$$\theta_p(a)(R + (1 - \gamma(a))(1 - \tau)\phi a) + (1 - \theta_p(a))(1 - \tau)\phi a = nw. \tag{3.11}$$

where $(1 - \gamma(a))$ is the proportion of collateral that is taken over by the bank in case the project succeeds. It is important to note that entrepreneurship is attractive not just because of the appropriable return R but also for the non-appropriable return M. If the latter is large enough, agents would be willing to choose entrepreneurship in exchange for their wealth even in the case when the project succeeds. Note that this formulation implicitly assumes that the optimal contract is one where all wealth is seized when the agent defaults. On the other hand, when the project succeeds, the minimum wealth $a\phi(1-\tau)(1-\gamma(a))$ that satisfies the zero profit condition of the bank is seized. It is easy to see that the pooling contract will take this form since this is the preferred contract for agents with high talent. Agents with high talent succeed with a higher probability and hence, relative to less talented agents, prefer contracts that are tougher in the bad state and yield a high payoff in the good state. Note that $\gamma(a)$ is increasing in a since banks would have to appropriate a larger share of wealth in the good state to satisfy the zero profit condition when the agent has lower wealth.

In both these regions, $\theta_p(a)$ is the average talent in the pool at wealth level a:

$$\theta_p(a) = \frac{1}{1 - F(\hat{\theta}(a))} \int_{\hat{\theta}(a)}^{1} \theta f(\theta) d\theta \tag{3.12}$$

and $\hat{\theta}(a)$ is the agent with the lowest talent in the pool, who must be indifferent between working for a wage and becoming an entrepreneur with the pooling contract. In the region where $R < r_p nw$ this is determined by

$$M - (1-\tau)(1 - \hat{\theta}(a)\gamma(a))a = w. \tag{3.13}$$

Plugging (3.12) in (3.11), the system of two equations (3.11) and (3.13) simultaneously determines the $\gamma(a)$ which can be thought of as the pooling interest rate and the lower bound $\hat{\theta}(a)$ of types that could choose the pooling contract $(\gamma(a), a)$ if they have wealth a. However, there exists a lower bound of wealth below which banks are not willing to offer such a contract. Note that credit contracts can only be offered when

$$\theta_p(a)R + (1-\tau)\phi a \geq nw.$$

This condition only holds when agents have sufficient wealth. This in turn defines the wealth level $\underline{a}_p$, such that agents with wealth less than this threshold will not be offered a pooling contract. Note that at this wealth level $\gamma(a) = 0$ must hold since agents would have to forgo their entire wealth in order to secure the credit contract.

$$\underline{a}_p = \frac{nw - \theta_p(\underline{a}_p)R}{\phi(1-\tau)}. \tag{3.14}$$

Substituting this, and $\gamma(\underline{a}_p) = 0$ into the occupational choice condition (3.13) of the marginal agent who is indifferent, we find at this wealth level, all agents choose entrepreneurship. $\hat{\theta}(\underline{a}_p) = 0$ and

$$\theta_p(\underline{a}_p) = \int_0^1 \theta f(\theta) d\theta. \tag{3.15}$$

This implies that at the lowe$\theta_p(\underline{a}_p) = \int \theta f(\theta)d\theta$ at is consistent with the pooling contract, all agents prefer to become entrepreneurs.

Lemma 3.2. *The lower bound of talent in a pool at any given wealth class is weakly increasing in wealth.*

Proof. In the appendix. □

In words, starting from $\underline{a}_p$, an agent of a higher wealth class receives a lower interest rate but has a greater loss in case of failure, and this second effect always dominates for an agent at the bottom of the talent distribution. Hence entrepreneurship is more attractive to less talented agents when they have less wealth, since they have less to lose in case of default. Since these agents prefer working for a wage at high levels of wealth, the quality of the pool of borrowers is weakly increasing in wealth. The maximum wealth level for which a pooling contract can be acceptable is given by

$$\overline{a} = \frac{nw}{(1-\tau)\phi} \tag{3.16}$$

such that the pooling interest rate drops to 1. This is the level of collateral that will be charged in a pooling contract when the interest rate equals one.

3.4 Equilibrium

In the previous section we have discussed the types of credit contracts that can exist in the economy. We are now ready to characterise the equilibrium.

3.4.1 Equilibrium in the Credit Market

We have shown that both pooling and separating contracts are viable. Given that banks can introduce any contract $(r(a), a)$ we will now characterise the equilibrium in the model. We will use the Rothschild Stiglitz equilibrium concept where an equilibrium is characterised by the conditions: *i)* all the contracts in the equilibrium set make non negative profits and *ii)* non existence of a contract that can be introduced that will make a strictly positive profit. We will assume that $\underline{a}_p > 0$. It is easy to check that $\underline{a}_p < \underline{a}_s < \overline{a}$. Hence there is no contract that can be offered to (and accepted by) an agent with wealth $a < \underline{a}_p$ that will make non negative profits.

Lemma 3.3. *There exists a level of wealth $\hat{a}_p$ defined by $\underline{\theta}_s = \hat{\theta}(\hat{a}_p)$ where $\overline{a} > \hat{a}_p > \underline{a}_s$ such that the only contract in the equilibrium set for $a < \hat{a}_p$ can be a pooling contract.*

Proof. Recall that $\hat{\theta}(a)$ is the level of talent such that an agent with this talent is indifferent between becoming an entrepreneur with the pooling contract $(r_p(a), a)$ and working for a wage. Since the distribution of wealth is continuous, there exists a level of wealth $\hat{a}_p$ such that an agent with talent $\underline{\theta}_s = \hat{\theta}(\hat{a}_p)$ is indifferent between both these alternatives and the separating contract $(r(a(\underline{\theta}_s)), a(\underline{\theta}_s))$. At $\underline{a}_s$ the agent with type $\underline{\theta}_s$ prefers the pooling to the separating contract since she receives a cross subsidy. At $\hat{a}_p$ the attractiveness of the cross subsidy disappears since the collateral requirement becomes too high. Hence even though a separating contract is feasible at $\underline{a}_s$ it is not incentive compatible for an agent with type $\underline{\theta}_s$ to accept it. It becomes incentive compatible only when the agent has wealth $a \geq \hat{a}_p$ at which point he prefers $(r_s(a(\underline{\theta}_s)), a(\underline{\theta}_s)$ to $(r_p(\hat{a}_p), \hat{a}_p))$ □

Lemma 3.4. *In the region of wealth $a \in (\hat{a}_p, \overline{a})$ there exists a level of talent $\hat{\theta}_s(a)$ such that agents with talent $\theta > \hat{\theta}_s(a)$ prefer the pooling contract and agents with talent $\theta \leq \hat{\theta}_s(a)$ prefer the separating contract.*

Proof. Note that for $a \in (\hat{a}_p, \overline{a})$ a fully separating contract schedule is not available since the collateral required for full separation of types is $\overline{a}$. $\frac{\partial \hat{\theta}(a)}{\partial a}$ implies that the attractiveness of the pooling contract is increasing in type. This is obvious since it simply captures the fact that more wealth is better for screening than less. This implies the existence of a cutoff talent $\hat{\theta}_s(a)$ for level of wealth $a \geq \hat{a}_p$ such that it becomes possible to offer agents with talent $\theta \leq \hat{\theta}_s(a)$ a separating contract that they prefer to the pooling contract. Note that $\hat{\theta}_s(\hat{a}_p) = \hat{\theta}(\hat{a}_p) = \underline{\theta}_s$ and $\hat{\theta}_s(\overline{a}) = 1$ □

Proposition 3.1. *[Existence and Uniqueness] A unique credit market equilibrium exists such that agents with wealth a:*

- $a \geq \overline{a}$*: are offered separating contracts*
- $\overline{a} > a > \hat{a}_p$*: are offered both pooling and separating contract*
- $\hat{a}_p > a > \underline{a}_p$*: are offered pooling contracts*
- $\underline{a}_p > a$*: are credit constrained*

Proof. $a < \underline{a}_p$ are credit constrained since no contract that makes non negative profits can be offered to these agents. Lemma 3.3 shows that only a pooling contract can exist in the region of wealth $a < \hat{a}_p$. Lemma 3.4 shows that in the region of wealth $\overline{a} > a > \hat{a}_p$ agents with talent $\theta \leq \hat{\theta}_s(a)$ a separating contract and $\theta > \hat{\theta}_s(a)$ accept a pooling contract. For the region of wealth $a \geq \overline{a}$ a fully separating schedule of contract exists that is offered and accepted by agents. This is a unique equilibrium since the zero profit pooling and separating contract schedules are unique. □

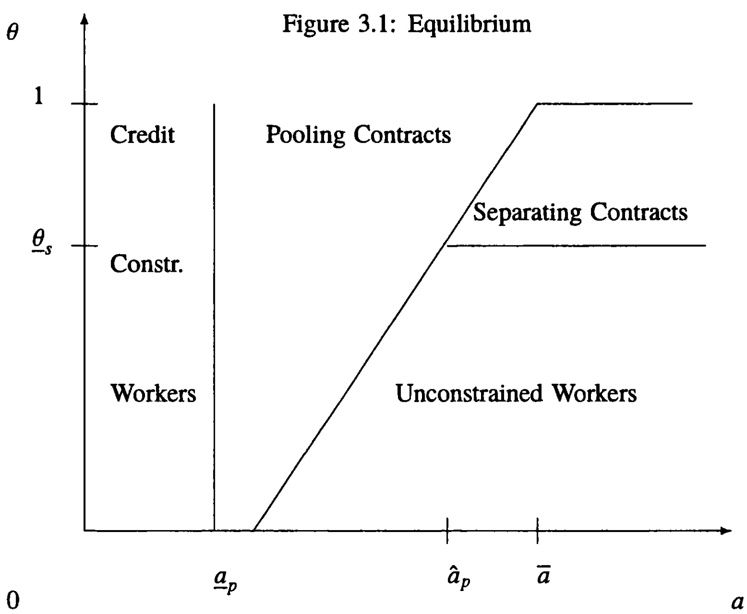
Proposition 3.2 (Occupational Choice). *Agents with wealth:*

- $\underline{a}_p > a$ *become workers*
- $\hat{a}_p > a \geq \underline{a}_p$ *and talent* $\theta \geq \hat{\theta}(a)$ *accept the pooling contract and become entrepreneurs and the rest become workers*
- $\overline{a} > a > \hat{a}_p$ *and talent* $1 \geq \theta > \hat{\theta}_s(a)$ *accept the pooling contract and become entrepreneurs; and talent* $\hat{\theta}_s(a)) \geq \theta \geq \underline{\theta}_s$ *accept the separating contract and become entrepreneurs, and the rest become workers*
- $a \geq \overline{a}$ *and talent* $\theta \geq \underline{\theta}_s$ *accept the separating contract and become entrepreneurs and the rest become workers*

Proof. Follows from lemma 3.2 to lemma 3.4 and proposition 3.1. □

The following figure presents a graphical representation of the equilibrium. As seen in the figure, we can conveniently analyse the equilibrium in terms of four regions of wealth:

Figure 3.1: Equilibrium



- Region 1: $\underline{a}_p > a$ are credit constrained and become workers;

- Region 2: $\hat{a}_p > a \geq \underline{a}_p$ and talent $\theta \geq \hat{\theta}(a)$ accept the pooling contract and become entrepreneurs and the rest become workers;
- Region 3: $\overline{a} > a > \hat{a}_p$ and talent $1 \geq \theta > \hat{\theta}_s(a)$ accept the pooling contract and become entrepreneurs; and talent $\hat{\theta}_s(a)) \geq \theta \geq \underline{\theta}_s$ accept the separating contract and become entrepreneurs, and the rest become workers;
- Region 4: $a \geq \overline{a}$ and talent $\theta \geq \underline{\theta}_s$ accept the separating contract and become entrepreneurs and the rest become workers.

At wealth level $\underline{a}_s$ it is possible to offer a separating contract to the agent with talent $\underline{\theta}_s$. However at this wealth level the agent with talent $\underline{\theta}_s$ will always accept the pooling contract since he pledges the same level as collateral but receives a lower interest rate with the pooling contract because of the cross subsidy. As the wealth of this agent increases, the pooling contract that is offered becomes less attractive since the pooling contract always requires an agent to pledge all his wealth as collateral. At wealth level $\hat{a}_p$ the agent prefers to take the separating contract with collateral $\underline{a}_s$ rather than take the pooling contract with wealth $\hat{a}_p$. Hence though separating contract is feasible from wealth level $\underline{a}_s$, in equilibrium they are only seen from wealth level $\hat{a}_p$. Because of this reason, as the level of wealth rises the talent of the least talented agent who accepts the pooling contract also rises. Similar to region 2 where there is no separating contract, this happens because agents with higher wealth prefer to become workers due to the high collateral requirement for being an entrepreneur. In region 3 however, this happens due to the high collateral requirement of the pooling contract relative to the separating contract.

In region 3 take a specific wealth level a. The agent with talent $\hat{\theta}_s(a)$ is indifferent between the separating contract that is offered to him and the pooling contract and accepts the separating contract. The agents with talent $\hat{\theta}_s(a) > \theta > \underline{\theta}_s$ strictly prefer the respective separating contracts they are offerred. Now consider an agent with talent greater than $\hat{\theta}_s(a)$. I will show that all agents in this group prefer the pooling contract rather than accepting the separating contract offered to the agent with talent $\hat{\theta}_s(a)$. Agent with talent $\hat{\theta}_s(a)$ (lets call this $\hat{\theta}$ to ease notation) who is indifferent between pooling and separating contract implies:

$$v_s(\hat{\theta}) = \hat{\theta}(R - r_s(a_s)nw) + M + \hat{\theta}(1-\tau)a_s + (1-\tau)(a - a_s) \tag{3.17}$$

where a_s and $r_s(a_s)$ are the collateral and interest rate for the separating contract defined by (3.8). The value for this agent from the pooling contract is

$$v_p(\hat{\theta}) = \hat{\theta}(R - r_p(a)nw) + M + \hat{\theta}(1 - \tau)a. \tag{3.18}$$

Equating $v_s(\hat{\theta}) = v_p(\hat{\theta})$ we get

$$r_p(a)nw = r_s(a)nw - \frac{(1-\tau)(1-\hat{\theta})}{\hat{\theta}}(a - a_s). \tag{3.19}$$

Take a $\tilde{\theta} > \hat{\theta}$. I will now show that agent with talent $\tilde{\theta}$ will prefer the pooling contract to the separating contract offered to agents with talent $\hat{\theta}$. Using the expression for the pooling interest rate in equation (3.19) we have:

$$v_p(\tilde{\theta}) = \tilde{\theta}\left(R - r_s(a_s)nw + \frac{(1-\tau)(1-\hat{\theta})}{\hat{\theta}}(a - a_s)\right) + M + \hat{\theta}(1 - \tau)a. \tag{3.20}$$

On the other hand the value from mimicking $\hat{\theta}$ and accepting the pooling contract is

$$v_s^{\hat{\theta}}(\tilde{\theta}) = \tilde{\theta}(R - r_s(a_s)nw) + M + \tilde{\theta}(1 - \tau)a_s + (1 - \tau)(a - a_s). \tag{3.21}$$

Equating $v_p(\tilde{\theta})$ and $v_s^{\hat{\theta}}(\tilde{\theta})$ we find that the pooling contract dominates for all $\tilde{\theta}$. Hence types greater than $\hat{\theta}$ prefer to post the higher collateral and get the pooling contract rather than take the separating contract offered to agent $\hat{\theta}$.

In the region of wealth $a < \hat{a}_p$ the talent of the least talented agent is $\hat{\theta}(a)$. However when a separating contract becomes feasible the nature of this function that determines the talent of the least talented agent in the pool changes somewhat hence we call it $\hat{\theta}_s(a)$. $\hat{\theta}_s(a)$ is defined by 2 conditions: The first condition determines the feasibility of the separating contract, i.e equation (3.8). Since equation (3.8) is monotonically increasing in θ, it is invertible. Expressing equation (3.8) as a function of theta we have: $\theta_s(a)$. The second condition that determines $\hat{\theta}_s(a)$ is the condition that determines the indifference between the payoffs from the pooling and separating contract for the agent. Let us call this $\hat{\theta}(a)$. Hence in region 3 we have:

$$\hat{\theta}_s(a) = \max\{\hat{\theta}(a), \theta_s(a)\} \tag{3.22}$$

This is because it is possible for either of the two constraints to be slack in this region. It is possible that agents prefer the separating contract but the separating contract is simply not feasible in which case $\theta_s(a)$ would bind. Alternatively it is possible that the separating contract is feasible but agents at the lower end prefer the pooling contract. In this case $\hat{\theta}(a)$ would bind.

Now consider the threshold at which region 3 begins. If $\theta_s(a)$ binds here then we have $\hat{a}_p = \underline{a}_s$. However this is not possible for the following reason. If $\theta_s(a)$ binds at the threshold this implies that $\hat{\theta}(\underline{a}_s) \leq \theta_s(\underline{a}_s)$. This implies that even though the agent with talent $\underline{\theta}_s$ is made to post the same level of collateral and receives a lower interest rate with the pooling contract, he still prefers the separating contract. This is not possible. Hence at the beginning of region 3 $\hat{\theta}(a)$ binds and we have $\hat{a}_p > \underline{a}_s$. Note that as $\hat{\theta}(a)$ follows continuously from from region 2, the transition from region 2 to 3 from $\hat{\theta}(a)$ to $\hat{\theta}_s(a)$ is continuous. Thereafter $\hat{\theta}_s(a)$ is continuous since both $\hat{\theta}(a)$ and $\theta_s(a)$ are continuous. Note that it is possible that there could be finite points where $\hat{\theta}(a)$ and $\theta_s(a)$ cross each other making $\hat{\theta}_s(a)$ non differentiable. However this does not affect the monotonicity property of $\hat{\theta}_s(a)$ since both $\hat{\theta}(a)$ and $\theta_s(a)$ are monotonically increasing in a.

3.4.2 Equilibrium in the Labour Market

The labour market is perfectly competitive. An equilibrium is characterised by the demand equalling supply. It is much easier to characterise the equilibrium by thinking of the labour demand of a firm instead of the labour demand by an entrepreneur. A firm demands 1 unit of entrepreneurial and n units of non entrepreneurial labour. Supply is 0 for wage $w < \underline{w}$, and 1 at $w = \underline{w}$. Labour demand is given by:

$$L_d = (n+1)\left(\int_{\underline{a}_p}^{\hat{a}_p}\left(1 - F(\hat{\theta}(a))\right)g(a)da + (1 - F(\underline{\theta}_s))(1 - G(\hat{a}_p))\right) \quad (3.23)$$

Proposition 3.3. *The equilibrium wage is $\underline{w}$ when $L_d(\underline{w}) \leq 1$ $w > \underline{w}$ when $L_d(\underline{w}) > 1$*

Proof. Note that Labour demand is monotonically decreasing in the wage:

$$\frac{\partial L_d}{\partial w} = (n{+}1)\left(-g(\underline{a}_p)\frac{\partial \underline{a}_p}{\partial w} - f(\underline{\theta}_s)\frac{\partial \underline{\theta}_s}{\partial w}(1 - G(\hat{a}_p)) - \int_{\underline{a}_p}^{\hat{a}_p} f(\hat{\theta}(a))\frac{\partial \hat{\theta}(a)}{\partial w}g(a)da\right) < 0 \quad (3.24)$$

since

$$\frac{\partial \underline{a}_p}{\partial w} > 0 \qquad \frac{\partial \underline{\theta}_s}{\partial w} > 0 \qquad \frac{\partial \hat{\theta}(a)}{\partial w} > 0 \quad (3.25)$$

If Labour demand is less than 1, there is excess supply of labour in the economy and the wage must equal $\underline{w}$ which is the outside option to working for a wage. If the labour demanded at $w = \underline{w}$ is more that 1, then the economy is tight in the sense that no one is engaged in the subsistence sector, and the wage must increase to equilibriate demand and supply.

□

The proof shows that there are two effects that create a labour demand that is monotonically decreasing in wage.

Firstly as the wage increases, the amount that entrepreneurs need to borrow also increases. This drives up the credit constraint.

Secondly, as wage increases, the agent with the lowest talent that was previously indifferent between entrepreneurship and paid employment now prefers paid employment. This is because the increase in wage tips his occupational choice constraint. Both these effect imply a reduction in labour demand for an increase in wage.

Note that in this economy $M > w$ is necessary and sufficient for there to be a credit constraint. If the equilibrium wage rises above this then the bank's zero profit condition is satisfied even at 0 wealth. We will assume that the equilibrium wage is lower than M since the problem without credit constraint is not interesting to analyse.[8]

3.5 Credit Market Institutions

The argument we make is that when interest groups are created in an imperfect market, then this can lead to an inefficient choice of institutional reform. In the first best world where talent is observable, the best institutions are chosen. As we move away from the first best world, there is not only a market inefficiency created by the unobservability of talent, but also a political inefficiency through the creation of class structure in the electorate that votes in favour of inefficient institutions.

The parameter τ captures the strength of enforcement of property rights. A high τ implies that law enforcement is poor and assets are likely to be stolen by thieves or taken over by the local strongman. Hence a straightforward way to think about τ is how tough government is on property related crime and how well it enforces the claims of someone dispossessed of their property. Alternatively, τ can also be thought of as how well the titling system works. To the extent it is easy to bribe the local bureaucrat to get the name on someone's land title changed, τ would be high and vice versa.

The parameter ϕ measures the efficiency of contractual institutions. The treatment of ϕ is somewhat different since it is the proportion of collateralized wealth that can be liquidated. If an agent pledges wealth a as collateral to become an entrepreneur, and his project fails, the bank only recovers ϕa. Hence $(1-\phi)a$ is pure inefficiency and consequently there is a strong case for thinking that $\phi = 1$ will be the surplus maximising policy. However under certain

[8]It should be noted that in contrast to Ghatak et al. (2007) there are no multiple equilibria since firm level labour demand is constant at n. This implies that in our model the what drives the labour demand is the extensive margin effect.

conditions, this effect may be dominated through the inefficiencies caused in the occupational choices since a high ϕ can end up making entrepreneurship attractive to agents who should optimally become workers.

ϕ and τ are parameters that capture institutional frictions that reduce the efficiency of market transactions involving wealth.[9] This can be illustrated with the following example. To fix ideas let us think of wealth as land. Consider a scenario where there's an agent who wishes to rent out his land. This landlord would consider two things when entering into a rental contract with a potential tenant. Firstly he would consider how secure his property rights are. When τ is high, the landlord realises that his property rights over the land he is renting out are not very secure. This dampens the incentives for renting the land since the landlord worries about a potential capture by the tenant. Independently, a low ϕ implies that enforcement of contracts is costly. The landlord anticipates that in the event a tenant refuses to vacate the land as per the terms of the rental contract, the landlord would need to approach the courts for enforcement of his contractual rights. Even if property rights are fully secure, if ϕ is low, the court costs would be substantial. Therefore a low ϕ would also dampen the incentives to put land to its productive use.

The distinction between the two institutions is heuristic.[10] In most applications one can think of, ϕ and τ would interact together creating aggregate transaction costs that would dampen the incentives for market transactions involving wealth. For example in the model presented here, both enter multiplicatively when agents post their wealth as collateral to become entrepreneurs. The credit market takes into account both the insecurity of the property right over the collateral and the costs of enforcing the credit contract in case of default.

3.5.1 Institutions in the First Best World

We now show that in the first best world the surplus maximising institutions are chosen.

Proposition 3.4. *When talent is observable, voters unanimously choose surplus maximising institutions.*

[9]We have focused only on institutional frictions involving wealth because wealth is the instrument that banks can use to mitigate the inefficiencies due to the unobservability of talent, and we want to show that the political process can fail to choose the right reforms even when there is no redistributive objective.

[10]In Besley (1995) three channels through which property rights affects investment incentives are laid out. These are the security of tenure, the use of property as collateral, and the benefits of gains from trade. Of these we feel that the first and the third are channels through which τ would affect investment incentives whereas the second channel relating to the use of land as collateral is affected by an interaction of τ and ϕ as is the case in the model. Of course wealth in our model is exogenous and therefore the issue of investment incentives does not arise.

Proof. Total surplus in the economy is maximized when the most talented agents become entrepreneurs regardless of their wealth. This is equivalent to the quality of the pool of entrepreneurs being maximised. Under the first best the total surplus in the economy is:

$$W_{fb} = R\int_{\theta^*}^{1} \theta f(\theta)d\theta + \frac{M}{n+1} + \int_{0}^{\infty}(1-\tau)ag(a)d(a) \tag{3.26}$$

By inspecting this expression it is clear that the total surplus is decreasing in τ. Hence $\tau = 0$ is the surplus maximising. Since all agents lose a part of their wealth as τ increases, agents unanimously vote for τ equal to zero. Since ϕ does not appear in (3.26), all values of ϕ are surplus maximising, and hence the proposition is trivially true for ϕ. □

When talent is observable, the preferences of the electorate are unanimously aligned with surplus maximisation. Hence a $\tau = 0$ is chosen because better property rights increase the expected payoff of all agents. Similarly the optimal ϕ would be chosen to the extent there are any contractual transactions involving wealth. Note that in the first best in our model there are no contractual transactions involving wealth since talent is observable and wealth has no use as a screen. Hence all values of ϕ are optimal in the first best world.

3.5.2 Institutions in the Second Best World

In the last subsection we showed that in the first best world the preferences of the electorate are unanimously aligned with surplus maximisation. We will show that as soon as there's a departure from the first best, the inefficiency of the market gets further amplified by the choices of the electorate that is created in the inefficient market. In the second best world with unobservable talent, the total surplus is:

$$W_{sb} = R\left(\int_{\underline{a}_p}^{\hat{a}_p}\int_{\hat{\theta}(a)}^{1} \theta f(\theta)g(a)d\theta da + \int_{\hat{a}_p}^{\infty}\int_{\underline{\theta}_s}^{1} \theta f(\theta)g(a)d\theta da\right) \tag{3.27}$$

$$+M\left(\int_{\underline{a}_p}^{\hat{a}_p}(1-F(\hat{\theta}(a)))g(a)da + (1-G(\hat{a}_p))(1-F(\underline{\theta}_s))\right)$$

$$+\underline{w}\left(1-(n+1)\int_{\underline{a}_p}^{\hat{a}_p}(1-F(\hat{\theta}(a)))g(a)da + (1-F(\underline{\theta}_s))(1-G(\hat{a}_p))\right) + \int_{0}^{\infty}(1-\tau)ag(a)da$$

$$-(1-\tau)(1-\phi)\left(\int_{\underline{a}_p}^{\hat{a}_p}\int_{\hat{\theta}(a)}^{1}(\theta(1-\gamma(a))+(1-\theta))\,af(\theta)g(a)d\theta da+\int_{\hat{a}_p}^{\infty}\int_{\underline{\theta}_s}^{1}a(1-\theta)f(\theta)g(a)d\theta da\right)$$

In this economy there are two productive activities, the subsistence sector where a worker produces $\underline{w}$, and the hi tech sector where n workers and 1 entrepreneur of ability θ produce R with probability θ and 0 with probability $(1-\theta)$. The project also yields a non expropriable return M to the entrepreneur. The wage paid to the worker in the hi tech sector is simply a transfer from the entrepreneur to the worker which doesn't enter the total surplus. In the world with full information, the first best is guaranteed, where all agents are engaged in the hi tech sector either as a worker or entrepreneurs. This is what equation (3.26) captures.

In the second best world it is possible that there are agents that work in the subsistence sector. The mass of agents engaged in the hi tech sector is $n+1$ times the mass of entrepreneurs. The rest of the agents work in the subsistence sector where they produce $\underline{w}$. This is captured in the third part of equation (3.27) which takes a positive value when $w = \underline{w}$ and 0 otherwise.

The fourth part of the expression captures the loss of wealth when τ is greater than 0. Similarly when ϕ is less than one there is some loss of collateral in case of default. The first best could be achieved if $\underline{a}_p = 0$ and $\hat{\theta}(a) = \underline{\theta}_s = \theta^*$. In such a case none of the agents in the economy are engaged in the subsistence sector and hence the second term in the expression drops out.

It is easy to see why the first best is never possible when talent is unobservable. Even when there is no credit constraint, at low enough levels of wealth, separation is not possible. At the bottom of the wealth distribution where $a = 0$, the credit market can only offer a pooling contract. With a pooling contract at $a = 0$, the talent of the least talented agent that chooses entrepreneurship is always lower than θ^* since θ^* is the talent of the least talented agent that accepts her actuarially fair contract in the full information case. Since the least talented agent receives a cross subsidy with the pooling contract but not a separating contract, the talent of the marginal agent with 0 wealth is lower when talent is unobservable. But since the mass of entrepreneurs is bounded at $\frac{1}{n+1}$, and at the lower end of the wealth distribution agents with talent less than θ^* are entrepreneurs, then at wealth $a \geq \hat{a}_p$, $\underline{\theta}_s$ must be greater than θ^*. That is, agents that would become entrepreneurs in the first best world, choose to work for a wage. This drives the inefficiency in the model. If credit constraint exists then there is the added inefficiency of agents with high talent but low wealth that are excluded from entrepreneurship.

The first best can only be replicated in the world with incomplete information if all agents have sufficient wealth and can be offered a separating contract.

Therefore if the average wealth in this economy is greater than the threshold level of wealth required for separation, a policy of redistribution can restore full efficiency in this economy. If the total level of wealth is insufficient or if the instruments for conducting such a redistribution are unavailable then there will always be some inefficiency since there would at the same time be agents with talent less than θ^* who choose entrepreneurship and talent greater than θ^* that choose working for a wage.

Observation 3.1. *A non-zero level of credit constraint may be optimal in this economy.*

Given this discussion, it is possible to envisage distributions of wealth and talent such that there exists a non zero "natural level of credit constraint". That is, the total surplus may not always be maximised when the credit constraint is pushed down. Though reducing the credit constraint allows agents with low wealth to become entrepreneurs, this has an effect through the labour market of increasing the wage. Increasing the wage may in turn reduce the number of high type entrepreneurs with high wealth.

To discuss whether endogenous institutions can bring the economy in the direction of higher welfare or not, suppose that all agents can vote in a binary election between a status quo institution (status quo ϕ or τ) and an alternative. When faced with a binary choice, each agent votes sincerely.

One obvious remark we will make, without making distributional assumptions, is that an alternative policy that is aimed at maximising total surplus may not win when put to majority vote. This result in itself is not particularly surprising. Since redistributive instruments are lacking it is to be expected that agents inefficiently use institutions to redistribute rather than to maximise surplus. Indeed such a choice of institutions is not inefficient in the paretian sense. What is interesting here however is that the alignment of interest groups is itself created by the existence of market failure and this alignment takes the economy away even from the second best world with market failures. In other words, the inefficiency of market failure is further amplified by the political alignments it creates.

The cornerstone to understanding why agents choose non surplus maximising institutions is the following: in this economy there are always at least $\frac{n}{n+1}$ workers. Since $n \geq 1$, a policy that increases wage has support of at least half the population. However policies that increase the wage may not increase the quality of the pool of entrepreneurs. This is the insight that we will use to generate the results in the rest of this section. Thus efficient institutions are those that increase the quality of the pool of entrepreneurs whereas institutions that increase wage are politically feasible.

Support for improvement in judicial enforcement

The parameter ϕ in the model denotes the amount of collateral that banks can liquidate in case of default and is the parameter that denotes the quality of the judiciary. Instead of a cost that is proportional to the collateral in dispute, the quality of the judiciary could be modelled as a fixed cost that need to be paid for approaching the judiciary. In such a model ϕ would be a fixed cost and interest rate would instead be determined by the following zero profit condition:

$$r(a)nw\theta + ((1-\tau)a - \phi)(1-\theta) = nw \tag{3.28}$$

The idea we wish to capture with ϕ is the efficiency of the judiciary in expropriating assets of a defaultor and handing them over to the creditor at the least possible cost. This idea is captured in both these formulations. Given the discussion on efficiency and political feasibility, we have:

Proposition 3.5. *A policy aimed at increasing ϕ is guaranteed majority support but may not always be surplus maximising.*

Proof. There are two parts to this proposition. The first part is that a policy of increasing ϕ is guaranteed majority support. This is proven in the appendix. The second part is that such a policy is not guaranteed to be surplus maximising. This is proven by construction of an example in the final extension where increasing ϕ reduces total surplus. □

The intuition for the result is the following. It is easy to show that the equilibrium wage is non decreasing in ϕ, and hence the proposal for increasing ϕ is supported by the majority. However, total surplus may not be increasing in ϕ since the effect of an increase in ϕ on the quality of the pool of entrepreneurs is ambiguous.

This result is quite striking when contrasted against the standard intuition about contracting institutions. Here improving the quality of contracting institutions (increasing ϕ) is not always good since that makes entrepreneurship more attractive and this induces low types to become entrepreneurs. This result arises because there are inherent externalities when agents borrow money: the low type entrepreneurs by their very existence impose an externality on the high types. Our result can be easily understood when seen in the light of the theory of second best.

Support for Improvement in Property Rights

Imperfect protection of property rights reduces the value of wealth. This in turn makes entrepreneurship more attractive since agents do not place as much weight on default and consequent loss of collateral.

The political support for a change in τ is ambiguous because the effect on the wage is ambiguous. We can see this from the following:

$$\frac{\partial L_d}{\partial \tau} : (n+1)\left(-g(\underline{a}_p)\frac{\partial \underline{a}_p}{\partial \tau}(1 - F(\underline{\theta}_p)) - f(\underline{\theta}_s)\frac{\partial \underline{\theta}_s}{\partial \tau}(1 - G(\hat{a}_p)) - \int_{\underline{a}_p}^{\hat{a}_p} f(\hat{\theta}(a))\frac{\partial \hat{\theta}(a)}{\partial \tau}g(a)da\right) \tag{3.29}$$

The sign of this expression is ambiguous. This is because:

$$\frac{\partial \underline{a}_p}{\partial \tau} > 0 \qquad \frac{\partial \underline{\theta}_s}{\partial \tau} = 0 \qquad \frac{\partial \hat{\theta}(a)}{\partial \tau} < 0 \tag{3.30}$$

$$\frac{\partial \underline{a}_p}{\partial \tau} = \frac{nw - \theta_p(\underline{a}_p)R}{\phi(1-\tau)^2} - \frac{\partial \theta_p(\underline{a}_p)}{\partial \tau}\left(\frac{R}{\phi(1-\tau)}\right) > 0 \tag{3.31}$$

since

$$\frac{\partial \underline{\theta}_p}{\partial \tau}|_{\underline{a}_p} = 0 \tag{3.32}$$

The credit constraint is increasing in τ. When τ increases, the effective wealth of an agent decreases, and the interest rate at all levels of wealth increases. This is intuitive since an increase in τ decreases the value of wealth as a screen. Since agents are likely to have their wealth expropriated anyway, posting a high collateral is less effective in revealing an agent's type. Take the limiting case where τ goes close to 1, in this case, the credit market correctly anticipates that all agents are equally eager to post any collateral since they know that their wealth will be expropriated and hence don't attach any value on recovery of collateral in the event of success and consequent repayment of the loan.

There are two opposing effects on wage of a decrease in τ. Firstly decreasing τ reduces the level of credit constraint. This increases the number of entrepreneurs. Decreasing τ also decreases the attractiveness of entrepreneurship for marginal agents ($\hat{\theta}(a)$), who were previously accepting the pooling contract to become entrepreneurs due to the cross subsidy from higher types within their wealth level. Since there are two opposite effects on wage, the precise effect on total surplus of a change in τ would depend on the assumptions on the distribution of wealth and talent. However in case these two effects exactly cancel each other out, it is possible then to characterise the effect on total surplus.

Proposition 3.6. *If the wage remains unchanged as a result of a change in τ, then decreasing (increasing) τ increases (decreases) total surplus*

Proof. If wage remains unchanged as a result of a decrease in τ then the new equilibrium pareto dominates the previous equilibrium. All agents who remain workers are unaffected, all entrepreneurs are made better off due to a reduction in the interest rate. Additionally there are agents who were previously credit constrained who can now become entrepreneurs for whom the policy is a strict improvement over status quo. Since it is a pareto improvement, it must also increase total surplus. Similarly if an increase in τ keeps the wage unchanged, it must reduce the total surplus since workers are unaffected, entrepreneurs are made worse off due to the increase in the interest rate, and there are at least some agents who are denied credit as a result of the increase in the credit constraint who are made strictly worse off. □

By continuity we can extend this proposition to mean that if the change in wage as a result of an improvement in property rights is small enough, then total surplus must have increased. It is possible to push this result further.

Proposition 3.7. *If the change in wage as a result of improvement (deterioration) in property right is negative (positive) then total surplus must increase (decrease).*

Proof. Note first that the average quality of the pool of entrepreneurs is a sufficient statistic for gauging changes in total surplus. If the wage decreases as a result of an decrease in τ, it must be the case that the effect on labour demand through $\hat{\theta}(a)$ dominates the reduction in the credit constraint. Now note that the average talent at the lowest level of wealth where a pooling contract is offered is lower than the average talent of the pool. This is true because the distribution of wealth and talent are independent and the talent of the least talented agent within a wealth level is increasing in wealth.

Now note that is always possible to construct a distribution of wealth such that the pre reform average talent is the same but post reform the credit constraint is relaxed more to the extent that the two opposing effects on wage cancel each other out and wage remains unchanged. In this case, the average talent post reform would be lower than the case where the wage went down. However, given the previous result, the total surplus would still increase. Since the initial average quality of the pool of entrepreneurs is the same by construction, this implies that the ex post level of talent must have increased in the case where the wage decreases. □

This result brings into sharp relief the trade-off between political feasibility and efficiency of institutional reform. Only reforms that increase wages are politically feasible but these may not correspond to reforms that are surplus maximising. In case of property rights institutions, when worsening them (increasing τ) is politically feasible, they have an unambiguously negative effect on the total surplus. The political feasibility of τ depends on the distribution of wealth. if the median voter is a worker with very low wealth she would care more about the effect on wage and would therefore vote in favour of worsening property rights.

3.6 Conclusion

To summarise our result on institutional efficiency and feasibility, we find that improving contractual institutions is always feasible but may not always be efficient since improving contracting induces too many low type agents to choose entrepreneurship. On the other hand we find that if worsening property rights institutions is politically feasible then it unambiguously reduces total surplus. Similarly if improving property rights is politically infeasible then it unambiguously increases total surplus. These results bring into sharp focus the tension between political feasibility and surplus maximisation.

When there's a market failure, the competitive equilibrium is no longer guaranteed to be on the Pareto frontier. Our model makes the point that in the event of a market failure, competitive markets can passively play a political role of creating constituencies. These constituencies can have a preference for inefficient policies. This leads to the inefficiencies of market failure being further amplified by the policy choices that constituencies created in a flawed market make. In this sense this chapter provides an additional reason to worry about market failure; market failure may lead to a political failure even in a fully representative democracy.

Appendix

Proof for Lemma 3.1

Proof. If $R < r_s nw$ then there are insufficient appropriable returns to cover the interest payments from the loan. Hence the only way the banks can break even is if entrepreneurs pledge a portion of their collateral even in the state where the project is successful. Let us call the proportion of collateral that banks seize in the good state $(1 - \gamma(a))$. The new zero profit condition for banks when they lend to an agent whose declared type is $\tilde{\theta}$ is

$$\tilde{\theta}(R + (1 - \gamma(a))(1 - \tau)\phi a) + (1 - \tilde{\theta})(1 - \tau)\phi a = nw$$

Rearranging this, we get

$$\gamma(a) = \frac{\tilde{\theta}R - nw + (1 - \tau)\phi a}{\tilde{\theta}(1 - \tau)\phi a}. \tag{3.33}$$

The left hand side of the occupational choice constraint for an agent with talent θ who accepts a separating contract (designed for an agent with talent $\tilde{\theta}$) is

$$M - (1 - \theta\gamma(a))(1 - \tau)a$$

Substituting the value for $\gamma(a)$ from equation (3.33) we get:

$$v_\theta(\tilde{\theta}) = M + \theta\left(\frac{\tilde{\theta}R - nw + (1-\tau)\phi a}{\tilde{\theta}\phi}\right) - (1-\tau)a.$$

We can now differentiate this equation with respect to the declaration $\tilde{\theta}$ to see whether the agent has an incentive to declare his type truthfully. It is easy to check that in the relevant range, the payoff of the agent $v_\theta(\tilde{\theta})$ is increasing in his declaration $\tilde{\theta}$. Hence agents will always overstate their type and a separating contract cannot exist. □

Derivation of $a(\tilde{\theta})$ from section 3.3.2.

An agent of type θ maximises his payoff from entrepreneurship by choosing the declaration that maximises $v_\theta(\tilde{\theta})$.

$$\operatorname*{argmax}_{\tilde{\theta}} v_\theta(\tilde{\theta}) = \theta. \tag{3.34}$$

The first order condition for this problem evaluated at $\tilde{\theta} = \theta$ is:

$$\frac{nw}{\tilde{\theta}(1-\tilde{\theta})(1-\tau)(1-\phi)} - \frac{a(\tilde{\theta})\phi}{\tilde{\theta}(1-\tilde{\theta})(1-\phi)} - a'(\tilde{\theta}) = 0. \tag{3.35}$$

Let $\frac{nw}{\tilde{\theta}(1-\tilde{\theta})(1-\tau)(1-\phi)}$ be $Q(\tilde{\theta})$ and $\frac{\phi}{\tilde{\theta}(1-\tilde{\theta})(1-\phi)}$ be $P(\tilde{\theta})$. This is a differential equation of the following form:

$$a'(\tilde{\theta}) + P(\tilde{\theta})a(\tilde{\theta}) = Q(\tilde{\theta})$$

which is characterised by the solution

$$e^{\int P(\tilde{\theta})d\tilde{\theta}}a(\tilde{\theta}) = \int e^{\int P(\tilde{\theta})d\tilde{\theta}}Q(\tilde{\theta})d\tilde{\theta} + C$$

Solving this for $a(\tilde{\theta})$ we find that:

$$a(\tilde{\theta}) = \frac{nw}{\phi(1-\tau)}\left(1 + C\left(\frac{1-\tilde{\theta}}{\tilde{\theta}}\right)^{\frac{\phi}{1-\phi}}\right),$$

where C is the constant of integration. Since lower bound values of $\tilde{\theta} = \underline{\theta}_s$ and $a(\tilde{\theta}) = \underline{a}_s$ we can solve for the particular solution. This is

$$a(\tilde{\theta}) = \frac{nw}{\phi(1-\tau)}\left(1 - \frac{(R-nw)}{nw}\left(\frac{\underline{\theta}_s}{1-\underline{\theta}_s}\right)^{\frac{1}{1-\phi}}\left(\frac{1-\tilde{\theta}}{\tilde{\theta}}\right)^{\frac{\phi}{1-\phi}}\right).$$

It is possible to check that the second order condition is satisfied. The second order condition for this problem is:

$$-\frac{1}{\theta^2}(nw - \phi(1-\tau)a(\theta)) - \frac{2\phi(1-\tau)}{\theta}a'(\theta) - a''(\theta)(1-\tau)(1-\theta)(1-\phi) < 0$$

By inspection it is possible to check that this equation always holds, and hence the function is globally concave.

Proof of Lemma 3.2

Proof. Let us first consider the region where $R \geq r_p(a)nw$ is satisfied. In this region $\hat{\theta}(a)$. Totally differentiating this equation, and rearranging we find

$$\frac{d\hat{\theta}(a)}{da}\left(R - r(a)nw + (1-\tau)a - \hat{\theta}(a)\frac{dr(a)}{d\hat{\theta}}\right) = (1-\tau)\left(1 - \frac{\hat{\theta}(a)}{\theta_p(a)}(\theta_p(a) + (1-\theta_p(a))\phi)\right). \tag{3.36}$$

Note that

$$R - r(a)nw \geq 0, \qquad \frac{dr(a)}{d\hat{\theta}} < 0, \qquad \hat{\theta}(a) < \theta_p(a). \tag{3.37}$$

Hence

$$\frac{d\hat{\theta}(a)}{da} > 0. \tag{3.38}$$

□

Proof of Proposition 3.5

Proof. For proving political support it is sufficient to show that wage is non decreasing in ϕ.

$$\frac{\partial L_d}{\partial \phi} : (n+1)\left(-g(\underline{a}_p)\frac{\partial \underline{a}_p}{\partial \phi}(1 - F(\underline{\theta}_p)) - f(\underline{\theta}_s)\frac{\partial \underline{\theta}_s}{\partial \phi}(1 - G(\hat{a}_p)) - \int_{\underline{a}_p}^{\hat{a}_p} f(\hat{\theta}(a))\frac{\partial \hat{\theta}(a)}{\partial \phi}g(a)da\right) \geq 0 \tag{3.39}$$

This can be demonstrated by showing that

$$\frac{\partial \underline{a}_p}{\partial \phi} < 0, \qquad \frac{\partial \underline{\theta}_s}{\partial \phi} < 0, \qquad \text{and} \qquad \frac{\partial \hat{\theta}(a)}{\partial \phi} < 0. \tag{3.40}$$

$$\frac{\partial \underline{a}_p}{\partial \phi} = -\frac{nw - \theta_p(\underline{a}_p)R}{\phi^2(1-\tau)} - \frac{\partial \theta_p(\underline{a}_p)}{\partial \phi}\frac{R}{(1-\tau)\phi} < 0 \tag{3.41}$$

since

$$\frac{\partial \theta_p(\underline{a}_p)}{\partial \phi} = 0. \tag{3.42}$$

Also

$$\frac{\partial \underline{\theta}_s}{\partial \phi} = -\frac{M-w}{R} < 0. \tag{3.43}$$

And lastly,

$$\frac{\partial \hat{\theta}(a)}{\partial \phi} < 0 \tag{3.44}$$

by inspection.

This implies that the equilibrium wage is non decreasing in ϕ. This ensures that the policy enjoys majority support. The inequalities are strict when $w > \underline{w}$. If there is a subsistence sector, then there is no effect of ϕ on the labour demand. The effect of increasing ϕ on total surplus is ambiguous. □

We now construct an example where the credit constraint effect dominates, that is, the surplus maximising ϕ in the economy is less than one. Assume that the distribution of wealth is discrete. There are three classes in the population: the rich, the middle, and the poor of size p_r, p_m, p_p with wealth a_r, a_m, a_p respectively. Assume:

$$q(p_r + p_m) < \frac{1}{n+1}$$

It turns out that if there is a subsistence sector in the economy then it is always surplus enhancing to locally increase ϕ. Hence to make the problem interesting assume that there is no subsistence sector in the economy. The two feasible values for ϕ will be $\{\underline{\phi}, 1\}$. In this economy the credit constraint will be higher with $\phi = 1$.

The change in total surplus as a result of increasing $\underline{\phi}$ to 1 is:

$$\Delta TS = TS(\underline{\phi}) - TS(1) = qp_m(1-\theta)R - (1-\theta)(1-q)(1-\underline{\phi})(p_m a_m \lambda_m(\underline{\phi}) + p_r a_r \lambda_r(\underline{\phi})) \tag{3.45}$$

The first term in the expression represents the increase in the total surplus due to replacement of some low type entrepreneurs by high types as a result of access to credit due to reduction in the credit constraint. The second term represents the reduction in the surplus due to destruction of a proportion of assets in case of default due to imperfect judiciary. In the second term $\lambda_i(\underline{\phi})$ is the proportion of low type entrepreneurs with wealth i that choose entrepreneurship in equilibrium.

Lemma 3.5. *If credit constraint worsens as a result of an increase in ϕ from $\underline{\phi}$ to 1, then $\lambda_m(\underline{\phi}) = 1$*

Proof. Assume this is not true. Then there are two possibilities: either $\lambda_m(\underline{\phi}) = 0$ or $0 < \lambda_m(\underline{\phi}) < 1$. Consider $\lambda_m(\underline{\phi}) = 0$. This implies that $\lambda_r(\underline{\phi}) = 0$ since $\frac{\partial \lambda}{\partial a} < 0$. This implies that all entrepreneurs are high types. This contradicts Assumption 1. If $0 < \lambda_m(\underline{\phi}) < 1$ then the interest rate for agents with wealth a_m is:

$$r_s(a_m, (\underline{\phi}) = \frac{\theta R + M - w(\underline{\phi}) - (1-\theta)a_m}{nw(\underline{\phi})\theta}$$

Substituting this into the equation that determines the credit constraint:

$$R - r_s(a_m, \underline{\phi})nw(\underline{\phi}) \geq 0$$

it is easy to check that the credit constrain is decreasing in equilibrium wage. Since the equilibrium wage is monotonically increasing in ϕ and hence the credit constraint with $\phi = 1$ must be lower than the credit constraint with $\underline{\phi}$ but this is a contradiction. □

Hence the change in total surplus simplifies to:

$$\Delta TS = qp_m(1-\theta)R - (1-\theta)(1-q)(1-\underline{\phi})(p_m a_m + p_r a_r \lambda_r(\underline{\phi})) \tag{3.46}$$

Now we can back out $\lambda_r(\underline{\phi})$ since we know that the proportion of entrepreneurs in the economy is $\frac{1}{n+1}$.

$$\lambda_r(\underline{\phi}) = \left(\frac{1}{(n+1)p_r} - \frac{p_m}{p_r} - q\right)\frac{1}{1-q}$$

Substituting this into the expression for the change in total surplus, we find that $\Delta TS > 0$ if:

$$R > \frac{1-\underline{\phi}}{q}\left((1-q)a_m + \left(\frac{1}{(n+1)p_m} - 1 - q\frac{p_r}{p_m}\right)a_r\right) \tag{3.47}$$

This equation ensures that if the credit constraint worsens as a result of an increase in ϕ, the loss of efficiency through reduction in the quality of the pool of entrepreneurs dominates the loss of collateral during recovery with a lower ϕ. The credit constraint worsens if:

$$R - nw(\underline{\phi})r_p(a_m, \underline{\phi}) > 0 > R - nw(1)r_p(a_m, 1)$$

Solving the model to derive the equilibrium wage rate, and interest rate at wealth level a_m for both values of ϕ we find:

$$n\frac{\theta(\theta R + M) + (n+1)p_r q(1-\theta)(\theta R + M - a_r)}{(n+1)(\theta + (1-\theta)p_r q)} - (1-\theta)(1-q)a_m >$$

$$(q + (1-q)\theta)R >$$

$$n\frac{\theta(\theta R + M - (1-\underline{\phi})a_r)(1 - p_m(n+1)) + p_r q(n+1)((1-\theta)(\theta R + M) - (1-\theta(1-\underline{\phi}))a_r)}{(n+1)(\theta(1-p_m(n+1)) + (1-\theta)qp_r)}$$

$$-(1-\theta)(1-q)\underline{\phi}a_m \tag{3.48}$$

Proposition 3.8. *For any constellation of parameter values for which equations* (3.48) *and* (3.47) *are satisfied,* $\phi = 1$ *is suboptimal.*

Proof. Equation (3.48) implies that the credit constraint worsens as a result of an increase in ϕ from $\underline{\phi}$ to 1. This implies that there are fewer high type entrepreneurs with $\phi = 1$. Equation (3.47) ensures that assuming the credit constraint worsens, the change in total surplus is negative for an increase in ϕ from $\underline{\phi}$ to 1. Taken together they imply that the credit constraint worsens, and enough high type entrepreneurs are credit constrained such that total surplus is diminished. □

Bibliography

Acemoglu, Daron, "The Form of Property Rights: Oligarchic vs. Democratic Societies," *MIT Department of Economics Working Paper*, September 2003, *3* (34).

Anderlini, Luca and Leonardo Felli, "Costly Bargaining and Renegotiation," *Econometrica*, 2001, *69* (2).

_ , _ , and Andrew Postlewaite, "Active Courts and Menu Contracts," *CESifo Working Paper*, 2006, (1852).

_ , _ , and _ , "Should Courts Always Enforce What Contracting Parties Write?," *CESifo Working Paper*, 2006, (1847).

Ausubel, Lawrence M., Peter Cramtom, and Raymond J Deneckere, "Bargaining with Incomplete Information," *Handbook of Game Theory*, 2002.

Banerjee, Abhijit, "Contracting Constraints, Credit Markets and Economic Development," *MIT Department of Economics Working Paper*, September 2001, (02-17).

_ and Andrew Newman, "Occupational Choice and the Process of Development," *Journal of Political Economy*, 1993, *101.*

Bardhan, Pranab, *The Economic Theory of Agrarian Insititutions*, Oxford University Press, 1989.

Bebchuk, Lucian Arye, "Litigation and settlement under imperfect information," *Rand Journal of Economics*, 1984, *15* (3).

Besley, Tim, *Principled Agents? The Political Economy of Good Government*, Oxford University Press, 2006.

Besley, Timothy, "Property Rights and Investment Incentives: Theory and Evidence from Ghana," *Journal of Political Economy*, 1995, *103* (5).

Biais, Bruno and Thomas Mariotti, "Credit, Wages and Bankruptcy Laws," *C.E.P.R. Discussion Papers*, 2003, *3996.*

Boyer, Marcel and Jean-Jacques Laffont, "Toward a Political Theory of the Emergence of Environmental Incentive Regulation," *Rand Journal of Economics*, 1999, *30* (1).

Caselli, Francesco and Nicola Gennaioli, "Economics and Politics of Alternative Institutional Reforms," *Quarterly Journal of Economics*, 2008, *123* (3).

Chatterjee, Kalyan and William Samuelson, "Bargaining under Incomplete Information," *Operations Research*, 1983, *31.*

Clark, Derek J. and Christian Riis, "Contest Success Functions: An Extension," *Economic Theory*, 1998, *11.*

Cooter, Robert D and Daniel L Rubinfeld, "Economic Analysis of Legal Disputes and their Resolution," *Journal of Economic Literature*, 1989, *27.*

Daughety, Andrew and Jennifer F Reinganum, "Settlement Negotiations with Two-Sided Asymmetric Information: Model Duality, Information Distribution, and Efficiency," *International Review of Law and Economics*, 1994, *14* (3).

Djankov, Simeon, Rafael La Porta, Florencio Lopez-De-Silanes, and Andrei Shleifer, "The Regulation of Entry," *Quarterly Journal of Economics*, February 2002, *117* (1).

Fearon, James D., "Rationalist explanations for war," *International Organisation*, 1995, *49* (3).

Friedman, Daniel and Donald Wittman, "Litigation with Symmetric Bargaining and Two-Sided Incomplete Information," *Journal of Law Economics & Organization*, 2006, *23* (1).

Galor, Oded and Joseph Zeira, "Income Distribution and Macroeconomics," *Review of Economic Studies*, 1993, *60.*

Ghatak, Maitreesh and Timothy Guinnane, "Economics of Lending with Joint Liability: Theory and Practice," *Journal of Development Economics*, 1999, *60* (1).

_ , Massimo Morelli, and Tomas Sjostrom, "Entrepreneurial Talent and Occupational Choice," *Journal of Economic Theory*, 2007.

Gould, J, "The Economics of Legal Conlicts," *Journal of Legal Studies*, 1973, *2.*

Grossman, Sanford, "The Informational Role of Warranties and Private Disclosure About Private Quality," *Journal of Law and Economics*, 1981, *24*.

Groves, T, "Incentives in Teams," *Econometrica*, 1973, *41*.

Hay, Bruce L., "Effort, Information, Settlement, Trial," *Journal of Legal Studies*, 1995, *24*.

_ **and Kathryn E. Spier**, "Litigation and Settlement," in Peter Newman, ed., *New Palgrave Dictionary of Economics and Law*, Palgrave MacMillan, 1998.

Hirshleifer, Jack and Evan Osborne, "Truth, Effort, and the Legal Battle," *Public Choice*, 2001, *108*.

Hua, Xinyu and Kathryn E. Spier, "Information and Externalities in Sequential Litigation," *Journal of Institutional and Theoretical Economics*, 2005, *161* (2).

Landes, W, "An Economic Analysis of Courts," *Journal of Law and Economics*, 1971, *14*.

Lanjouw, Jean Olson and Mark Schankerman, "Protecting Intellectual Property Rights: Are Small Firms Handicapped?," *Journal of Law and Economics*, 2004.

Lerner, Joshua, "The Importance of Patent Scope: An Empirical Analysis," *Rand Journal of Economics*, 1994, *25*.

Levy, Gilat, "Careerist Judges and the Appeals Process," *Rand Journal of Economics*, 2005, *36* (2).

Myerson, Roger, "Optimal Coordination Mechanisms in Generalised Principal-Agent Problems," *Journal of Mathematical Economics*, 1982, *10*, 67–81.

_ **and Mark A Satterthwaite**, "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, 1983, *28*.

Nalebuff, Barry, "Credible Pretrial Negotiation," *Rand Journal of Economics*, 1987, *18* (2).

Nish, Ian, *The Origins of the Russo-Japanese War*, London: Longman, 1985.

Okuno-Fujiwara, Masahiro, Andrew Postlewaite, and Kotaro Suzumura, "Strategic Information Revelation," *Review of Economic Studies*, 1990, *57* (1).

Perotti, Enrico C. and Paolo F. Volpin, "Lobbying on Entry," *CEPR Discussion Paper*, August 2004, *4519.*

P'ng, Ivan P. L., "Strategic Behavior in Suits, Settlement, and Trial," *The Bell Journal of Economics*, 1983, *14* (2).

Posner, R, "An Economic Approach to Legal Procedure and Judicial Administration," *Journal of Legal Studies*, 1973, *2.*

Priest, George L. and Benjamin Klien, "The Selection of Disputes for Litigation," *Journal of Legal Studies*, 1984, *13.*

Rajan, Raghuram and Luigi Zingales, "The Persistence of Underdevelopment: Institutions, Human Capital, or Constituencies?," *NBER Working Paper*, March 2006, *12093.*

Robson, Alex and Stergios Skaperdas, "Costly Enforcement of Property Rights and the Coase Theorem," *Economic Theory*, 2008, *36.*

Rubin, Edward L., "Toward a General Theory of Waiver," *UCLA Law Review*, 1980-1981, *28.*

Schweizer, Urs, "Litigation and Settlement under Two-sided Incomplete Information," *Review of Economic Studies*, 1989, *56.*

Shavell, Steven, "Sharing of Information prior to Settlement or Litigation," *Rand Journal of Economics*, 1989, *20* (2).

_ , "Fundamental Divergence Between the Private and the Social Motive to Use the Legal System," *Journal of Legal Studies*, June 1997, *26.*

Skaperdas, Stergios, "Contest Success Functions," *Economic Theory*, 1996, *7.*

_ , "Bargaining versus Fighting," *Defence and Peace Economics*, 2006, *17* (6).

Spier, Kathryn E., "The Dynamics of Pretrial Negotiation," *Review of Economic Studies*, 1992, *59.*

_ , "Pretrial Bargaining and the Design of Fee-Shifting Rules," *Rand Journal of Economics*, 1994, *25* (2).

Weitzman, Martin, "The New Soviet Incentive Model," *The Bell Journal of Economics*, 1976, *7.*

White, J.A., *The Diplomacy of the Russo-Japanese War*, Princeton University Press, 1964.

Yale Law, Journal, "Fairness, Flexibility, and the Waiver of Remedial Rights by Contract," *The Yale Law Journal*, 1978, *87* (5).

Yildiz, Mohamet, "Bargaining Without Common Prior - An Immediate Agreement Theorem," *Econometrica*, 2003, *71* (3).

_ , "Waiting to Persuade," *Quarterly Journal of Economics*, 2004, *119* (1).