Title
Essays on communication games with multiple informants and their applications to legal systems

Permalink
https://escholarship.org/uc/item/33c7r355

Author
Kim, Chulyoung

Publication Date
2011

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA, SAN DIEGO

Essays on Communication Games with Multiple Informants and Their Applications to Legal Systems

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Economics

by

Chulyoung Kim

Committee in charge:

Professor Joel Sobel, Chair
Professor Vincent Crawford
Professor Roger Gordon
Professor Samuel Kernell
Professor Branislav Slantchev

2011
The dissertation of Chulyoung Kim is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2011
DEDICATION

To Jesus Christ, My Savior
TABLE OF CONTENTS

Signature Page ........................................ iii
Dedication ........................................... iv
Table of Contents ..................................... v
List of Figures ........................................ vii
List of Tables ......................................... viii
Acknowledgements .................................... ix
Vita ...................................................... x
Abstract of the Dissertation ......................... xi

Chapter 1  Introduction ................................ 1

Chapter 2  The Value of Information in Legal Systems ............ 6
  2.1  Introduction ........................................ 8
  2.2  Model ............................................... 12
  2.3  Adversarial System ................................ 16
      2.3.1  Communication Stage ......................... 16
      2.3.2  Investigation Stage .......................... 18
      2.3.3  Adversarial Equilibrium .................... 21
  2.4  Inquisitorial System ............................. 27
  2.5  Comparison of Systems ........................... 29
  2.6  Concluding Remarks and Discussion ................. 34
  2.7  Appendix ......................................... 38
      2.7.1  Proof for Proposition 5 .................... 38
      2.7.2  Proof for Proposition 6 .................... 38

Chapter 3  Partisan Advocates .......................... 40
  3.1  Introduction ...................................... 42
  3.2  The Model ........................................ 44
      3.2.1  Dewatripont and Tirole’s Model ............ 44
      3.2.2  Information-based Rewards and Partisans .... 47
  3.3  Optimality of Partisan Advocacy .................. 49
      3.3.1  Inquisitorial System ....................... 49
      3.3.2  Adversarial System ......................... 52
      3.3.3  Comparison ................................ 53
  3.4  Partisan Advocacy under Private Information ........ 55
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure 2.1:</th>
<th>The adversarial equilibria</th>
<th>23</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.2:</td>
<td>Example for Inquisitorial System</td>
<td>32</td>
</tr>
<tr>
<td>Figure 2.3:</td>
<td>Example for Adversarial System</td>
<td>33</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 3.1: List of Possible Manipulation .......................... 56
ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my wonderful advisor, Joel Sobel for his continuous encouragement, guidance, and support during my Ph.D. studies. I truly appreciate and value his patience and commitment as advisor, and respect his passion and enthusiasm on the research. He was the only person at the Department of Economics whom I could always count on with confidence.

I thank my committee members, Vincent Crawford, Roger Gordon, Samuel Kernell and Branislav Slantchev for their time and effort and for making themselves available during their busy schedules. Thanks to Navin Kartik for his guidance and support in my early, important years at the Department of Economics.

I thank my dear wife, Hyuna, and our parents whose consistent prayers have been the ultimate driving force to help me become the person who I am today.

Lastly but definitely not the least, I thank Jesus Christ who strengthen me and love me without reservations. I give you all the honor and praise.
VITA

2001 B. A. in Economics,
Yonsei University

2003 M. A. in Economics,
Yonsei University

2011 Ph. D. in Economics,
University of California, San Diego
This dissertation contains three chapters of my research. Chapter 1 studies the properties of the two most commonly used legal institutions, the inquisitorial system and the adversarial system. In the former, the decision maker takes a decision based on her own acquired information, whereas under the latter the decision maker requires the advocates to present their acquired information prior to making a decision. When information is both verifiable and costly to collect, I show that the decision maker expects to make fewer decision making errors in the adversarial system than in the inquisitorial system. The main factor behind this result is that the advocate with the burden of proof values information more and, consequently, works harder to collect information under the adversarial system than the impartial
decision maker under the inquisitorial system. This larger effort exerted by the advocates leads to more informed decision making under the adversarial system in spite of the advocates’ incentives to distort information.

Chapter 2 studies the problem of an uninformed decision maker who acquires expert advice prior to making a decision. I show that it is less costly to hire partisan agents than impartial agents, especially under advocacy, and that the decision maker prefers partisan advocacy to other forms of institutions. I also extend the literature, originating with Dewatripont and Tirole (1999), to a setting with contracts that condition on information provided and not just the decision made.

Chapter 3 studies the robustness of fully revealing equilibria (FRE) in multidimension-multisender cheap talk games. A FRE is outcome-robust (strategy-robust) if there is an equilibrium whose outcome (strategy) is close to the FRE outcome (strategy) when the noise in senders’ observations is small. I show that there is no outcome-robust FRE in the model of Levy and Razin (2007), and discuss the connections between these new notions of robustness and the existing stability concepts studied in the literature.
Chapter 1

Introduction
In many instances, a person who should make a decision is different from one who possesses information relevant for the decision-making, and, naturally, inefficiency arises. This has motivated economists to investigate how direct communication between the two parties, the uninformed decision maker and the informed expert, may enhance efficiency. A seminal paper by Crawford and Sobel (1982) was the first to pose this question in an economic model and laid the ground for 30 years of research to come. To the extent that the preference of the decision maker differs from that of the expert, they show that loss of information in communication is inevitable and that the equilibrium structure takes a very intuitive form. While their paper focuses on the games where information is non-verifiable ("cheap talk games"), another important paper by Milgrom (1981) is designed to investigate the case where information is verifiable ("persuasion games"). Based on their basic models, until 1990s, the literature has studied mainly the situation of one uninformed decision maker communicating with one informed expert, with few exceptions.

In the 2000s, the literature has shown much interest in the situation where the uninformed decision maker interacts with multiple informed experts. In the cheap talk game literature, Battaglini (2002) showed that the existence of another informed expert guarantees construction of the so-called fully revealing equilibrium, where the decision maker can extract all the information that the experts possess. This result has brought about a few papers that ask whether this special equilibrium is a reasonable outcome. One of my essays, Chapter 4, contributes to this literature by organizing these arguments according to the new concepts of "reasonableness" and deriving a new result.

Although the cheap talk literature has evolved around rather theoretical questions, the persuasion game literature has seen burgeoning interests in applications, especially for legal disputes, following Dewatripont and Tirole (1999). This is because legal disputes naturally involve two competing parties (defendant and plaintiff) and also because a large punishment on perjury makes the model with verifiable information more suitable for such applications. Two of my essays, Chapters 2 and 3, contribute to this literature by adding some more realistic features to the existing models.

In Chapter 2, "The Value of Information in Legal Systems," I study the properties of the two most commonly used legal institutions, the inquisitorial system and the
adversarial system. In the former, the decision maker takes a decision based on her own acquired information, whereas under the latter the decision maker requires the advocates to present their acquired information prior to making a decision. The former system is widely used in continental European countries where the decision maker takes the initiative in collection and evaluation of relevant pieces of information. Some legal scholars such as Tullock (1975, 1980, 1988) favor this type of system on the ground of its impartiality toward the dispute resolution. On the other hand, the latter system is mostly used in America and U.K., where evidence collection is assumed by the interested parties, or the advocates, who drive the organization and development of the case. Posner (1988, 1999) argues that the competitive nature of the adversarial process gives the advocates greater incentives to search for evidence than in a system where the judge or investigator is the only searcher. He concludes that this benefits the fact-finding process of the adversarial system.

When information is both verifiable and costly to collect, I show that the decision maker expects to make fewer decision making errors in the adversarial system than in the inquisitorial system. The main factor behind this result is that the advocate with the burden of proof values information more and, consequently, works harder to collect information under the adversarial system than the impartial decision maker under the inquisitorial system. This larger effort exerted by the advocates leads to more informed decision making under the adversarial system in spite of the advocates’ incentives to distort information.

Chapter 3 contains the second essay, "Partisan Advocates," which is a companion piece to Chapter 2. In this essay, I study the same two legal institutions in a principal-agent model and show that the agency cost of generating information is lower in the adversarial system than in the inquisitorial system.

Although the objects of study are the same, I take a somewhat different approach in modeling these institutions in this second essay. In the inquisitorial system, instead of collecting information herself as in the first essay, the decision maker hires and pays for one single agent who collects information for both causes. In the adversarial system, the decision maker hires and pays for two distinct agents who collect information for each cause. So, the main difference between the two systems in this model is whether
to use two specialized forces, one for each cause, or just one investigative body. This approach to viewing these legal institutions is based on Dewatripont and Tirole (1999) who attempted to provide a rationale for advocacy.

In this environment, I show that, first, it’s better for the decision maker to use biased agents, even in the inquisitorial system. This is because the biased agents each have a stake in the decision maker’s decision, so they are willing to provide information even when they expect little compensation for their work. And second, I show that it’s better for the decision maker to use the adversarial process, because the problem of conflict of interests between the decision maker and the biased agents is smaller under such process, in which each agent is charged with only one specific cause to pursue. So I show that in terms of wage compensation paid to the agents, it’s best for the decision maker to adopt the adversarial system and hire two biased agents for information gathering, which I call partisan advocacy.

My third essay, "Non-robustness of Fully Revealing Equilibria in Cheap Talk Games," is contained in Chapter 4, where I study whether fully revealing equilibria are reasonable outcomes. As mentioned before briefly, this debate began with Battaglini (2002) who argues that the construction of this special equilibrium is possible if the decision maker interacts with multiple informed experts. A crucial assumption behind this striking result is that the experts know the true state without error. What if they have only noisy estimates of the true state? If a small amount of noise in their knowledge of the true state moves the set of equilibria "far from" the set of fully revealing equilibria, we can see that these special equilibria are indeed "special", with their existence not continuous with respect to the possibility of the experts’ error. Thus I say a fully revealing equilibrium is not reasonable, or not robust, if it is far from the set of equilibria we get under a small amount of noise.

To this end, I define two kinds of robustness as follows: a fully revealing equilibrium is outcome-robust (strategy-robust) if there is an equilibrium whose outcome (strategy) is close to the fully revealing equilibrium outcome (strategy) when the noise in the experts’ knowledge is small. I show that there is no outcome-robust fully revealing equilibrium in the model of Levy and Razin (2007), and discuss the connections between these new notions of robustness and the existing stability concepts.
studied in the literature.

This dissertation is organized as follows. Chapter 2 contains the first essay, "The Value of Information in Legal Systems," Chapter 3 contains the second essay, "Partisan Advocates," and Chapter 4 contains the last essay, "Non-robustness of Fully Revealing Equilibria in Cheap Talk Games".
Chapter 2

The Value of Information in Legal Systems
Abstract

This paper studies the properties of the two most commonly used legal institutions, the inquisitorial system and the adversarial system. In the former, the decision maker takes a decision based on her own acquired information, whereas under the latter the decision maker requires the advocates to present their acquired information prior to making a decision. When information is both verifiable and costly to collect, I show that the decision maker expects to make fewer decision making errors in the adversarial system than in the inquisitorial system. The main factor behind this result is that the advocate with the burden of proof values information more and, consequently, works harder to collect information under the adversarial system than the impartial decision maker under the inquisitorial system. This larger effort exerted by the advocates leads to more informed decision making under the adversarial system in spite of the advocates’ incentives to distort information.
2.1 Introduction

Consider a decision maker who has to settle a dispute between two parties. Should she collect the relevant information by herself or delegate the task to the parties instead? This question is of interest to those concerned with designing efficient judicial decision making procedures, as well as the organizational structure of firms and regulatory bodies. In continental Europe, this issue has been resolved in favor of the decision maker taking the initiative in collection and evaluation of relevant pieces of evidence, and is referred to as the inquisitorial system. In America and U.K., in contrast, the adversarial system is in use, whereby evidence collection is assumed by the interested parties, or the advocates, who drive the organization and development of the case. This paper is concerned with comparing these two decision making procedures in terms of the accuracy of the decision.

One of the main arguments in favor of the inquisitorial method, put forward by its proponents such as Tullock (1975, 1988), is its impartiality toward the dispute resolution. Since an judge or investigator appointed by the government is in charge of developing the case, one can expect greater impartiality in collecting and evaluating pieces of evidence pertaining to the dispute. In contrast, the adversarial process often leads to two clashing positions, resulting from an independent partisan search and presentation of the facts, which often involves evidence tampering that is seen as a hindrance to correct decision making.

On the other hand, one of the main arguments in support of the adversarial system is the high initiatives of the advocates in shaping the fact-finding process (Posner, 1988). The competitive nature of the adversarial process gives the advocates greater

---

1Tullock also argues that the Anglo-Saxon legal procedure is inferior because “if the litigant is likely to overinvest socially, the opponent will do the same and the outcome will depend to a considerable extent on how expensive the lawyers are,” which points out the excessive litigation cost under the adversarial system. Also see Tullock (1980).

2In practice, such evidence tampering may occur in many different forms including shredding documents, as in the Arthur Anderson case (http://news.bbc.co.uk/2/hi/business/2047122.stm), or providing a large volume of documents, as in the Goldman Sachs case, in which the Wall Street firm flooded the Financial Crisis Inquiry Commission with 2.5 billion pages of records (http://www.reuters.com/article/idUSTRE65637X20100608).

3Posner also doubts whether the expansion of the government sector by giving more initiatives to the judges under the inquisitorial system is beneficial for the society. For other criticisms on the inquisitorial system, see Posner (1999).
incentives to search for evidence than in a system where the judge or investigator is the only searcher. Also, it is often noted that, while the advocates do not need to be incentivized to work for their causes, it is not an easy task to create incentives for the judge to work with “diligence and excellence.”

This debate suggests that there exist trade-offs in adopting one system over the other. On the one hand, the adversarial system suffers from the communication problem, in the form of evidence distortion, between the informed advocates and the uninformed decision maker. On the other hand, although the inquisitorial system does not exhibit such a communication problem, it may fall short of providing enough incentives for information collection to be on a par with the adversarial system. Studying how such a trade-off may arise in a game-theoretic model and which institution is favored in terms of the accuracy of decision making is the focus of this paper.

To this end, I study a simple binary decision model in the framework of a persuasion game. The decision maker takes a binary decision to uphold the cause of one of two interested parties, the plaintiff and the defendant. The interested parties want the ruling in their favor whereas the decision maker wants to get at the truth. It is common knowledge that there exists one piece of evidence relevant to the dispute, whose value is unknown to everyone until it is discovered. In the inquisitorial system, the decision maker must exert effort to have access to this information and takes a decision according to the result of her investigation. The advocates play no role in this situation, passively waiting for the decision. In contrast, the decision maker is passive in the adversarial system in that now it is the advocates’ role to uncover the relevant information and present it to the decision maker who takes a decision. Shin (1998) also models a dispute resolution procedure as a persuasion game and compares the two legal institutions on the grounds of decision making errors. While he assumes that the information is exogenously given to the players, I relax this feature of the persuasion game to study the trade-off between the collection and communication of information.

One of the important assumptions in my model is that the decision maker is a Bayesian player of the game. Although it seems reasonable in the framework of a persuasion game, this approach is in contrast to other works that study non-Bayesian decision makers. Milgrom and Roberts (1986) and Froeb and Kobayashi (1996) show
that unsophisticated decision makers faced with the advocates’ incentives to distort information will reach a full information decision as long as the litigants’ interests are sufficiently opposed. Daughety and Reinganum (2000a,b) argue that the trial process cannot be purely Bayesian because the courts are constrained by the evidentiary rules along with other characteristics of the trial process, and take an axiomatic approach to model the court decision making. Rather than modeling the decision maker’s choice explicitly, one body of literature simply assumes a contest function which may be seen as a reduced-form approach to the courts’ behaviors. Such functions generate the winning probability of each litigant in a trial, depending on their litigation expenses.

Another important assumption in my model is that information is verifiable à la Milgrom (1981), which creates a communication problem between the advocates and the decision maker under the adversarial system. Thus the actors in a trial may conceal a body of evidence if it is harmful for their causes but cannot falsify the evidence presented to the decision maker. In contrast, Emons and Fluet (2009a,b) study a signaling game in which players may falsify their information by paying some cost. Although models with verifiable information seem reasonable in a trial setting in which falsification of evidence brings large penalties upon the party when caught, studying how the possibility of falsification may affect the litigants’ strategies along with the trial outcome is an interesting future research area.

Although there exists a communication problem under the adversarial system, the advocates have higher incentives to search for information. This can be explained by looking at how information is valued by each player in the game. Suppose the information is represented by a random variable and consider two possible realizations, $x$ and $x'$, where $x$ provides strong evidence in favor of the plaintiff and $x'$ provides only weak evidence in favor of him. Since the decision maker wants to get at the truth, she cares about the “informativeness” or “quality” of the information when she decides how much effort to exert under the inquisitorial system. If, fixing the informativeness of $x$, $x'$ were to provide more precise evidence in support of the plaintiff, the decision maker

---


5 For other works on this modeling approach, see Demougin and Fluet (2008), Dewatripont and Tirole (1999), Milgrom and Roberts (1986), and Shin (1998).
would become more confident in her decision to uphold the plaintiff’s cause under $x'$, which would give her a higher expected payoff than before. Thus, the decision maker exerts more effort to collect information as the information becomes more accurate. In contrast, the advocates do not care about the information’s quality. The plaintiff values these two realizations, $x$ and $x'$, equally as long as he wins the verdict by presenting either of them and, therefore, increasing the quality of $x'$ does not affect his valuation of the information. Since his goal is to persuade the decision maker to rule in favor of him, the information’s quality does not matter to him as long as his winning probability is the same.

Although there exists such a trade-off in my model, the main result shows that the adversarial system is better at generating more accurate decisions by the decision maker. This suggests one rationale for the use of the “partisan system,” in which the large effort exerted by the advocates in searching for information dominates the loss resulting from their incentives to distort information. This result is reminiscent of those from the principal-agent model, which identify the environments in which a principal prefers to hire a biased agent to increase the effort for information collection, in spite of the communication problem with her agent. I discuss the relevant papers and their connection to my results in more details in Section 2.5.

Dewatripont and Tirole (1999) ask similar questions to mine in a principal-agent setting in which the uninformed decision maker acquires expert advice prior to making a decision. They argue that the adversarial system generates information with lower cost because, in so far as the rewards for agents depend only on the final decision, the incentive for evidence manipulation also exists in the inquisitorial system, which makes it harder to create incentives for collecting evidence. Whereas their paper is based on the model with contracts which condition on the final decision made by the decision maker, Kim (2010) extends the model to a setting with contracts that condition on information provided by agents, and studies how hiring biased agents may benefit the decision maker, especially under the adversarial system.

Froeb and Kobayashi (2001) model the two legal systems in terms of estimators about the unknown parameter. Parisi (2002) models the choice between the two systems

---

as a continuous, rather than discrete, variable and studies how the increase in the weight
given to a system may affect litigation effort and outcomes. Emons and Fluet (2009b)
compare the two institutions in a setting in which costly falsification of evidence is
possible and study under which conditions one institution is favored to the other when
the decision maker faces a trade-off between the accuracy of her decision and the
litigation cost of the advocates. For experimental studies comparing the two institutions,
see Block et al. (2000) and Block and Parker (2004).

The article is organized as follows. Section 4.2 lays out the basic model. Section
2.3 studies the adversarial system, Section 2.4 analyzes the inquisitorial system, and
Section 2.5 compares the equilibrium outcomes across the two systems. Section 4.5
concludes with discussions on the model and directions for future research. Proofs, not
in the main body of paper, are to be found in Appendix.

2.2 Model

Consider a situation in which a decision maker has to uphold one of two
interested parties’ causes. The interested parties have diametrically opposed interests
and I refer to them as the plaintiff and the defendant. The decision maker is a
disinterested party whose concern is to get at the truth with a binary decision, \( a \in \{P, D\} \),
where \( P \) is to find in favor of the plaintiff and \( D \) is to find in favor of the defendant. The
decision maker’s payoff depends on whether her decision matches the defendant’s true
type, \( t \in \{h, l\} \). If the defendant is of high type, \( t = h \), the decision maker gets payoff 1
if she rules in favor of the defendant and payoff 0 otherwise. Similarly, if the defendant
is of low type, \( t = l \), the decision maker gets payoff 1 if she rules in favor of the plaintiff
and payoff 0 otherwise. The interested parties want the decision maker to rule in their
favor regardless of the defendant’s type. The defendant gets payoff 1 if the decision
maker upholds his cause and 0 otherwise. Similarly, the plaintiff gets payoff 1 if the
decision maker rules in favor of him and 0 otherwise.

I assume that the defendant knows his type whereas the plaintiff does not know
the defendant’s type. Although the plaintiff also has some information about the
defendant’s true behavior in some civil cases, it seems reasonable to assume that he
has much weaker information regarding the defendant’s type than the defendant himself in most civil and criminal cases. This assumption helps me to study how the interested party’s private information may help to improve decision making with the use of the burden of proof, which will be defined more precisely in a subsequent section. It should be clear from the analysis that, if the interested parties are symmetric in terms of this private information about the defendant’s type, I have stronger results.

There is one piece of evidence pertaining to the dispute, which is summarized by a signal realization $x \in X$, where $X$ is a countable set which contains all possible pieces of evidence. This evidence, if presented to the decision maker, gives her imperfect information about the defendant’s true type, and, therefore, the decision maker still faces uncertainty when taking her binary decision based on the evidence. The realization of $x$ is drawn from the conditional density $p(x|t)$ which depends on the defendant’s type $t \in \{h, l\}$. Unless this density is identical across the defendant’s types, the evidence generates useful information for decision making. I assume $p(x|t) > 0$ for all $x$ and $t$.

The decision maker forms her posterior using Bayes’ formula. This assumption has the following implication on the information structure. The set of evidence $X$ can be partitioned into $\{X^p, X^d\}$ such that

\[
x \in X^p \iff \mu_h p(x|h) < \mu_l p(x|l) \\
x \in X^d \iff \mu_h p(x|h) \geq \mu_l p(x|l)
\]

where $\mu_h (= 1 - \mu_l)$ is the prior probability that the defendant is of high type. I say $x \in X^p$ is favorable evidence for the plaintiff and $x \in X^d$ is favorable evidence for the defendant. As shown more precisely later, by applying Bayes’ rule, once the decision maker observes $x \in X^p$, there is a preponderance of evidence that favors the plaintiff. Thus, if the decision maker takes a binary decision based on the preponderance of evidence standard, she would rule in favor of the plaintiff after observing such evidence. Similarly, upon observing $x \in X^d$, the decision maker rules in favor of the defendant. Recall that I assume the two types of errors – ruling in favor of the plaintiff when the defendant is of high type, and upholding the defendant’s cause when he is of low type – give the same payoff 0 to the decision maker, which induces the decision maker to rule on the preponderance of evidence. If the decision maker were more averse to making the first type of error, the set of realizations favorable to the plaintiff, $X^p$, would be smaller.
since the decision maker now requires a higher standard of proof, \( k\mu_h p(x|h) < \mu_l p(x|l) \) where \( k > 1 \), to rule in favor of the plaintiff.

The key issue in this paper is which mode of decision making procedure is better for the accuracy of the decision, that is, whether the decision maker herself should collect relevant information for the decision to be made, or whether the interested parties should collect information and communicate what they learned to the decision maker. The former is called the inquisitorial system which is one of the characteristics of civil law countries, whereas the latter is called the adversarial system exercised mostly in common law countries. In the inquisitorial system, an impartial judge or investigator appointed by the government is responsible for developing the case and is supposed to work for both causes. In contrast, in the adversarial system, the case is organized and the evidence is developed by the interested parties’ initiatives, and the decision maker, either the judge or jury, takes a relatively passive role, reaching a decision based on the evidence presented by the advocates.\(^7\)

In the inquisitorial system, the decision maker collects information herself. She exerts effort \( e_{dm} \in [0, 1] \) with cost \( C(e_{dm}) \) and observes the evidence \( x \) with probability \( e_{dm} \). Based on her acquired information, the decision maker rules in favor of one of the interested parties. If she observed \( x \) her decision is given by \( a(x) \), and if she could not observe any evidence she takes a decision given by \( a(\phi) \) where \( \phi \) indicates the event in which no evidence is available.

In the adversarial system, the plaintiff and the defendant should exert private\(^8\) effort to gather information.\(^9\) I assume that the cost function is the same for the advocates and the decision maker. One may argue that the interested parties have lower cost of information search because they are “closer” to the fact pertaining to the dispute. Also, the inquisitorial investigation may be more costly because it is financed by public funds and controlled by the government, which may put excessive burden on the society.

---

\(^7\)In common law countries in which the adversarial procedure is usually exercised, the decision maker is usually constrained by rules such as the exclusionary rule. See, for example, Demougin and Fluet (2005) for this line of analysis. I abstract from such rules in this paper.

\(^8\)Unless I mention explicitly, I assume that, when taking her decision under the adversarial system, the decision maker does not observe the levels of effort chosen by the interested parties. See Section 4.5 for the observable effort case.

\(^9\)There is only one piece of evidence which both advocates are after. Thus, if both of them observe a piece of evidence, they observe the same evidence.
It should be clear from the analysis that my results still hold in such a case.

If the defendant is of type $t \in \{h, l\}$ and exerts effort $e'_d \in [0, 1]$ with cost $C(e'_d)$, he observes the evidence $x$ with probability $e'_d$. Similarly, the plaintiff’s effort, $e_p \in [0, 1]$, which costs $C(e_p)$, enables him to observe $x$ with probability $e_p$. Notice that the plaintiff exerts the same level of effort whereas the defendant’s effort depends on his type, because only the defendant possesses the private information about his true type. They may choose to suppress rather than to reveal their acquired evidence to the decision maker. If the evidence $x$ is revealed by any party, the decision maker takes a decision based on the evidence, $a(x)$, whereas if no evidence is revealed by the parties, the decision maker takes a decision, $a(\phi)$, forming a Bayesian posterior.

Specific assumptions on the cost function are as follow. The cost function is given by $C : [0, 1] \to \mathbb{R}_+$ such that $C(0) = C'(0) = 0$, strictly increasing, strictly convex and differentiable. I also assume that it is impossible to observe the evidence for sure since it is too costly, $\lim_{e \to 1} C'(e) = \infty$.

The adversarial decision making procedure is composed of two stages: the “investigation stage” in which the interested parties collect information and the “communication stage” in which they decide whether to reveal their acquired evidence (if they could not acquire any evidence they reveal nothing) to the decision maker who takes a binary decision based on the evidence presented to her. In the subsequent analysis, I analyze the adversarial system first followed by the inquisitorial system. Note that the communication stage in the adversarial system is not a proper subgame because the effort choices of advocates are not observed by the decision maker. The solution concept is perfect Bayesian equilibrium which is referred to simply as equilibrium.

Without loss of generality, I assume that the prior is non-degenerate and favors the defendant, i.e., $0 < \mu_l < \mu_h$. I assume that both $X^d$ and $X^p$ are not empty since the fact-finding procedure is meaningless otherwise.

**Assumption 1.** $X^d$ and $X^p$ are not empty.\(^{10}\)

I also assume that the information structure is informative in that the set of favorable evidence for the defendant is more likely when the defendant is of high type:\(^{11}\)

\(^{10}\)This is satisfied if there exists a piece of arbitrarily precise evidence for both causes and the prior is non-degenerate.

\(^{11}\)This implies that the favorable evidence for the plaintiff is more likely when the defendant is of low
Assumption 2.

\[ \sum_{x \in X^d} p(x \mid l) < \sum_{x \in X^d} p(x \mid h). \]

For example, if \( x \in X^d \) is associated with high numerical values, one may think of \( p(x \mid h) \) as first-order stochastically dominating \( p(x \mid l) \).

### 2.3 Adversarial System

#### 2.3.1 Communication Stage

In the communication subgame of the adversarial system, \((e^h_d, e^l_d, e_p)\) are given.\(^{12}\) The plaintiff’s strategy is summarized by a set of evidence which he reveals to the decision maker. The defendant’s strategy is analogous. The decision maker’s strategy is a function of available evidence, \(a(x)\) if \( x \in X \) is revealed or \( a(\phi)\) if no evidence is available.

If \( x \in X^p \) is submitted to the decision maker (either by the defendant or the plaintiff), it induces her to rule in favor of the plaintiff since by the definition of \( X^p \),

\[
\hat{\mu}_l(x) \equiv \frac{\mu_l p(x \mid l)}{\mu_l p(x \mid l) + \mu_h p(x \mid h)} > \frac{\mu_h p(x \mid h)}{\mu_l p(x \mid l) + \mu_h p(x \mid h)} \equiv \hat{\mu}_h(x)
\]

which means that the decision maker’s posterior on the defendant being of low type, \( \hat{\mu}_l(x) \), is higher than the defendant being of high type, \( \hat{\mu}_h(x) \). Thus, if the evidence is \( x \in X^p \), the decision maker rules in favor of the plaintiff. Similarly, if \( x \in X^d \) is submitted to the decision maker, it induces her to rule in favor of the defendant. As a tie-breaking rule, when indifferent, \( \hat{\mu}_h(x) = \hat{\mu}_l(x) \), the decision maker is assumed to rule in favor of the defendant.

**Lemma 1.** In the communication stage of the adversarial system, if the evidence is presented to the decision maker, she rules in favor of the plaintiff if and only if the evidence is favorable to the plaintiff.

\(^{12}\)Strictly speaking, I’m abusing terminology by referring to this as a “subgame” because the decision maker does not observe \((e^h_d, e^l_d, e_p)\). Also, these should be thought of as her belief over the interested parties’ search effort. I assume that this belief is in pure strategy, which is without loss of generality since the advocates do not mix in choosing their effort levels in the investigation stage.
Anticipating this, the defendant and the plaintiff have weakly dominant strategies when they acquired the evidence: the defendant reveals $x$ if and only if $x \in X^d$ and the plaintiff reveals $x$ if and only if $x \in X^p$. If they do not have any evidence, they reveal nothing to the decision maker. Note that it is possible that the defendant or the plaintiff is indifferent between revealing and hiding a particular $x'$, which is the case if $a(x') = a(\phi)$. If the decision maker rules in favor of the defendant in the absence of any evidence, the defendant is indifferent between revealing and hiding the favorable evidence because the decision maker’s choice is the same in both cases.\textsuperscript{13} As a tie-breaking rule, I assume that the advocates submit their evidence when indifferent. As will be shown in the later section, this tie-breaking rule does not matter in equilibrium.

**Lemma 2.** In the communication stage of the adversarial system, the interested parties reveal the evidence if and only if it is favorable to their causes.

It remains to determine the decision when there is no evidence presented in the communication stage, $a(\phi)$. In this situation, the decision maker forms her Bayesian posterior, incorporating her belief of the advocates’ investigative activities, $(e_h, e_l, e_p)$, as well as their disclosure behaviors. This belief about the levels of effort should be consistent with the interested parties’ search behavior in equilibrium.

**Lemma 3.** In the communication stage of the adversarial system, if no evidence is presented to the decision maker, she rules in favor of the plaintiff if and only if her posterior on the defendant being of high type is less than $\frac{1}{2}$.

The posterior of the decision maker in the absence of any evidence is given by, for $t \in \{h, l\}$,

\begin{align*}
\hat{\mu}_t(\phi) & \equiv \frac{\mu_t q_t}{\mu_t q_t + \mu_h q_h} \quad (2.1) \\
q_t & \equiv (1 - e_d)(1 - e_p) + e_d(1 - e_d) \sum_{x \in X^d} p(x|t) \\
& \quad + e_p(1 - e'_d) \sum_{x \in X^d} p(x|t) \quad (2.2)
\end{align*}

\textsuperscript{13}Note that when the defendant suppresses the favorable evidence, he expects that the decision maker does not observe any evidence from both advocates since the plaintiff has the same piece of evidence (if he has one) which is suppressed by the plaintiff.
where $\hat{\mu}_t(\phi)$ is the posterior belief on the defendant being of type $t$, and $q_t$ is the conditional probability of the event $\phi$ given the defendant’s type $t$. As shown in (2.2), the event $\phi$ occurs if both the defendant and the plaintiff did not obtain $x$ (first term), the defendant alone observed and suppressed $x \in X^p$ (second term), or the plaintiff alone observed and concealed $x \in X^d$ (last term). If $a(\phi) = D$, I say the burden of proof rests on the plaintiff: he can uphold his cause only when he provides supporting evidence to the decision maker. Unless the plaintiff presents a supporting document in the communication stage, the decision maker rules in favor of the defendant. Similarly, I say the burden of proof falls on the defendant if $a(\phi) = P$.

**Definition 1.** The burden of proof is said to rest on the plaintiff if $a(\phi) = D$ and on the defendant if $a(\phi) = P$.

### 2.3.2 Investigation Stage

To study the information collection effort exerted by the advocates, recall that the defendant knows his type whereas the plaintiff does not know the defendant’s type. The plaintiff’s objective is given by

$$
\max_{e_p} \mathbb{1}_{\{a(\phi) = D\}} \left[ e_p \sum_{x \in X^p} \left( \mu_h P(x|h) + \mu_l P(x|l) \right) \right] + \mathbb{1}_{\{a(\phi) = P\}} \left[ 1 - \sum_{x \in X^d} \left( \mu_h e_h^p p(x|h) + \mu_l e_l^p p(x|l) \right) \right] - C(e_p)
$$

where $\mathbb{1}_{\{\cdot\}}$ is an indicator function. If the burden of proof falls on the plaintiff, $a(\phi) = D$, the plaintiff’s objective function is given by the expression in the first bracket, which is the expected gain from information collection, minus the cost of investigation:

$$
\max_{e_p} \left[ e_p \sum_{x \in X^p} \left( \mu_h P(x|h) + \mu_l P(x|l) \right) \right] \quad \text{(2.3)}
$$

where the first term is the probability that the plaintiff finds favorable evidence for his cause. This is also the probability of winning because the plaintiff wins the decision only when he provides favorable evidence if the burden of proof falls on him. Since the payoff of 1 accrues to the plaintiff upon winning the decision and he gets payoff
0 otherwise, (2.3) is the expression for the expected payoff of the plaintiff under the burden of proof. Thus in this case the plaintiff chooses his effort level equating the marginal gain and the marginal cost of exerting effort:

\[ \sum_{x \in X^p} \left( \mu_h p(x|h) + \mu_l p(x|l) \right) = C'(e_p). \tag{2.4} \]

The LHS of (2.4) is positive and less than 1 due to Assumption 2.\textsuperscript{14} Since LHS is positive, the above condition gives \( e_p > 0 \) since \( C'(e_p) \) is increasing and takes value 0 at \( e_p = 0 \). I also have \( e_p < 1 \) since \( \lim_{e \to 1} C'(e) = \infty \). Therefore (2.4) gives us an interior solution, \( e_p \in (0, 1) \).

On the other hand, if the burden of proof rests on the defendant, the plaintiff’s objective function is given by

\[
\frac{1 - \sum_{x \in X^p} \left( \mu_h e^h_d p(x|h) + \mu_l e^l_d p(x|l) \right)}{\text{prob. of winning}} - \frac{C(e_p)}{\text{cost}} \tag{2.5}
\]

where the first term is the probability that the defendant cannot find favorable evidence for his cause. Note that the probability of winning for the plaintiff does not depend on \( e_p \) in this case. Since the burden of proof falls on the defendant, the plaintiff wins the case as long as the defendant fails to obtain supporting documents which vindicate his burden. Thus the probability of winning for the plaintiff in this case only depends on the defendant’s effort, not on the plaintiff’s. Therefore the plaintiff always chooses \( e_p = 0 \) in such a situation.

Also observe that the plaintiff’s choice of effort affects the indicator functions through its effects on equilibrium posteriors \( \hat{\mu}_t(\phi) \) for \( t \in \{h, l\} \). In equilibrium, the plaintiff could end up taking the burden if the advocates’ effort choices are consistent with the decision maker’s posteriors.

The defendant’s objective can be analogously stated as

\[
\max_{e_d'} \ 1_{\{a(\phi)=D\}} \left[ 1 - e_p \sum_{x \in X^p} p(x|t) \right] + 1_{\{a(\phi)=P\}} \left[ e_d' \sum_{x \in X^d} p(x|t) \right] - C(e_d')
\]

\textsuperscript{14}If the summation in LHS is over \( X \) (instead of \( X^p \)) then LHS is equal to 1 since \( \sum_x p(x|h) = \sum_x p(x|l) = 1 \) and \( \mu_h + \mu_l = 1 \). Since \( X^d \) is not empty by Assumption 2, \( X^p \) is a proper subset of \( X \), which implies that LHS is strictly less than 1.
for \( t \in \{h, l\} \). Since the defendant knows his own type, he knows the underlying information structure which generates supporting material. As argued before, the defendant opts out of investigative activities, \( e'_d = 0 \), if he does not bear the burden of proof. If he does bear the burden, his effort satisfies following first-order conditions, depending on his private information:

\[
\sum_{x \in X^d} p(x|h) = C'(e'_h) \quad (2.6)
\]

\[
\sum_{x \in X^d} p(x|l) = C'(e'_l). \quad (2.7)
\]

Since the evidence supporting the defendant’s claim is more easily obtainable when the defendant is of high type – Assumption 2 – we can see that the high type defendant has higher incentive to exert effort to collect information. Since the defendant knows whether the underlying information structure is favorable for his cause or not, his search activities are “variable” depending on the true state, in contrast to the plaintiff. In the next subsection, I analyze the equilibrium under the adversarial system and discuss how the variability of the defendant’s search behavior may benefit the decision making.

Studying the interested parties’ search behavior in this subsection shows that only the party with the burden of proof has incentive to search for information.\(^{15}\) This result is, obviously, not general and one can imagine a situation in which both advocates want to look for the evidence. For example, if the costly investigation of the advocates generates different pieces of evidence (\( x_d \) for the defendant and \( x_p \) for the plaintiff), both of them will search.\(^{16}\) One caveat in formulating such a model is that the information structure, as well as the searchers, differ across the two decision making regimes because there are two pieces of information (\( x_d \) and \( x_p \)) collected by the advocates in the adversarial system whereas there is only one piece of information (\( x_{dm} \)) gathered by the decision maker in the inquisitorial system. In contrast, by fixing the information structure in my model, I can focus on how the search behavior performed by different players of the game may have different effects on the accuracy of the decision.\(^{17}\)

\(^{15}\)Somewhat related results are shown in different settings. See, for example, Emons and Fluet (2009b) who show that the advocates do not testify at the same time in a signaling model in which costly falsification of evidence is possible.

\(^{16}\)Since the marginal cost is null when the level of effort is 0, both parties will exert positive amount of effort.

\(^{17}\)Another possible modeling approach is to assume that both \( x_d \) and \( x_p \) are accessed by the decision
2.3.3 Adversarial Equilibrium

Suppose the burden of proof falls on the plaintiff in equilibrium. Then the argument from the previous subsection shows that the following are necessary conditions for an equilibrium:

\[ e_p^* > 0 \]
\[ e_d^h = e_d^l = 0 \]

where \( e_p^* \) satisfies (2.4). Since the burden rests on the plaintiff, the “burdened” plaintiff engages in investigative activities while the defendant without the burden opts out of such activities. Thus \( (e_d^h, e_d^l, e_p^*) \) which satisfy above necessary conditions is a part of our tentative equilibrium.

It remains to confirm that the advocates’ strategies are consistent with the burden of proof resting on the plaintiff. Under \( (e_d^h, e_d^l, e_p^*) \), Lemma 3 shows that, in the absence of any evidence, the decision maker rules in favor of the defendant if \( \hat{\mu}_l(\phi) \leq \hat{\mu}_h(\phi) \),

\[
\hat{\mu}_l(\phi) \leq \hat{\mu}_h(\phi) \\
\iff \mu_l q_l \leq \mu_h q_h \\
\iff \mu_l \left[ 1 - e_p^* + e_p^* \sum_{x \in X} p(x|l) \right] \leq \mu_h \left[ 1 - e_p^* + e_p^* \sum_{x \in X} p(x|h) \right].
\]

Since the prior favors the defendant and Assumption 2 holds, the last inequality is true. This confirms that the advocates’ strategies are indeed consistent with the burden of proof resting on the plaintiff. Thus Lemmas 5-3, along with above arguments, prove my first Proposition.

**Proposition 1.** If the prior favors the defendant, there exists an adversarial equilibrium such that the defendant does not provide evidence, the plaintiff provides the evidence favorable for his cause with positive probability, and the decision maker rules in favor of the plaintiff if and only if the plaintiff provides favorable evidence for his cause.

In this equilibrium, only the plaintiff collects information. Since the plaintiff only provides the evidence favorable for his own cause whenever possible, this implies...
that the decision maker’s observing \( x \in X^d \) is out of equilibrium. The out-of-equilibrium path belief should be such that the decision maker upholds the defendant’s cause, which sustains the equilibrium. Also, notice that, if we consider only the equilibria in which the burden falls on the plaintiff, the multiplicity problem arises only from such out-of-equilibrium path belief, which need not be uniquely defined.

The defendant’s little incentive to exert effort is obvious since he does not benefit from information collection. Whatever evidence he may find in the investigation stage has no value to him in the communication stage, since he suppresses unfavorable evidence and the submission of favorable evidence for his cause does not change the decision which favors the defendant anyway. Thus the defendant does not value information at all. This result is reminiscent of Hay and Spier (1997), who argue that only the plaintiff has incentive to submit a supporting document to the decision maker when the burden of proof falls on him.

On the other hand, the plaintiff values information but values only “contrary” evidence: his expected payoff only depends on \( x \in X^p \) which contradicts the decision maker’s prior assessment which favors the defendant. If such evidence is more likely to be collected upon investigation, the plaintiff has higher incentive to exert effort. When there is no evidence on the table, the decision maker is unsure whether the plaintiff could not observe the evidence or he is hiding the unfavorable evidence for his cause. However, since the correct beliefs under both possibilities support the defendant, the decision maker optimally rules in favor of the defendant in the absence of any evidence, putting the burden of proof on the plaintiff.

Now suppose the burden of proof falls on the defendant in equilibrium. Then, the following are necessary conditions for an equilibrium:

\[
e^*_h, e^*_l > 0 \quad \quad \quad e^*_p = 0
\]

where \( e^*_t \) for \( t \in \{h,l\} \) satisfies (2.6) or (2.7). It remains to show that the advocates’ strategies are consistent with the burden falling on the defendant. Lemma 3 implies,
Recall that $q_t$ is the probability of the event in which no evidence is available to the decision maker in the communication stage, conditional on the defendant’s type $t \in \{h,l\}$.

As can be seen from the panel (a) in Figure 2.1, $q_t$ is a decreasing function of the defendant’s effort and $q_l$ is always higher than $q_h$ for the same value of effort level. This is because the defendant is more likely to observe $x \in X^p$ when he is of low type, which he conceals in the communication stage. The panel (b) shows that $\mu_h q_h$ is higher than $\mu_l q_l$ regardless of efforts exerted by either type of the defendant if the following

$$
\hat{\mu}_l(\phi) > \hat{\mu}_h(\phi)
\iff
\mu_l q_l > \mu_h q_h
\iff
\mu_l \left[ 1 - e_d^{h*} + e_d^{l*} \sum_{x \in X^p} p(x|h) \right] > \mu_h \left[ 1 - e_d^{h*} + e_d^{h*} \sum_{x \in X^p} p(x|h) \right].
$$

Figure 2.1: The adversarial equilibria

under $(e_d^{h*}, e_d^{l*}, e_p^{*})$, it must be true that $\hat{\mu}_l(\phi) > \hat{\mu}_h(\phi)$,
condition holds:
\[ \mu_l \leq \mu_h \left[ \sum_{x \in X^p} p(x|h) \right]. \]  \hfill (2.8)

When is this condition satisfied? Investigation of RHS of (2.8) shows that there are two effects of the prior. On the one hand, if the prior strongly favors the defendant, this condition is likely to hold since the first term in RHS increases whereas LHS decreases. On the other hand, a stronger prior reduces the set of evidence favorable for the plaintiff’s cause, \(X^p\), which lowers the second term in RHS. When the decision maker believes that the defendant is likely to be of high type, there are not many pieces of evidence which may persuade the decision maker to rule against the defendant’s claim. Thus whether (2.8) holds or not depends on the specific information structure. However, we can still see that a strong prior on the defendant may help to sustain this condition as long as the probability of getting the plaintiff’s favorable evidence when the defendant is of high type is not very small. If (2.8) does not hold, there may exist another adversarial equilibrium in which the defendant, in spite of strong support from the decision maker’s prior belief, may bear the burden of proof, which is described in the panel (c) in Figure 2.1.

**Proposition 2.** If the prior favors the defendant and the condition (2.8) holds, the burden of proof falling on the defendant cannot be an equilibrium.

Proposition 2 suggests that, when the prior favors the defendant, the adversarial equilibrium with the burden of proof on the plaintiff is “robust.” If there exist multiple equilibria, the decision maker may use the burden of proof as an instrument to coordinate on the better equilibrium. In the remaining part of this subsection, I present one rationale why the decision maker may prefer the equilibrium described in Proposition 1. Since the decision maker wants to get at the truth, her expected payoff is larger if the following quantity is smaller:
\[\mu_h \alpha + \mu_l \beta \]  \hfill (2.9)

where \(\alpha\) is the probability of ruling against the defendant when in fact he is of high type and \(\beta\) is the probability of ruling in favor of the defendant when his type is low. If we

---

\(^{18}\)In general, the decision maker may also be concerned about reducing the litigation cost as well as the decision making errors. See Section 4.5 for discussion on this issue.
view the defendant being of high type as the “null hypothesis” and the defendant being of low type as the “alternative hypothesis,” \( \alpha \) and \( \beta \) can be interpreted as type I and type II errors, respectively, and the quantity in (2.9) is the average of those two types of errors.

Under the adversarial equilibrium with the burden on the plaintiff, \( \alpha \) and \( \beta \) can be expressed as (denoting them by \( \alpha^p \) and \( \beta^p \), respectively)

\[
\alpha^p = P(\text{the decision maker observes } x \in X^p|t = h)
\]
\[
\beta^p = P(\text{the decision maker does not observe } x \in X^p|t = l).
\]

To see more clearly, suppose the defendant’s true type is high. Type I errors occur when the plaintiff reveals \( x \in X^p \) since the decision maker rules in favor of the plaintiff if and only if the plaintiff proves himself with supporting documents. Likewise, if the defendant’s true type is low, type II errors occur when there is no evidence on the table in the communication stage since the decision maker rules in favor of the defendant in such a case. Thus (2.9) becomes

\[
\mu_h \alpha^p + \mu_l \beta^p
\]
\[
= \mu_h \cdot P(\text{the decision maker observes } x \in X^p|t = h)
+ \mu_l \cdot P(\text{the decision maker does not observe } x \in X^p|t = l)
\]
\[
= \mu_h \left[ e^*_p \sum_{x \in X^p} p(x|h) \right] + \mu_l \left[ 1 - e^*_p + e^*_p \sum_{x \in X^p} p(x|l) \right].
\]

Similarly, under the adversarial equilibrium with the burden on the defendant, I have (denoting the errors by \( \alpha^d \) and \( \beta^d \), respectively)

\[
\mu_h \alpha^d + \mu_l \beta^d
\]
\[
= \mu_h \cdot P(\text{the decision maker does not observe } x \in X^d|t = h)
+ \mu_l \cdot P(\text{the decision maker observes } x \in X^d|t = l)
\]
\[
= \mu_h \left[ 1 - e^*_d + e^*_d \sum_{x \in X^d} p(x|h) \right] + \mu_l \left[ e^*_d \sum_{x \in X^d} p(x|l) \right].
\]

By comparing the average errors in these two equilibria, one can see under which conditions the decision maker would prefer one equilibrium outcome over the other in
terms of (2.9). Consider the equilibrium with the burden on the plaintiff. Ideally, in order to reduce the decision making errors, the decision maker wants the plaintiff to exert low effort when the true state is high, which lowers $\alpha$, and high effort when the true state is low, which reduces $\beta$. However, the plaintiff’s lack of knowledge of the defendant’s private information induces him to exert the same level of effort regardless of the true state. In contrast, under the equilibrium with the burden on the defendant, the defendant behaves just as the decision maker wants. When the burden of proof rests on the defendant, the decision maker wants the defendant to exert high effort when he is of high type and low effort otherwise. The defendant’s private information enables him to do exactly that: when the defendant is of high type he has higher incentives to engage in the information collection activities since he knows that favorable evidence for his cause abounds; when the defendant is of low type he knows his investigation effort would be less likely to produce favorable evidence, which lowers his incentives for the search of information. If the defendant’s effort is sufficiently variable, this will increase the benefit of giving the defendant the burden of proof.

On the other hand, it is risky to have the defendant bear the burden of proof. Since the prior favors the defendant, the type I error – ruling in favor of the plaintiff when the defendant is of high type – weighs more heavily than the type II error – ruling in favor of the defendant when he is of low type. Thus, unless the high type defendant can easily vindicate her burden, the decision maker would refrain from placing the burden of proof on the defendant.

As an example, suppose the defendant’s effort is not variable at all and assume $e_{d}^{h*} = e_{d}^{l*} = e_{p}^{*}$. Then above expressions imply $\mu_{h}\alpha^{P} + \mu_{l}\beta^{P} < \mu_{h}\alpha^{D} + \mu_{l}\beta^{D}$.$^{19}$ Since the latter is strictly higher than the former, the average error is lower under the burden on the plaintiff as long as the defendant’s effort is not sufficiently variable.

In the discussion in this subsection, I show that, when the prior favors the defendant, the adversarial equilibrium with the burden of proof on the plaintiff is robust in that it always exists, whereas the other equilibrium exists only under some conditions. Also I argue that, when the latter also exists, it may be preferred only when the

$^{19}$This is because $\mu_{h}\alpha^{P} + \mu_{l}\beta^{P} = \mu_{h}[e^{*}\sum_{x \in X^{P}} p(x|h)] + \mu_{l}[1 - e^{*} + e^{*}\sum_{x \in X^{D}} p(x|l)] = \mu_{h}[e^{*}\sum_{x \in X^{P}} p(x|h)] + \mu_{l}[1 - e^{*} + e^{*}\sum_{x \in X^{D}} p(x|l)] < \mu_{h}[1 - e^{*} + e^{*}\sum_{x \in X^{P}} p(x|h)] + \mu_{l}[e^{*}\sum_{x \in X^{D}} p(x|l)] = \mu_{h}\alpha^{D} + \mu_{l}\beta^{D}$. 
information structure is sufficiently “far apart” so that the defendant’s information search behavior is highly variable. This may provide one rationale for why the burden of proof is seldom put in question and the adversarial courts hardly appreciate the arguments for giving the defendant the burden of proof.

2.4 Inquisitorial System

The decision maker herself searches for information in the inquisitorial system. Note that there is no issue of information manipulation in the inquisitorial system since the collection and evaluation of information are performed by the decision maker. Thus, if the decision maker observes the evidence, she rules in favor of the party which her acquired evidence favors. Since there is no manipulation of information, the event $\phi$ induces the decision maker to base the decision on her prior, which favors the defendant.

**Lemma 4.** In the inquisitorial system, the decision maker rules in favor of the plaintiff if and only if she observes the evidence favorable for the plaintiff.

Prior to making a decision, the decision maker engages in the information collection, anticipating the outcome after the investigation. The decision maker’s objective is given by

$$
\max_{e_{dm}} e_{dm} \cdot \left[ \sum_{x \in X^d} \left( \mu_l p(x|l) + \mu_h p(x|h) \right) \cdot \hat{\mu}_h(x) + \sum_{x \in X^h} \left( \mu_l p(x|l) + \mu_h p(x|h) \right) \cdot \hat{\mu}_l(x) \right] \\
+ (1 - e_{dm}) \cdot \mu_h - C(e_{dm}).
$$

To see more clearly, consider (a) first. The expression in the parentheses is the probability of obtaining a particular $x \in X^d$: $\mu_l p(x|l)$ is the probability of observing $x$ when the defendant is of low type and $\mu_h p(x|h)$ is the corresponding probability with the high type defendant. Note that the decision maker needs to consider both cases since she does not observe the defendant’s true type. The second expression, $\hat{\mu}_h(x)$, is the decision maker’s posterior on the defendant being of high type after observing $x$, which is her expected gain from ruling in favor of the defendant. Since observing
If the decision maker does not observe any evidence, she rules in favor of the defendant according to her prior, which is the expression denoted by \((c)\). The last term is the cost of effort. After canceling terms, the decision maker’s objective function simplifies to

$$\max_{e_{dm}} e_{dm} \cdot \left[ \sum_{x \in X^d} \mu_h p(x|h) + \sum_{x \in X^p} \mu_l p(x|l) \right] + (1 - e_{dm}) \cdot \mu_h - C(e_{dm}).$$

Therefore, the decision maker optimally chooses \(e^*_{dm}\) which satisfies the following first-order condition:

$$\sum_{x \in X^d} \mu_h p(x|h) + \sum_{x \in X^p} \mu_l p(x|l) - \mu_h = C'(e^*_{dm}) \quad (2.10)$$

which again gives us an interior solution, \(e^*_{dm} \in (0, 1)\). Rearranging LHS of the above first-order condition gives us

$$\sum_{x \in X^d} \mu_h p(x|h) + \sum_{x \in X^p} \mu_l p(x|l) - \mu_h = C'(e^*_{dm})$$

which is strictly positive if and only if \(X^p\) is non-empty. This expression tells us two important aspects of the information search behavior in the inquisitorial system. First, the benefit of effort comes solely from the contrary evidence. Recalling that the decision maker’s prior favors the defendant, the marginal gain from her effort depends only on observing \(x \in X^p\) which upsets her prior assessment. Second, the informativeness of such contrary evidence matters. The difference \(\mu_l p(x|l) - \mu_h p(x|h)\) measures the information content of \(x \in X^p\): if this difference is large, the decision maker’s posterior
strongly favors the plaintiff upon observing such \( x \). Since the decision maker upholds the plaintiff’s cause upon observing any \( x \in X^p \), the higher expected payoff the decision maker may receive, the more informative \( x \) is. If the information content of \( X^p \) is large, the decision maker exerts more effort in the investigation stage.

**Proposition 3.** If the prior favors the defendant, there exists a unique inquisitorial equilibrium such that the decision maker provides the evidence with positive probability and rules in favor of the plaintiff if and only if she observes evidence favorable for the plaintiff.

There is no explicit burden of proof in the inquisitorial system since the collection and evaluation of the information are performed by the decision maker herself. But the inquisitorial equilibrium in Proposition 3 shows “implicit” burden of proof on the plaintiff. If the decision maker fails to collect any relevant information to the dispute, she rules against the plaintiff as if the burden of proof rests on him. Thus the quantity (2.9) under the inquisitorial equilibrium takes the same expression as that under the adversarial equilibrium with the burden of proof on the plaintiff:

\[
\mu_h \alpha + \mu_l \beta \\
= \mu_h \cdot P(\text{the decision maker observes } x \in X^p | t = h) \\
+ \mu_l \cdot P(\text{the decision maker does not observe } x \in X^p | t = l) \\
= \mu_h \left[ e_{dm}^* \sum_{x \in X^p} p(x|h) \right] + \mu_l \left[ 1 - e_{dm}^* \sum_{x \in X^d} p(x|l) \right].
\]

2.5 **Comparison of Systems**

In this section I compare the adversarial equilibrium in Proposition 1 and the inquisitorial equilibrium in Proposition 3 and show that the average error is strictly lower under the former. Since the former always exists in the adversarial system and the latter is a unique equilibrium in the inquisitorial system, I conclude that the average error is lower in the adversarial system.

Recall that, when assuming the prior favors the defendant, only the plaintiff and the decision maker have incentives to search for information. Reproducing their
incentives to investigate the evidence here gives us

\[ \sum_{x \in X} (\mu_l p(x|l) + \mu_h p(x|h)) = C'(e_p^*) \]

\[ \sum_{x \in X} (\mu_l p(x|l) - \mu_h p(x|h)) = C'(e_{dm}^*) . \]

Comparing the above two conditions tells us that the plaintiff has higher incentive to exert effort than the decision maker since LHS of the first equation is greater than LHS of the second equation. The decision maker’s value of information depends on the information content of contrary evidence against the prior. The more informative the contrary evidence is, the more search effort the decision maker exerts. In contrast, the plaintiff’s value of information does not depend on such information content. Since his goal is to persuade the decision maker to rule in favor of him, the information’s quality does not matter to him as long as he wins the decision.

Proposition 4. The plaintiff with the burden of proof in the adversarial system exerts higher effort for information collection than the decision maker in the inquisitorial system.

Note that higher effort is exerted by the plaintiff, who is a partisan agent, than the decision maker, who is an impartial agent in the model. Similar results are reported in Kim (2010) who shows how hiring partisan agents may benefit decision making, especially under the adversarial system. In a different setting, Che and Kartik (2009) also show that a partisan agent has higher incentives to search for information, which induces the decision maker to optimally hire a partisan agent in spite of the communication problem. In general, these papers identify the trade-off between collection and communication of information. On the one hand, to have better communication between the informed and the uninformed agents, it is necessary to reduce conflict of interests. On the other hand, it is often observed that non-congruent preferences create incentives for the agents to search for information harder. These papers including mine show that the interests of the decision maker is often served by a partisan agent, who balances the benefit of more information search and the loss of less information communication.\(^{20}\)

\(^{20}\)This should not read as partisan agents always having higher incentives to search for information. For
To compare the equilibrium outcomes under the two decision making procedures, recall that the average of two types of errors in both systems is expressed by

\[
\mu_h \alpha + \mu_l \beta = \mu_h \cdot P(\text{the decision maker observes } x \in X^p|t = h) + \mu_l \cdot P(\text{the decision maker does not observe } x \in X^p|t = l).
\]

To reduce the type I error, investigation effort should be low when the defendant is of high type and, to reduce the type II error, investigation effort should be high when the defendant is of low type. Proposition 4 tells us that there is a trade-off in doing this between the two systems. Higher effort by the plaintiff lowers the type II errors but raises the type I errors under the adversarial equilibrium. This trade-off between the two systems arises naturally from the trade-off between collection and communication of information discussed before. In spite of this trade-off, however, Proposition 5 shows that the adversarial system is better at increasing the accuracy of decision making.

**Proposition 5.** The average of the two types of errors is lower under the adversarial system than the inquisitorial system.

**Proof.** See Appendix.

To understand Proposition 5, let us think about the following example. Suppose \( X = \{h, l\} \) and the probabilities are given by \( p(h|h) = 0.7, p(l|h) = 0.3, p(h|l) = 0.2 \) and \( p(l|l) = 0.8 \), respectively. Assume the prior favors the defendant’s claim, i.e., \( \mu_h = 0.6 \). If the decision maker observes \( x = h \), she updates her belief so that the defendant is of high type with probability 0.84, which induces her to rule in favor of the defendant. Similarly, if the decision maker observes \( x = l \), her updated belief favors the plaintiff with probability 0.36 on the defendant being of high type and, therefore, she upholds the plaintiff’s cause.

In the inquisitorial equilibrium, the decision maker has direct access to the evidence. Assume that she observes the evidence with probability 0.7. For simplicity, I disregard the search cost in calculating the expected payoff. See Figure 2.2 for this case.
The mass on the left side of the figure indicates the probability that the decision maker observes $x = l$: she observes evidence with probability 0.7 and there is 50% chance to observe $x = l$ given that she observes the evidence. Since observing $x = l$ induces the decision maker to uphold the plaintiff’s cause, she expects to get $1 - 0.36$ which is the probability of the defendant being of low type. Thus, the expected payoff from observing $x = l$ is given by $0.7 \times 0.5 \times 0.64$ which is the first term of the expected payoff in the figure. Other terms can be similarly calculated and, therefore, her expected payoff from the inquisitorial system is 0.698 as shown.

Now consider the adversarial equilibrium with the burden on the plaintiff and assume that the plaintiff has access to the evidence with the same probability as the decision maker, 0.7. As the first panel in Figure 2.3 shows, the decision maker’s expected payoff is the same as under the inquisitorial equilibrium. To see why, first observe that the expected payoff from $x = l$ is the same as in the inquisitorial system because the plaintiff has the same probability of observing $x = l$ as the decision maker and he does not manipulate such evidence. In the absence of any evidence, the decision maker’s Bayesian posterior is a convex combination of the prior and the posterior under $x = h$ since she is not sure whether the plaintiff could not observe the evidence or he is hiding $x = h$. Since her posterior on the defendant being of high type in this case is 0.73, she rules in favor of the defendant, and Bayes’ rule tells us that the expected payoff in this case is equal to the sum of those under $\phi$ and $x = h$ in the inquisitorial system. This example demonstrates that this type of information manipulation does not hurt the decision maker as long as the decision maker is a
sophisticated player of the game in a binary decision model. Of course, this result is not general and information manipulation will be detrimental for decision making in a general environment. However, it suggests that the decision maker’s sophistication operates to mitigate the interested parties’ adverse incentives to distort information for their causes.

As there is more information collection under the adversarial regime, the second figure in Figure 2.3 shows that the plaintiff has access to the evidence with higher probability, 0.8, which increases the chance of observing \( x = 1 \) for the decision maker. In the absence of any evidence, the decision maker’s posterior is now higher, 0.74, since there is more chance of the plaintiff hiding any unfavorable evidence for his cause. As we can see from the figure, this effect operates to spread out the probability masses toward extreme posteriors, which is beneficial for the decision making since the decision maker is now more sure whether the defendant is of high or low type. Thus, the expected payoff is higher in the adversarial system.

It is interesting to observe that making contrary evidence more available, suppressing all evidence supporting the prior assessment, helps decision making. This finding suggests that it may be beneficial for decision making to distort the information structure. Meyer (1991) also finds a somewhat related result in the context of sequential sampling by a decision maker. In her model, the decision maker distorts the information structure so that the contrary evidence from subsequent signals is more valuable. If the first signal reports “high type,” this distortion makes “low type” report under the second
signal less likely than before. This makes the contrary evidence from the second signal very informative since the realization of the rare “low type” report strongly suggests the true type being low, which raises the expected payoff of the decision maker from the sequential sampling.

Another interesting implication is that the plaintiff’s opportunistic behavior in a criminal court setting may be beneficial for justice.\footnote{In this paper I assume that the decision is made based on a preponderance of evidence. It is typically supposed that the loss from convicting the innocent is larger than acquitting the guilty in a criminal case, which requires a much higher standard of proof for the decision maker to rule in favor of the prosecution, i.e., \( k \mu_h p(x|h) < \mu_l p(x|l) \) where \( k > 1 \). This can be done by changing the decision maker’s payoff appropriately in my model.} Although “[t]he duty of the prosecutor is to seek justice, not merely to convict” according to the Criminal Justice Standards of the American Bar Association, many legal scholars and political scientists are concerned about an elected or appointed prosecutor’s desire to retain office or seek employment outside of the prosecutor’s office, which induces the prosecutor to seek conviction rather than justice.\footnote{Prosecutors are often modeled to maximize the sentence on the defendant due to their inherent preferences or career concerns. Landes (1971) is the first paper which views the prosecutor as maximizing the expected punishment. See, for example, Boylan (2005) for the prosecutor’s opportunistic behavior and Gordon and Huber (2009) for a survey on the political economy of prosecution.} However, Proposition 5 suggests that the prosecutor’s incentives to convict the defendant may help to increase the accuracy of judicial decision making. In doing so, he supplies the type of information which the trier of fact would want to look for herself.

\section*{2.6 Concluding Remarks and Discussion}

This paper has explored the trade-off between the two most common legal institutions, the inquisitorial and the adversarial system, and argued that the adversarial system is better at reducing the decision making error due to the advocates’ high incentives to search for information, in spite of the communication problem stemming from its competitive character. However, it does not come without strong assumptions. I conclude by discussing my assumptions and results, and suggesting directions for future research.

Throughout the paper, I assumed that the advocates’ effort choices are not
observed by the decision maker under the adversarial system. In practice, however, the decision maker may observe some signals correlated with the advocates’ search activities. How does this affect my equilibrium analysis? To this end, assume that the decision maker observes the advocates’ effort choices. The following Proposition shows that the decision maker prefers the adversarial system because there exists a separating equilibrium in which the decision maker can identify the type of defendant from his effort level and, therefore, take the correct decision. Considering that the advocates’ effort choices are private and unlikely to be fully observed by the decision maker in practice, however, this seems to be just a theoretical construct and not a description of reality.

**Proposition 6.** Assume that effort choices of the advocates are observed by the decision maker in the adversarial system. There exists an adversarial equilibrium in which the low type defendant exerts zero effort and the high type defendant exerts strictly positive amount of effort.

**Proof.** See Appendix.

In the model, I assume that all the players have the same payoff of 1 from getting their most preferred outcomes, i.e., winning the decision for the advocates and taking the correct decision for the decision maker. Since the cost function is the same for all players, this is more than just a normalization. For example, if the payoff 2 accrues to the decision maker upon her correct decision, for a given effort level this will increase her marginal gain from investigation two-fold whereas her marginal cost remains the same. However, all the results still hold as long as the advocates value their most preferred outcomes more than the decision maker values hers since it will induce the parties to exert more effort to collect information under the adversarial system. Since, as Posner (1999) argues, the adversarial system is characterized by the high initiatives of the interested parties while it is not an easy task to create incentives for the judge to work with diligence and excellence in the inquisitorial system, this seems to be not a bad description of the reality.

In the binary decision model studied in this paper, only one advocate searches for information if he bears the burden of proof. This feature would not survive if one
considers a model in which the decision maker may be able to finely tune her decision. For example, consider a situation in which the decision maker should decide how much compensation to be paid to the plaintiff. In this formulation, the “quality of the case” now matters. The interested parties want to win with a stronger case which induces a more favorable decision for their own causes, rather than just to win the case. Since the marginal cost is null when the level of effort is 0, both advocates will engage in information collection activities as long as there is some prospect of finding a piece of favorable evidence which may sway the final verdict in favor of their causes. The study of strategic interaction among players and how to define the burden of proof properly in this context will be an interesting line of future research.

Another important assumption in this paper is that the information is verifiable so that the advocates may suppress their acquired evidence but cannot falsify it. This assumption seems reasonable in light of large penalties and sanctions if the advocate is found to have committed perjury. However, as pointed out by Shin (1998), the assumption of verifiability may need to be reconsidered in a case which involves specialist or scientific evidence, since the submissions of the evidence rely crucially on current scientific understanding, including the possible controversies and uncertainties in place at the time of the dispute. If the evidence becomes harder to verify, it exacerbates the communication problem between the decision maker and the advocates, which may work against the case for the adversarial regime in terms of the decision’s accuracy.

In comparing the two decision making systems, I show that the decision is more accurate under the adversarial system. Since the decision’s accuracy is strictly higher under this regime, the decision maker would still prefer this partisan system as long as she cares about reducing mistakes sufficiently more than other aspects of legal outcomes such as deterrence and litigation costs. However, what if other goals than minimizing the decision making errors are deemed very important for the society? For example, high litigation cost under the adversarial system is often criticized by many proponents of the inquisitorial system such as Tullock (1988). Although my model is not well suited to address these questions in the full measure, it is possible that the high litigation cost in the adversarial system may operate as a “threat” to the interested parties, which may
increase the deterrence and decrease the expected litigation cost.

Suppose the defendant can choose between the high and low types in the beginning of the game. For instance, if the defendant is a physician, he may choose to give more care (i.e., high type) in performing an operation on the patient or to be negligent (i.e., low type). Other than the evidence $x$ to be collected in the fact-finding procedure, the plaintiff (a patient in this example) may bring a piece of evidence to the decision maker to initiate the dispute resolution process. This initial piece of evidence will shape the prior belief of the decision maker. For simplicity, assume that the burden of proof rests on the party who is against the prior belief in the adversarial equilibrium. If there exists sufficient correlation between the defendant’s care and the initial evidence brought in to initiate the dispute resolution process, the low type defendant is likely to be against the prior belief based on the initial evidence and, therefore, to bear the burden of proof under the adversarial system. Thus the dispute resolution is more expensive to the low type defendant under the adversarial than the inquisitorial system since he should incur litigation costs to prove his claim under the former while the decision maker incurs the cost under the latter. Therefore, if the litigation cost is large, the defendant is more likely to behave properly, or in this example the physician is more likely to exert due care in the operation, under the adversarial system.\textsuperscript{23} Furthermore, if the deterrence effect is large enough, it may also lead to lower expected litigation costs under the adversarial system. I leave more careful investigation of this issue to future research.

\textsuperscript{23}This example is related to an interesting topic called \textit{defensive medicine} in which the physicians administer treatments without clear medical benefits to protect themselves from the costly lawsuits. See, for example, Kessler and McClellan (1996). My example suggests that defensive medicine may pose a much more serious problem in common law countries in which the adversarial system is mostly employed.
2.7 Appendix

2.7.1 Proof for Proposition 5

Under both systems, the average error is given by

\[ \mu_h \alpha + \mu_l \beta \]

\[ = \mu_h \cdot P(\text{the decision maker observes } x \in X^p | t = h) \]

\[ + \mu_l \cdot P(\text{the decision maker does not observe } x \in X^p | t = l) \]

\[ = \mu_h \left[ e^*_k \sum_{x \in X^p} p(x|h) \right] + \mu_l \left[ 1 - e^*_k + e^*_k \sum_{x \in X^d} p(x|l) \right] \]

\[ = \mu_l + e^*_k \left[ \sum_{x \in X^p} \mu_h p(x|h) + \sum_{x \in X^d} \mu_l p(x|l) - \mu_l \right] \]

\[ = \mu_l - e^*_k \sum_{x \in X^p} (\mu_l p(x|l) - \mu_h p(x|h)) \]

where \( k \in \{p, dm\} \) depending on the systems. Since \( X^p \) is not empty by Assumption 2 and Proposition 4 shows \( e^*_p > e^*_{dm} \), this proves that the average error is strictly lower under the adversarial system.

2.7.2 Proof for Proposition 6

Consider \( e^*_d > 0 \) and \( e^*_{d} = e^*_p = 0 \) such that

\[ 1 - C(e^*_d) = 0 \] (2.11)

along with an off-equilibrium path belief that the decision maker believes the defendant is of low type if she observes any \( \hat{e}_d \notin \{0, e^*_d\} \). Note that there exists such \( e^*_d \) satisfying the above equation due to the assumption on the cost function. The high type defendant is best responding since his expected payoff is 0 while any deviation generates non-positive expected payoff for him due to the off-equilibrium path belief. Similarly, the low type defendant is also best responding since his expected payoff is 0 under zero effort while any deviation generates non-positive expected payoff for him. The plaintiff is best responding by exerting no effort since his search effort does not affect the decision in a separating equilibrium. Depending on whether she observes \( e^*_d \)
or $e_d^{x}$, the decision maker can correctly identify the defendant’s type, which leads to the full-information decision. Also, note that the fact-finding procedure is meaningless since the decision maker learns the defendant’s type from his effort choice. Since there is no incentive to deviate for any player, this constitutes an equilibrium.
Chapter 3

Partisan Advocates
Abstract

This paper studies the problem of an uninformed decision maker who acquires expert advice prior to making a decision. I show that it is less costly to hire partisan agents than impartial agents, especially under advocacy, and that the decision maker prefers partisan advocacy to other forms of institutions. I also extend the literature, originating with Dewatripont and Tirole (1999), to a setting with contracts that condition on information provided and not just the decision made.
3.1 Introduction

The collection and diffusion of evidence is perhaps the most important stage in a trial process. Accurate assessment of the case is a necessary condition for efficient decision making in a court trial, which has led many economists as well as legal scholars to study how to design an institution best serving such a purpose. The main object of study in this regard has been the relative performance of two existing systems, the adversarial and the inquisitorial, in the fact-finding stage of a trial. In his early contribution, Tullock (1988) argues in favor of the inquisitorial system, widely used in civil law countries of continental Europe, on two grounds. First, as an impartial investigator employed by government is responsible for developing the evidence, the inquisitorial system is less prone to evidence manipulation as opposed to the adversarial system in which the advocates support their own causes. Second, the litigation cost expensed in the fact-finding process is much higher under the adversarial system due to the competition between the advocates, which is deemed socially undesirable to many scholars. Alternatively, Posner (1999) argues that in the adversarial system, usually adopted in common law countries such as the United States, (i) the rules of adversarial court ameliorates the problem of socially excessive search activity, (ii) the bias in decision making may pose a less serious problem than in the inquisitorial system and (iii) the competition operates to correct incentive problems created by the inquisitorial system.

As Posner suggests, the fact-finding stage in legal procedures can be fruitfully modeled in economic models, which helps scholars better understand the pros and cons in different legal institutions. In an early theoretical analysis on this issue, Milgrom and Roberts (1986) show that competition between two interested parties elicits all relevant information to even less-sophisticated decision makers. Shin (1998) shows that the allocation of the burden of proof within the adversarial system enables the decision maker to extract information even when both contestants cannot provide convincing evidence. In the framework of a persuasion game, these papers assume that the advocates investigate their causes and collect hard evidence in the beginning of the game. Thus the main issue they study is diffusion of collected evidence. However, if the investigation is costly, the participation constraints may not be satisfied, and, therefore,
explicit incentives may have to be devised by the decision maker. Dewatripont and Tirole (1999) (henceforth DT) consider such a situation and conclude that advocacy generates information with lower cost. In so far as the rewards for agents depend only on the final decision taken (indirect rewards), they show that the incentive for evidence manipulation also exists in the inquisitorial system, which requires more powerful incentive schemes in generating evidence.

Their basic model has been extended in various ways, generating more insights in the issues from which they abstracted. Palumbo (2001) argues that the adversarial system may not perform as well as the existing theory predicts because there exists an optimal amount of proof-taking above which the decision maker cannot benefit from additional evidence. Palumbo (2006) shows that the advocacy may benefit from the mutual monitoring of the interested parties, which reduces the scope for manipulation. Iossa and Palumbo (2007) consider the case in which the decision maker may not be impartial. They show that information provision by interested parties generates more efficient monitoring through the appeals process and less opportunism by the decision maker. Deffains and Demougin (2008) argue that there may exist “justice inequality" under the adversarial system because investigative expenditures are constrained by individual wealth levels.

This paper extends DT’s model in a more abstract way than others. I point out that they do not consider information-based rewards because both legal institutions are “equivalent” for the decision maker when rewards can be specified in terms of the information provided. This may be a concern to proponents of the advocacy system because information-based incentive schemes are pervasive in the economics literature. Such direct rewards for information collection and diffusion also matter in the decision making process because not only the final decision itself but also the quality of the case is important. The decision that upholds the status quo may have different effects on the investigator’s career, depending on whether the status quo prevails due to a lack of evidence or the existence of convincing evidence for both causes.

I show that the advocacy generates evidence with less cost under information-based rewards if the decision maker has access to a rich pool of agents including partisans who derive private benefits from the decision that upholds one of
two causes. First, the decision maker is shown to benefit from hiring a moderately biased investigator under the inquisitorial system. The biased investigator’s willingness to exert effort to his favorite cause operates to alleviate the incentive problems although he requires increasingly higher compensation for his search activity to the cause detrimental to him. In contrast, the latter effect does not exist in the adversarial system. Advocates’ inherent motives to marshal their causes help the decision maker to be less concerned with designing costly incentive schemes. I show that partisan advocacy, an institution in which the investigative tasks are allocated to two distinct partisan agents, is most preferred by the decision maker.

This paper is organized as follows. Section 4.2 presents DT’s model and identifies two important assumptions that drive their results. Their model is generalized by relaxing these two assumptions. Section 3.3 analyzes the model assuming the evidence acquired by agents is public information. The case of private information, which generates the manipulation incentives of agents, is analyzed in Section 3.4, and Section 4.5 concludes.

### 3.2 The Model

This section presents the generalization of the model by DT. In the following subsection their model is described and two important assumptions are identified. In the next subsection, I relax these two assumptions, which close the model.

#### 3.2.1 Dewatripont and Tirole’s Model

A risk-neutral decision maker takes one of three decisions: $L$, $R$ and the “status quo.” Decisions $L$ and $R$ are to be interpreted as favoring interest groups $L$ and $R$, respectively, and the status quo is a moderate decision. For example, $L$ and $R$ might be the plaintiff and the defendant competing in a court trial. The status quo would then correspond to the equal division of a pie, while the other two decisions favor one party over the other.

The decision maker’s preferences depend upon the realization of an unknown state of the world $\theta \in \{-1, 0, 1\}$ where $\theta = \theta_L + \theta_R$. The parameter $\theta_L$ takes the value
−1 with probability \( \alpha \) and 0 with the remaining probability. Similarly, \( \theta_R \) takes the value 1 with probability \( \alpha \) and 0 with the remaining probability. These two parameters are independently distributed, which implies

\[
Pr(\theta = 1) = Pr(\theta = -1) = \alpha(1 - \alpha)
\]

\[
Pr(\theta = 0) = 1 - 2\alpha(1 - \alpha).
\]

Under full information, the decision maker would take decision \( L \) in state \( \theta = -1 \), decision \( R \) in state \( \theta = 1 \), and the status quo otherwise. Thus, \( \theta_L = -1 \) and \( \theta_R = 1 \) are to be interpreted as pieces of information favorable to causes \( L \) and \( R \), respectively. The status quo prevails when either there is no case for either cause or there is favorable information to both causes.

Under imperfect information, the decision maker hires agents to collect information about \( \theta_L \) and \( \theta_R \). An agent must incur unverifiable disutility of effort \( K \) to collect information to cause \( i = L, R \). If he incurs \( K \) and \( |\theta_i| = 1 \), the agent acquires hard evidence \( P_i \) that \( |\theta_i| = 1 \). With the remaining probability, he either learns nothing (denoted \( \phi \)) or obtains two conflicting pieces of information—an argument \( P_i \) and a counterargument \( N_i \)—that nullify each other in that learning two conflicting arguments is informationally equivalent to learning nothing. Letting \( \phi_i = \{P_i, N_i\} \), I assume that \( \phi \) is obtained with probability \( \beta(1 - q) \) and \( \phi_i \) with probability \( (1 - \beta)(1 - q) \). If the agent incurs \( K \) and \( \theta_i = 0 \), then he obtains \( \phi \) with probability \( \beta \) and \( \phi_i \) with probability \( 1 - \beta \). The agent always learns nothing if he does not incur effort.

The decision maker takes the optimal decision conditional on the evidence gathered by the agents. Under imperfect information about her preferences, the decision maker may make one of three types of errors. Decision making exhibits \textit{inertia} when \( |\theta| = 1 \), and yet, the status quo prevails. Conversely, it may suffer from \textit{extremism} if \( \theta = 0 \) and one of two causes is embraced. Finally, there may be \textit{misguided activism} when \( \theta = -1 \) and cause \( R \) is selected or when \( \theta = 1 \) and cause \( L \) prevails. The losses associated with \textit{inertia}, \textit{extremism} and \textit{misguided activism} are denoted by \( l_I \), \( l_E \) and \( l_M \), respectively. Notice that these are losses when the decision maker makes an inefficient decision under full information. The corresponding expected losses conditional on imperfect information collected by agents are denoted by \( \hat{l}_I \) and \( \hat{l}_E \), which are calculated
below. In computing \(\hat{I}_I\) and \(\hat{I}_E\), I assume that the agents have exerted effort and the decision maker knows the information collected by the agents.\(^1\)

The posterior belief \(\hat{\alpha}\) that \(|\theta_i| = 1\) conditional on \(\phi\) or \(\phi_i\) having been discovered is given by
\[
\hat{\alpha} = \frac{\alpha - x}{1 - x} < \alpha
\]
where \(x = \alpha q\) is the unconditional probability of obtaining the hard evidence favorable to a cause when exerting effort. Thus, when the information is \((P_L, \phi)\) or \((P_L, \phi_R)\),\(^2\) the decision maker assigns probability one to \(\theta_L = -1\) and probability \(\hat{\alpha}\) to \(\theta_R = 1\). The expected loss from choosing cause \(L\) is \(\hat{\alpha} l_E\), while the expected loss from choosing the status quo is \((1 - \hat{\alpha}) l_I\). I assume that the optimal decision is to choose cause \(L\):

**Assumption 1.** \(\hat{I}_I \equiv (1 - \hat{\alpha}) l_I - \hat{\alpha} l_E > 0\).

Similarly, when the information is \((\phi, \phi)\), \((\phi_L, \phi)\), \((\phi, \phi_R)\) or \((\phi_L, \phi_R)\), the expected loss from choosing the status quo is \(2\hat{\alpha}(1 - \hat{\alpha}) l_I\), while the expected loss from choosing one of the causes is \([1 - 2\hat{\alpha}(1 - \hat{\alpha})] l_E + \hat{\alpha}(1 - \hat{\alpha}) l_M\). We assume that the expected loss from moving away from the status quo is positive in such cases:

**Assumption 2.** \(\hat{I}_E \equiv [1 - 2\hat{\alpha}(1 - \hat{\alpha})] l_E + \hat{\alpha}(1 - \hat{\alpha})(l_M - 2l_I) > 0\)

which is always satisfied when \(l_M \geq 2l_I\).

The agents are risk-neutral and protected by limited liability. Their reservation utility is normalized to zero. DT assume that the agents receive no private benefits from the final decision, and therefore, they must be rewarded monetarily to search for information. Furthermore, DT assume that information is not contractible and work with reward schemes that depend solely on the final decision.

**Assumption 3.** The agents receive no private benefits from the final decision.

**Assumption 4.** Information is not contractible.

\(^1\)As shown in the following sections, the optimal contract is such that the agents exert effort and do not manipulate acquired information when \(K\) is small. Because I only consider small values of \(K\), this assumption is without loss of generality.

\(^2\)That is, after investigation, \(P_L\) is collected from cause \(L\) and \(\phi\) (or \(\phi_R\)) from cause \(R\).
Such monetary payments are called *decision-based rewards*, whose foundations are provided in DT. An agent’s utility from receiving wage $w$ and exerting effort on $n$ tasks ($n \leq 2$) is therefore $w - nK$.

DT also assume that each cause is investigated by at most one agent due to the costly nature of the duplication of information. The key issue the decision maker faces, therefore, is whether the same agent investigates both causes (the inquiry), or whether tasks are allocated to two distinct agents (the advocacy).

Under these assumptions, DT show that advocacy generates a lower expected loss to the decision maker. Interested readers are referred to DT’s analysis for more details. In the next subsection, I discuss and relax Assumptions 3 and 4 to investigate this issue in a more general model.

### 3.2.2 Information-based Rewards and Partisans

In their model, DT assume that information is not contractible and focus on decision-based rewards because the inquisitorial system and the adversarial system generate the same expected loss to the decision maker when information is contractible:

Indeed, the organization would be indifferent between one and two agents if information-based rewards could be specified. It would then suffice to promise agents $K/x$ per piece of evidence. (DT, p15)

Because the study of information-based incentive schemes is pervasive in the economics literature, this equivalence of the two institutions may be of concern to the proponents of the adversarial system. I point out that this is due to Assumptions 3 and 4. In this paper, I show that if the decision maker has access to a rich pool of agents, the adversarial system still generates a lower expected loss to the decision maker than the inquisitorial system when information is contractible. For this purpose, I introduce new terminologies:

**Definition 1.** An L-agent (R-agent) is the agent who receives a positive private benefit $U_L$ ($U_R$) from the final decision L (R) and no private benefit from other decisions.

---

3See Palumbo (2006) for a relaxation of this assumption.
Together, we call them partisans. The agents who receive no private benefit from the final decision are called non-partisans.\(^4\)

Partisans are ubiquitous. In a criminal court, a prosecutor may receive positive benefit from a conviction due to either inherent preference or induced preference, such as career concern.\(^5\) With these terminologies, we can see that DT only consider non-partisans in the pool of agents to whom the decision maker has access. I replace Assumptions 3 and 4 with the following general assumptions and compare the relative performances of legal procedures in the fact-finding.

**Assumption 5.** The pool of agents consists of partisans as well as non-partisans.

**Assumption 6.** Information is contractible.

A few remarks are in order. First, as the subsequent analysis shows, the optimal information-based reward does not leave any rent to non-partisans and partisans with small private benefits.\(^6\) Because the analysis focuses on such a situation, I solely work with information-based rewards in this paper without loss of generality.\(^7\) Second, for simpler exposition, I sometimes restrict the set of information wages without affecting the results.\(^8\) In Section 3.3, I assume that the information agents collect is public so that the only decision agents make is whether to exert efforts as required by the decision maker or to shirk. In Section 3.4, the information agents collect is private. This is the case when agents may claim property rights on the information they collect. Therefore, the agents may not want to reveal their information if it is not in their interests to do so.

---

\(^4\)I may assume that the L-agent (R-agent) receives a positive private benefit from both the status quo and the decision \(R\) (\(L\)). As long as the L-agent (R-agent) has a higher private benefit from the decision \(L\) (\(R\)), I have qualitatively the same results.

\(^5\)See Gordon and Huber (2009) for a discussion on the prosecutor’s preferences and references therein.

\(^6\)That is, their participation constraints are merely met so that the agents receive just their reservation utility when induced to investigate the causes.

\(^7\)Since the decision maker and agents are risk-neutral and information-based reward does not leave any rent to the agents, any other types of rewards—such as a hybrid of information-based and decision-based rewards—are not strictly preferred by the decision maker.

\(^8\)Under the entire set of information wages, I have the same result with a larger set of optimal rewards.
3.3 Optimality of Partisan Advocacy

In this section, the information agents collect is public: the decision maker knows the results of the investigation by the agents. Thus the decision maker’s problem is to design the optimal reward that induces the “right” amount of efforts by agents with minimum cost. I assume that the decision maker chooses the optimal values of $w_{LR}$, $w_L$, $w_R$ and $w_0$ without loss of generality, where $w_{LR}$ is the wage for providing both pieces of hard evidence $P_L$ and $P_R$, $w_L$ ($w_R$) is the wage for providing only $P_L$ ($P_R$), and $w_0$ is the wage when none of these pieces of hard evidence is provided. I consider four possible institutions depending on whether one or two agents are hired and whether non-partisans or partisans are hired. They are presented below according to the order of analysis in the following subsections.

Definition 2. The following are four institutions considered in this paper:

- Non-partisan inquiry, with expected loss denoted by $V_N^I$, is an institution in which a single non-partisan agent investigates both causes.

- Partisan inquiry, with expected loss denoted by $V_P^I$, is an institution in which a single partisan agent investigates both causes.

- Non-partisan advocacy, with expected loss denoted by $V_N^A$, is an institution in which investigative tasks are allocated to two distinct non-partisan agents.

- Partisan advocacy, with expected loss denoted by $V_P^A$, is an institution in which investigative tasks are allocated to two distinct partisan agents.

3.3.1 Inquisitorial System

Under the inquisitorial system the same agent investigates both causes. In civil law countries, an investigator employed by the government is responsible for producing evidence while the advocates’ roles are generally limited. In this subsection, I consider two institutions, the non-partisan inquiry and the partisan inquiry, and compare their expected losses to the decision maker.
If the decision maker hires a non-partisan agent to collect information for both causes, she can promise the agent $\frac{K}{x}$ per piece of hard evidence, $P_L$ and $P_R$. This reward does not leave any rent to the agent and incentivizes him to investigate both causes. Following DT, I assume that $K$ is not too large relative to the stakes ($l_I$ and $l_E$), so that the decision maker always wants to induce the collection of information about both causes.\(^9\) Thus the decision maker’s expected loss in the non-partisan inquiry is

$$V^I_N = 2x(1-x)\hat{a}l_E + 2(1-x)^2\hat{a}(1-\hat{a})l_I + \frac{2K}{x}$$

where the decision loss is the expected loss from decision making errors based on evidence supplied by the agent, and the information cost is the expected payment made to the agent. If the agent presents both pieces of hard evidence to the decision maker, she becomes fully informed and thus makes no mistakes in the final decision. If only one favorable evidence is discovered for one of two causes (which occurs with probability $2x(1-x)$), the final decision is made according to the favorable evidence, which incurs the expected error $\hat{a}l_E$. Similarly, if no evidence is found to support both causes (which occurs with probability $(1-x)^2$), the status quo prevails and the decision maker suffers the expected error $2\hat{a}(1-\hat{a})l_I$. Summing up the terms gives us the expression for the decision loss. The decision maker pays $\frac{K}{x}$ for each piece of hard evidence, $P_i$ for $i = L, R$. Recalling that such an event occurs with probability $x$ for $i = L, R$ gives us $2K$ as the information cost.

Instead, suppose the decision maker hires an $L$-agent (or an $R$-agent without changing the result). Then she needs to find the optimal reward satisfying the following conditions:

$$H = 0 \quad (3.1)$$

$$H \geq x(w_L + U_L) + (1-x)w_0 - K \quad (3.2)$$

$$H \geq xw_R + (1-x)w_0 - K \quad (3.3)$$

$$H \equiv x^2w_{LR} + x(1-x)(w_L + U_L + w_R) + (1-x)^2w_0 - 2K$$

where the first condition says that the decision maker does not leave any rent to the $L$-agent, and the other conditions ensure that the $L$-agent investigates both causes

\(^9\)See Appendix for details.
rather than only one. Notice that if $U_L > K/x$, it is impossible to satisfy the first two conditions: this $L$-agent has a strong “partisan bias,” which makes it impossible to induce him to investigate both causes without leaving him positive rent. I assume that the partisan bias is not too strong, $U_L \leq K/x$, and, therefore, that it is possible to satisfy the above conditions. There is a set of rewards that satisfy the above conditions. For example, the following rewards scheme is one of them:

$$
\begin{align*}
w_{LR} &= 2K/x \\
w_L &= K/x - U_L \\
w_R &= K/x \\
w_0 &= 0.
\end{align*}
$$

This implies that I can calculate the information cost for the decision maker from (3.1):

$$
\text{information cost} = x^2w_{LR} + x(1-x)(w_L + w_R) + (1-x)^2w_0 \\
= 2K - x(1-x)U_L.
$$

Thus the decision maker’s expected loss in the partisan inquiry is

$$
V_P^l = 2x(1-x)\hat{\alpha}l_E + 2(1-x)^2\hat{\alpha}(1-\hat{\alpha})l_I + 2K - x(1-x)U_L.
$$

Comparing $V_N^l$ and $V_P^l$, it is immediate that the decision loss is the same because the same amount of information is provided under the non-partisan inquiry and the partisan inquiry. However, the information cost is lower under the partisan inquiry because the decision maker may pay less when she makes decision $L$ due to the positive benefit that the $L$-agent receives. This observation gives us the first Proposition:

**Proposition 1.** Assume that the partisan bias is not too strong, $U_L \leq K/x$. Then the decision maker prefers the partisan inquiry to the non-partisan inquiry.

---

10 This is because the RHS of (3.2) is strictly positive even when $w_L = w_0 = 0$.

11 This is without loss of generality because the decision maker would choose a partisan with a small bias if she had the option. To see this, observe that if $U_L > K/x$, then the information cost for the decision maker is strictly larger than $2K - x(1-x)U_L$ from (3.4). Thus choosing the $L$-agent with bias $U_L = K/x$ is the decision maker’s optimal choice of agent.
Intuitively, a strong partisan bias requires a large amount of compensation to guarantee the search effort to the cause that the partisan does not like. Thus the decision maker prefers a partisan to a non-partisan as long as the partisan’s bias is small, which reduces the decision maker’s expenses.

### 3.3.2 Adversarial System

Under the adversarial system tasks are allocated to two distinct agents. In a typical common law trial, the fact-finding process is driven by the initiatives of the interested parties. The process is developed by litigants before a passive judge or jury who hands out a verdict based on evidence and motions presented by the parties. Although the judge has some discretion over her participation in the fact-finding procedure, it is by no means comparable to that of her counterpart in the inquisitorial system. I compare the non-partisan advocacy and the partisan advocacy in this subsection and show the benefit of hiring partisans in the legal procedure.

DT observe that if the decision maker hires only non-partisans, then to her, the adversarial system is not different than the inquisitorial system: promising \( \frac{K}{x} \) per piece of hard evidence gives the decision maker the same expected loss in both systems. Because non-partisans do not receive private benefits from the final decision, they only care about wage compensation, so that investigations to causes \( L \) and \( R \) are the same in nature to them.\(^{12}\) Hence the expected loss under the non-partisan advocacy, \( V^A_N \), is equal to that under the non-partisan inquiry, \( V^I_N \). I summarize this observation as follows:

**Proposition 2.** The decision maker is indifferent between the non-partisan inquiry and the non-partisan advocacy.

Now suppose that the decision maker hires an \( L \)-agent and an \( R \)-agent for investigation to causes \( L \) and \( R \), respectively. Then the \( L \)-agent’s expected payoff given

\(^{12}\)If a non-partisan finds it increasingly difficult to investigate both causes due to a limited amount of resources, which I do not model here, the adversarial system may still do better. See Holmström and Milgrom (1991) for this line of analysis.
the effort exerted is\(^{13}\)

\[ x(w_L + (1 - x)U_L) - K \]

which implies a no-rent-leaving wage equal to

\[ w_L = \frac{K}{x} - (1 - x)U_L. \quad (3.5) \]

Similarly, I have

\[ w_R = \frac{K}{x} - (1 - x)U_R. \quad (3.6) \]

Note that \(w_L, w_R \geq 0\) because the agents are protected by limited liability. Thus the decision maker’s expected loss in the partisan advocacy is

\[ V^A_P = \underline{\text{decision loss}} + \underline{\text{information cost}} \]

\[ = 2x(1 - x)\hat{\alpha}l_E + 2(1 - x)^2\hat{\alpha}(1 - \hat{\alpha})l_I 
+ x\left( w_L \cdot I_{\{K/x(1-x) > U_L\}} + w_R \cdot I_{\{K/x(1-x) > U_R\}} \right) \]

where \(I_{\{\}}\) is an indicator function. Given that decision losses are the same in \(V^A_N\) and \(V^A_P\), the lower information cost due to “utility bonus” clearly indicates that the partisan advocacy generates a lower expected loss to the decision maker than the non-partisan advocacy.

**Proposition 3.** The decision maker prefers the partisan advocacy to the non-partisan advocacy.

### 3.3.3 Comparison

To have a complete ranking of possible institutions, I compare the partisan advocacy and the partisan inquiry in this subsection. Recall that I assumed that partisan biases are not too strong, \(U_L \leq K/x\) (and similarly \(U_R \leq K/x\)), in partisan inquiry. Under this assumption, the information cost in \(V^A_P\) becomes

\[ \text{information cost} = x(w_L \cdot I_{\{K/x(1-x) > U_L\}} + w_R \cdot I_{\{K/x(1-x) > U_R\}}) \]

\[ = x(w_L + w_R) \]

\[ = 2K - x(1 - x)(U_L + U_R) \]

\(^{13}\)I assume that the reward scheme is such that the decision maker pays a positive wage only when the \(L\)-agent presents a piece of hard evidence, \(P_L\). Since this does not leave any rent to the \(L\)-agent, this is without loss of generality.
while the information cost in $V^I_p$ is $2K - x(1-x)UL$. Therefore, the information cost is lower in $V^A_p$ because the decision maker may pay less when she makes either decision $L$ or $R$ due to partisan benefits. For larger values of the partisan bias, $UL > K/x$ (and similarly $UR > K/x$), the information cost in $V^I_p$ is higher than $2K - x(1-x)UL$ (see footnote 11), while the information cost in $V^A_p$ is lower than that. Thus the partisan advocacy is still better than the partisan inquiry when the partisan bias is large. Thus I have the following ranking between the two institutions.

**Proposition 4.** The decision maker prefers the partisan advocacy to the partisan inquiry.

The above analysis shows that a partisan with a small bias is useful in the inquisitorial system, contrary to the existing argument for its impartiality. His willingness to exert effort without significant monetary incentives reduces expenses by the decision maker in incentivizing the partisan to investigate both causes.\textsuperscript{14} However, as the partisan bias becomes stronger, the decision maker finds it increasingly costly to induce the partisan to search for the evidence that is detrimental to the partisan’s favorite decision. This induces the decision maker to select a “moderate” partisan to achieve the maximum fact-finding activity with the minimum cost under the inquisitorial system. This suggests that the presence of the bias in the inquisitorial investigator’s fact-finding activity may not pose a serious problem. Rather, by designing an optimal incentive scheme, the decision maker can achieve higher investigative activity with a lower cost.\textsuperscript{15}

However, the decision maker does not encounter any difficulty with hiring partisans under the adversarial system. In fact, she prefers to work with partisans with strong biases so that she does not need to incur any expenses to motivate their search activities. This finding confirms the idea that the advocates’ intrinsic motives operate to alleviate the incentive problems in fact-finding procedures.

Thus far, I have considered the case in which the agents cannot manipulate evidence they acquired in investigating the causes. Therefore, allowing such a

\textsuperscript{14}In his study of bureaucrats’ motivation, Prendergast (2007) argues that “their willingness to exert effort without significant monetary incentives is often attributed to such intrinsic motivation, and there is a large literature on the importance of hiring bureaucrats with the right preferences.”

\textsuperscript{15}A similar result can be found in Che and Kartik (2009), albeit in a different setting, who show that the decision maker optimally hires a prior-biased agent to induce more search activity.
possibility may pose a threat to the case for the partisan advocacy. In fact, “evidence tampering”—destruction, fabrication, and suppression of evidence—in the adversarial court is considered a serious threat to the justice by many legal scholars.\footnote{Such evidence tampering may occur in many different forms including shredding documents, as in the Arthur Anderson case (http://news.bbc.co.uk/2/hi/business/2047122.stm), or providing a large volume of documents, as in the Goldman Sachs case, in which the Wall Street firm flooded the Financial Crisis Inquiry Commission with 2.5 billion pages of records (http://www.reuters.com/article/idUSTRE65637X20100608).} I turn to this case in the next section.

3.4 Partisan Advocacy under Private Information

In this section, I investigate the optimality of partisan advocacy in a more general environment. Instead of assuming that the results of agents’ investigations are public, I assume that they are private. This is the case when agents may claim property rights on the information they collect. Therefore, the agents may not want to reveal their information truthfully if it is not in their interests to do so. Following DT, I assume that information can be concealed: that is, the agents can decide whether to report the $P_i$’s and $N_i$’s to the decision maker. Manipulation thus consists of either (i) announcing $\phi$ when one has hard evidence $P_i$ or (ii) announcing $\phi$ or revealing only $P_i$ when one in fact obtained the two conflicting pieces of information $\phi_i = \{P_i, N_i\}$.\footnote{I do not consider revealing only $N_i$ because it is equivalent to revealing that the agent received two conflicting pieces of information.} Notice that I allow the agents to conceal their information but not to forge them. This is a realistic assumption in light of the high penalty in the case that one is found to have presented false evidence for his cause. To simplify the analysis in this section, I assume that the decision maker chooses positive wages only for $P_L$, $P_R$, $\phi_L$ and $\phi_R$, denoted $w_L$, $w_R$, $\hat{w}_L$ and $\hat{w}_R$, respectively.

3.4.1 Inquisitorial System

Suppose the decision maker hires a non-partisan agent to collect information for both causes. Then she can promise the agent $K/(x + (1-x)(1-\beta))$ per piece of information, $P_L$, $P_R$, $\phi_L$ and $\phi_R$, where $x + (1-x)(1-\beta)$ is the unconditional probability
Table 3.1: List of Possible Manipulation

<table>
<thead>
<tr>
<th>Information</th>
<th>Profitable manipulation</th>
<th>Non-profitable manipulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (P_L, P_R) )</td>
<td>( (P_L, \phi)^a )</td>
<td>( (\phi, \phi) )</td>
</tr>
<tr>
<td>( (P_L, \phi_R) )</td>
<td>( (P_L, P_R)^b )</td>
<td>( (\phi, \phi_R), (P_L, \phi) )</td>
</tr>
<tr>
<td>( (P_L, \phi) )</td>
<td>( (\phi, \phi) )</td>
<td>( (\phi, P_R) )</td>
</tr>
<tr>
<td>( (\phi_L, P_R) )</td>
<td>( (P_L, P_R)^c ), ( (P_L, \phi)^d )</td>
<td>( (\phi, \phi_R), (\phi, \phi) )</td>
</tr>
<tr>
<td>( (\phi_L, \phi_R) )</td>
<td>( (\phi_L, \phi_R)^e ), ( (P_L, P_R)^f ), ( (P_L, \phi_R)^g )</td>
<td>( (\phi, \phi_R), (\phi, \phi), (P_L, \phi) )</td>
</tr>
<tr>
<td>( (\phi, \phi_R) )</td>
<td>( (\phi, P_R)^h )</td>
<td>( (\phi, \phi) )</td>
</tr>
<tr>
<td>( (\phi, \phi) )</td>
<td>( (\phi, P_R)^i )</td>
<td>( (\phi, \phi) )</td>
</tr>
</tbody>
</table>

of collecting such information for each cause. This reward does not leave any rent to the agent and incentivizes the agent to investigate both causes. Also notice that the compensation is the same for \( P_L \) and \( \phi_L \) (also for \( P_R \) and \( \phi_R \)), which keeps the non-partisan from manipulating information. Thus the decision maker’s expected loss in the non-partisan inquiry is

\[
V_N^I = \frac{2x(1-x)\alpha l_E + 2(1-x)^2\alpha(1-\alpha)l_I + 2K}{l_I + l_E} \tag{56}
\]

Instead, suppose the decision maker hires an \( L \)-agent. Table 3.1 lists all possible manipulation for each event. I say a manipulated report is **profitable** if it may increase the agent’s payoff under some rewards scheme. For example, consider \( a \) in Table 3.1: it is profitable because the agent’s payoff may be higher under the manipulated report \((P_L, \phi)\) than the truthful report \((P_L, P_R)\).\(^{18}\) I say a manipulated report is **non-profitable** if it is dominated by other possible reports under any rewards scheme. For example, when the outcome of investigation by the agent is \((\phi_L, \phi_R)\), reporting \((P_L, \phi)\) is non-profitable because it generates lower payoff for the agent than \((P_L, \phi_R)\) under any rewards scheme.\(^{19}\)

The following lemma shows a necessary condition on the reward that guarantees full revelation with minimum information cost:

\(^{18}\)This is the case if \( w_L + U_L > w_L + w_R \).
\(^{19}\)This is because the wage for \( \phi \) is set equal to 0.
Lemma 5. If a reward induces full revelation with minimum information cost to the decision maker, it should satisfy following conditions:

\[ w_R = \hat{w}_R = U_L \]  
\[ \hat{w}_L = w_L + U_L. \]  

(3.7)  
(3.8)

Proof. See Appendix. \( \square \)

To induce the \( L \)-agent to investigate both causes, the reward must satisfy the following conditions:

\[ G(w_L, \hat{w}_L, w_R, \hat{w}_R, U_L) \geq x(w_L + U_L) + (1 - x)(1 - \beta)\hat{w}_L - K \]  
\[ G(w_L, \hat{w}_L, w_R, \hat{w}_R, U_L) \geq xw_R + (1 - x)(1 - \beta)\hat{w}_R - K \]  

(3.9)  
(3.10)

where \( G \) is the \( L \)-agent’s expected payoff when he incurs two units of effort, whose formula is given in Lemma 6’s proof. The following lemma shows that if the investigation cost is small, it is possible to induce two search efforts from the \( L \)-agent.

Lemma 6. The conditions (3.9) and (3.10) are satisfied if and only if

\[ U_L[1 - \beta(1 - x) - x^2] \geq K. \]  

(3.11)

Proof. See Appendix. \( \square \)

I assume that \( K \) is small, as in the public information case, so that (3.11) holds. Then using Lemma 5, the \( L \)-agent’s expected payoff with two units of effort is

\[ G(w_L, \hat{w}_L, w_R, \hat{w}_R, U_L) = w_L[1 - \beta(1 - x)] + U_L[2(1 - \beta(1 - x)) - x^2] - 2K \]
\[ \geq U_L[2(1 - \beta(1 - x)) - x^2] - 2K \]
\[ > 0 \]

where the last strict inequality follows from (3.11). This means that the decision maker can set \( w_L = 0 \), which implies \( w_R = \hat{w}_R = \hat{w}_L = U_L \) by Lemma 5. Thus, I have a simple reward scheme for the \( L \)-agent, which induces two units of effort and full revelation, as follows:

\[ w_L = 0 \]
\[ w_R = \hat{w}_R = \hat{w}_L = U_L. \]
This implies the following expected loss in the partisan inquiry for the decision maker:

\[
V_P^I = 2x(1-x)\alpha l_E + 2(1-x)^2\alpha(1-\alpha)l_I + \left( x + 2(1-x)(1-\beta) \right) U_L
\]

Finally, comparing \(V_P^I\) and \(V_N^I\) gives us the following ranking.

**Proposition 5.** Assume (3.11) holds. Then the decision maker prefers the partisan inquiry to the non-partisan inquiry if and only if \(x < 1/2\) and

\[
\frac{K}{1-\beta(1-x)-x^2} \leq U_L < \frac{2K}{x+2(1-x)(1-\beta)}.
\]

**Proof.** See Appendix.

As in the public information case, the decision maker prefers the partisan only when his bias is not too strong. However, unlike in the previous analysis, the partisan investigator is preferred if and only if the hard evidence is less likely to be obtained. Intuitively, if \(P_R\) is likely to emerge after the investigation, the \(L\)-agent is reluctant to investigate the cause \(R\), which requires high monetary incentives to be designed. This suggests that the usefulness of the partisan investigator depends on the underlying information structure. If an investigator is assigned an investigative task that is unlikely to produce useful evidence, the decision maker needs to design high monetary incentives to induce the search activity of the impartial investigator. In such a case, a biased agent, say \(L\)-agent, may prove to be useful because his willingness to exert effort to his favorite cause reduces the incentive cost, \(w_L\), along with the full revelation cost, \(\hat{w}_L\), although he requires higher compensation for the investigation to the cause detrimental to him. Eventually, the latter effect dominates, and the decision maker may optimally choose a moderately biased investigator.

### 3.4.2 Adversarial System

Suppose the decision maker hires two non-partisans. Then she can promise each agent \(K/(x+(1-x)(1-\beta))\) per piece of information, \(P_L, P_R, \phi_L\) and \(\phi_R\), where \(x+(1-\)
\[ x)(1 - \beta) \] is the unconditional probability of collecting such information for each cause. Thus the decision maker’s expected loss in the non-partisan advocacy is

\[ V^A_N = 2x(1 - x)\hat{\alpha}l_E + 2(1 - x)^2\hat{\alpha}(1 - \hat{\alpha})l_I + \frac{2K}{\text{information cost}} \]

which implies that the decision maker is indifferent between the two legal systems when hiring non-partisans.

**Proposition 6.** The decision maker is indifferent between the non-partisan inquiry and the non-partisan advocacy.

Instead, suppose the decision maker hires the \( L \)-agent for cause \( L \) and the \( R \)-agent for cause \( R \). The optimal contract for \( L \)-agent should satisfy the following two conditions:

\[ x(w_L + (1 - x)U_L) + (1 - x)(1 - \beta)\hat{w}_L - K = 0 \]
\[ \hat{w}_L = w_L + (1 - x)U_L \]

where first condition is the no-rent-leaving condition, and second condition guarantees the full revelation. These conditions give us the following optimal contract for the \( L \)-agent:

\[ \hat{w}_L = \frac{K}{x + (1 - x)(1 - \beta)} \]
\[ w_L = \hat{w}_L - (1 - x)U_L. \]

Similarly, I have

\[ \hat{w}_R = \frac{K}{x + (1 - x)(1 - \beta)} \]
\[ w_R = \hat{w}_R - (1 - x)U_R \]

which implies the following expected loss in the partisan advocacy for the decision maker (assuming \( w_L, w_R \geq 0 \)):

\[ V^A_P = 2x(1 - x)\hat{\alpha}l_E + 2(1 - x)^2\hat{\alpha}(1 - \hat{\alpha})l_I + 2K - x(1 - x)(U_L + U_R) \]

Thus, comparing information costs in \( V^A_N \) and \( V^A_P \), we have the following ranking:
Proposition 7. The decision maker prefers the partisan advocacy to the non-partisan advocacy.

3.4.3 Comparison

To have a complete ranking over possible institutions, we compare the partisan advocacy and the partisan inquiry in this subsection. Recall that I assume (3.11) holds so that it is possible to induce two units of effort under the partisan inquiry. Clearly, if the conditions in Proposition 5 are not met, we have $V_N^I \leq V_P^I$, and therefore, by Propositions 6 and 7, the partisan advocacy is superior to the partisan inquiry. Now assume that those conditions hold, i.e., $x < 1/2$ and

$$\frac{K}{1 - \beta(1 - x) - x^2} \leq U_L < \frac{2K}{x + 2(1 - x)(1 - \beta)},$$

The partisan advocacy is superior to the partisan inquiry if the following difference is positive:

$$V_P^I - V_P^A = (x + 2(1 - x)(1 - \beta))U_L - (2K - x(1 - x)(U_L + U_R))$$

$$= U_L(2x - x^2 + 2(1 - x)(1 - \beta)) + x(1 - x)U_R - 2K.$$

This is positive because

$$U_L \geq \frac{K}{1 - \beta(1 - x) - x^2}$$

$$> \frac{K}{1 - \beta(1 - x) - \frac{x^2}{2}}$$

$$= \frac{2K}{2x - x^2 + 2(1 - x)(1 - \beta)}$$

$$> \frac{2K - x(1 - x)U_R}{2x - x^2 + 2(1 - x)(1 - \beta)}$$

where the last inequality proves the claim. Thus even in the private information case, the partisan advocacy is preferred by the decision maker.

Proposition 8. The decision maker prefers the partisan advocacy to the partisan inquiry.
3.5 Concluding Remarks

Partisans are ubiquitous in all societies and organizations. They are active and willing to exert effort without significant monetary incentives. The main contribution of this article is to point out the synergy between the partisan agents and the adversarial system, as well as to generalize the existing results in the literature. I show that the competition among partisans advocating different causes, originating from their strong interests in the causes, may benefit the organization by inducing more efforts. I hope that this article stimulates more discussion on the relationship between the nature of agents’ preferences and the institutional settings in a society or an organization, which may lead us to better understand the conflicts among various institutional settings observed in everyday lives.

3.6 Appendix

3.6.1 Details for the footnote 9

Suppose no effort is made. If the decision maker takes the status quo, the expected loss is $2\alpha(1 - \alpha)l_I$. If he embraces one of causes, the expected loss is $\alpha(1 - \alpha)l_M + [1 - 2\alpha(1 - \alpha)]l_E$. When $l_M \geq 2l_I$ (by which Assumption 2 is also satisfied) the optimal decision is the status quo.

Suppose effort is made only to cause $L$. Following is a table showing expected loss in each situation:

<table>
<thead>
<tr>
<th>information</th>
<th>decision</th>
<th>expected loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_L$</td>
<td>$L$</td>
<td>$\alpha l_E + w_L$</td>
</tr>
<tr>
<td></td>
<td>s.q.</td>
<td>$(1 - \alpha)l_I + w_L$</td>
</tr>
<tr>
<td>$\phi$ or $\phi_i$</td>
<td>$L$</td>
<td>$[\hat{\alpha}\alpha + (1 - \hat{\alpha})(1 - \alpha)]l_E + \alpha(1 - \hat{\alpha})l_M$</td>
</tr>
<tr>
<td></td>
<td>s.q.</td>
<td>$[\hat{\alpha}(1 - \alpha) + \alpha(1 - \hat{\alpha})]l_I$</td>
</tr>
</tbody>
</table>

Notice that one effort is beneficial only when hard information leads the decision maker to move in that direction, i.e. when he embraces the cause $L$ if $P_L$ is collected and the
status quo otherwise. This implies that expected loss from one effort is

\[ x(\alpha l_E + K/x) + (1 - x)[\hat{\alpha}(1 - \alpha) + \alpha(1 - \hat{\alpha})l_I] = \alpha^2 q l_E + \alpha(1 - \alpha)(2 - q)l_I + K. \]

With two efforts, the loss is

\[ x^2 \cdot 2K/x + 2x(1 - x)[\hat{\alpha}l_E + K/x] + (1 - x)^2 \cdot 2\hat{\alpha}(1 - \hat{\alpha})l_I = 2\alpha^2 q(1 - q)l_E + 2\alpha(1 - \alpha)(1 - q)l_I + 2K. \]

Thus, for example, if \( l_E \) and \( K \) are small relative to \( l_I \), two efforts are better than one or zero effort.

### 3.6.2 Proof for Lemma 5

The reward must be such that the agent does not have incentive to deviate from truthful report to profitable manipulation denoted by \( a \) through \( i \) in Table 3.1, which also guarantees that he does not deviate to non-profitable manipulation. To satisfy the incentive constraint for \( a \), the reward should satisfy

\[ w_R \geq U_L. \tag{3.12} \]

Similarly, to satisfy incentive constraints for \( e \) and \( i \), the reward should satisfy

\[ \hat{w}_R \geq w_R \tag{3.13} \]

and for \( g \) and \( h \), the reward should satisfy

\[ \hat{w}_L \geq w_L + U_L. \tag{3.14} \]

Note that the agent does not want to deviate to \( b \) under (3.13). The agent does not want to deviate to \( c \) and \( d \) under (3.14). Also the agent does not want to deviate to \( f \) under (3.13) and (3.14). Thus (3.12), (3.13) and (3.14) guarantee full revelation. Obviously, (3.7) and (3.8) achieve the minimum information cost under above conditions.
3.6.3 Proof for Lemma 6

It suffices to consider only (3.9) because, under Lemma 5, (3.10) is satisfied if (3.9) holds. The $L$-agent’s expected payoff with two search efforts is given as follows:

$$G(w_L, \hat{w}_L, w_R, \hat{w}_R, U_L) = x^2(w_L + w_R) + x(1-x)(1-\beta)(w_L + U_L + \hat{w}_R + \hat{w}_L + w_R) + x(1-x)\beta(w_L + U_L + w_R) + (1-x)^2(1-\beta)^2(\hat{w}_L + \hat{w}_R) + (1-x)^2(1-\beta)\beta(\hat{w}_L + \hat{w}_R) - 2K.$$  

Thus, using Lemma 5, (3.9) is equivalent to

$$U_L[1 - \beta(1-x) - x^2] \geq K.$$  

3.6.4 Proof for Proposition 5

To have $V^{I}_N > V^{I}_P$, the information cost must be lower under $V^{I}_P$ than $V^{I}_N$. This is equivalent to

$$2K > [x + 2(1-x)(1-\beta)]U_L.$$  

Since I assume the investigation cost is small (i.e., (3.11) holds), combining (3.11) and (3.15) gives us following condition:

$$\frac{K}{1 - \beta(1-x) - x^2} \leq U_L < \frac{2K}{x + 2(1-x)(1-\beta)}$$  

which is the second condition in Proposition 5. I conclude by noting that above inequalities hold if and only if $x < 1/2$. 
Chapter 4

Non-robustness of Fully Revealing Equilibria in Cheap Talk Games
Abstract

This paper studies the robustness of fully revealing equilibria (FRE) in multidimension-multisender cheap talk games. A FRE is outcome-robust (strategy-robust) if there is an equilibrium whose outcome (strategy) is close to the FRE outcome (strategy) when the noise in senders’ observations is small. I show that there is no outcome-robust FRE in the model of Levy and Razin (2007), and discuss the connections between these new notions of robustness and the existing stability concepts studied in the literature.
4.1 Introduction

The seminal paper by Crawford and Sobel (1982) introduces a class of signaling games, called cheap talk games. Assuming the state space is an interval, they show that there is no equilibrium where the receiver takes the optimal action under full information, called a FRE, as long as there exist conflict of interests between the receiver and the sender. However, Battaglini (2002) shows that, regardless of conflict of interests, there is a dimension of compromise between the receiver and the sender if the state space is a multidimensional Euclidean space. Exploiting these compromising dimensions, he constructs a FRE in a cheap talk game with multiple senders. Ambrus and Takahashi (2008) point out that Battaglini’s construction may fail if the state space is bounded, and derive conditions for existence of the truthful FRE in such cases.\(^1\)

These findings led researchers to study whether such equilibria are reasonable. To investigate this issue, recent papers introduce noise into the model and study whether the existence of a FRE is robust to noise in the senders’ observations. One concern in the findings of these papers is that they investigate robustness of a particular FRE, which does not necessarily imply robustness of other possible FRE’s. This leads us to study the entire set of FRE’s in a simple model. The other concern our paper addresses is what the “right” robustness concept is in this literature. To address this concern, we introduce new notions to this literature: we say a FRE is \textit{outcome-robust (strategy-robust)} if there is an equilibrium whose outcome (strategy) is close to the FRE outcome (strategy) when the noise in senders’ observations is small. Section 4.3 introduces the concepts and points out logical connections between them.\(^2\)

With help of these terminologies, we can classify the existing literature as follows. Battaglini (2004) shows that the FRE proposed by Battaglini (2002) is both outcome-robust and strategy-robust. This result requires that the state space is unbounded and that the prior is an (improper) uniform distribution. Assuming

\(^{1}\)A truthful FRE is a particular FRE where all senders tell the truth. All FRE’s are equivalent to the truthful FRE in terms of outcome.

\(^{2}\)Other than which objects are converging, these two robustness concepts differ also in which topologies are used to define convergence. The strategy-robustness uses the convergence in distribution while the outcome-robustness uses the convergence in mean squared error. This implies that, in order to be outcome-robust, a FRE needs to satisfy a more stringent requirement in terms of the topology used for convergence.
large conflict of interests, Levy and Razin (2007) (henceforth LR) show that the FRE proposed by Battaglini (2002) is not strategy-robust if the dimensions are dependent (i.e. communication on one dimension may reveal information on others). Ambrus and Lu (2010) consider the case where the senders’ biases are bounded and show that there exists a sequence of information structures such that the truthful FRE is outcome-robust as long as the size of state space is “large" compared to the senders’ biases.

In Section 4.4, we show that there is no outcome-robust FRE in LR’s model. Notwithstanding the seeming similarity, there are important differences between this paper’s and LR’s results. First, LR adopt the strategy-robustness criterion while we employ the outcome-robustness criterion in deriving the non-robustness result. Our result implies that the FRE proposed by Battaglini (2002) is strategy-robust but not outcome-robust under some conditions, so the definitions have different implications. Second, LR study the robustness of the FRE proposed by Battaglini (2002) while we study the robustness of the entire set of FRE’s. Although they show that Battaglini’s FRE is not strategy-robust, they don’t provide any evidence for whether their result implies non-strategy-robustness for other possible FRE’s. This is still an open issue. In contrast, we show that there is no outcome-robust FRE in LR’s model. Third, their non-robustness result does not hold if the dimensions are independent. However, our non-robustness result holds in that case too.

In relation to Ambrus and Lu (2010), our result shows that, regardless of the size of state space, no outcome-robust FRE exists if a sender’s bias is unbounded. Thus our result implies that their universal-boundedness assumption is important for their results.

This paper is organized as follows. Section 4.2 explains the model and Section 4.3 presents two robustness criteria and compares these criteria. Section 4.4 shows our result and Section 4.5 contains discussions.

Although they show this result under lexicographic preferences assumption, they show in the supplementary note that lexicographic preferences behave similarly to single-peaked preferences with large biases.
4.2 LR’s Model

We adopt the model from LR in this paper. Following their notation, we denote a random variable by $\tilde{x}$ with generic realization $x$.

Let $n \in \{1, 2, \ldots\}$ be given, which we use to index the information structure. The receiver has to choose an action $a = (a_x, a_y)$ in $\mathbb{R}^2$ which affects all players in the game. The appropriate choice of action depends on the realization of the unknown state $\theta = (\theta_x, \theta_y)$ in $\mathbb{R}^2$. The receiver has a prior density $f(\theta)$ with full support on $\mathbb{R}^2$ and expectation $\mu = (\mu_x, \mu_y)$. There are two senders. Each sender $i \in \{1, 2\}$ has a signal $\tilde{x}^n_i$ about $\theta$ with realization $x^n_i$ in $\mathbb{R}^2$. For a given $\theta$, assume these signals are distributed according to $f^n(x^n_1, x^n_2|\theta)$ with marginal densities $f^n(x^n_1|\theta)$ and $f^n(x^n_2|\theta)$. For a given $\theta$, we denote by $\tilde{x}^n_i(\theta)$ the random variable with density $f^n(x^n_i|\theta)$. We also make following assumption:

**Assumption 3.** $f^n(x^n_1, x^n_2|\theta)$ has full support for all $\theta$.

We discuss the relevance of this assumption after presenting our solution concept. The message space for senders is $\mathbb{R}^2$. After observing their private signal realizations, both senders transmit messages to the receiver simultaneously. This implies that, as LR note, each sender views the receiver’s action as a lottery $\tilde{a} = (\tilde{a}_x, \tilde{a}_y)$ since the action might depend on the information transmitted to the receiver by the other sender. We assume that one of senders has a lexicographic preference:

**Assumption 4.** Sender 1 prefers $\tilde{a}'$ to $\tilde{a}$ if $E(\tilde{a}'_x) > E(\tilde{a}_x)$ and is indifferent between $\tilde{a}'$ and $\tilde{a}$ otherwise. Sender 2 has a well-defined utility function $U(a, \theta)$.

Upon hearing from the senders, the receiver updates his belief about the unknown state and takes the optimal action. Rather than specifying the receiver’s preference, we assume that his action is the expectation of $\tilde{\theta}$ according to his posterior.

We briefly comment on our assumptions. We refer readers to Section 4.5 for discussions on dimensionality and unboundedness of the state space. The lexicographic preference assumption can be thought of as large conflict of interests assumption since LR’s supplementary note shows that single-peaked preferences converge to

---

4See Section 4.5 for generalization of this assumption.
lexicographic preferences as the bias becomes large. This large-biasedness assumption places strong restrictions on equilibrium structures as LR’s Lemma 1, 2 and S1 show. We use this feature to prove our Proposition 7 in Section 4.4. Without this assumption, it is not clear how a sequence of equilibria may behave as noise vanishes, which is our primary object of study, since very little is known about the equilibrium structures of cheap talk games with noise.\footnote{In general, one cannot expect to have usual “arbitrage” conditions and monotonic equilibrium structures – i.e. the set of sender types which send the same equilibrium message is an interval – even under single-crossing preferences. For non-monotonic equilibrium structure, see Chen (2009) who discusses this problem in a model with privately informed receiver.}

The strategy of sender $i$ given his type $x_i^n$, $m^n_i(\cdot|x^n_i)$, is a density function over $\mathbb{R}^2$ for $i = 1, 2$. To simplify notations, we let $m^n_i$ be a realized message according to $m^n_i(\cdot|x^n_i)$. The strategy of receiver is a function $a^n : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ which is his posterior expectation of $\tilde{\theta}$. The strategies of senders induce $h^n_i(m^n_i|\theta) = \int m^n_i(m^n_i|x^n_i)f^n(x^n_i|\theta)dx^n_i$ and $h^n_i(m^n_i) = \int h^n_i(m^n_i|\theta)f^n(\theta)d\theta$ which are the density of sender $i$’s messages for a given $\theta$ and the measure of a message $m^n_i$ sent by sender $i$ respectively. Similarly define $h^n(m^n_1,m^n_2|\theta)$ and $h^n(m^n_1,m^n_2)$. Denote by $g^n(m^n_j|x^n_i)$ the conditional density over sender $j$’s message given sender $i$’s signal realization. This is the sender $i$’s belief about the other sender’s messages. Similarly denote by $g^n(m^n_j,\theta|x^n_i)$ the conditional joint density over the other sender’s messages and the unknown state given sender $i$’s signal realization.

The standard solution concept in cheap talk games is weak perfect Bayesian equilibrium (WPBE). A WPBE is a strategy profile $\{a^n(\cdot), m^n_1(\cdot|x^n_1), m^n_2(\cdot|x^n_2)\}$ with belief $q^n(\theta|m^n_1,m^n_2)$ such that

1. $a^n(m^n_1,m^n_2) = E(\tilde{\theta}|m^n_1,m^n_2)$ for given $m^n_1(\cdot|x^n_1)$ and $m^n_2(\cdot|x^n_2)$ where $E(\tilde{\theta}|m^n_1,m^n_2) = \int \theta q^n(\theta|m^n_1,m^n_2)d\theta,$

2. if $m^n_1 \in \text{supp}(m^n_1(\cdot|x^n_1))$ then $m^n_1 \in \arg\max_m \int E(\tilde{\theta}|m,m^n_2)g(m^n_2|x^n_i)dm^n_2$ for given $a^n(\cdot)$ and $m^n_2(\cdot|x^n_2)$,

3. if $m^n_2 \in \text{supp}(m^n_2(\cdot|x^n_2))$ then $m^n_2 \in \arg\max_m \int U(a(m^n_1,m),\theta)g(m^n_1,\theta|x^n_2)dm^n_1d\theta$ for given $a^n(\cdot)$ and $m^n_1(\cdot|x^n_1)$, and
\( q^n(\theta | m^n_1, m^n_2) = \frac{h^n(m^n_1, m^n_2 | \theta) f^n(\theta)}{\int h^n(m^n_1, m^n_2 | \theta) f^n(\theta) d\theta} \) if \( h^n(m^n_1, m^n_2 | \theta) > 0 \) for some \( \theta \), or arbitrary otherwise.

Note that \( q^n(\theta | m^n_1, m^n_2) \) is well-defined for all \( m^n_1 \in M^n_1 \) and \( m^n_2 \in M^n_2 \) where \( M^n_i = \bigcup x^n_i \supp (m^n_i (\cdot | x^n_i)) \) and therefore \( E(\tilde{\theta} | m^n_1, m^n_2) \) is also well-defined. This is not true if we don’t have full support: there may exist a pair of messages \( (m^n_1, m^n_2) \) which is an out-of-equilibrium message profile – even though each message is on some equilibrium path – since there may not exist signal realization profiles inducing both \( m^n_1 \) and \( m^n_2 \). In this case, expectations may depend on the specification of out-of-equilibrium belief.

Denote by \( \tilde{m}^n_i (\theta) \) the random variable for sender \( i \)’s message given \( \theta \) with density \( h^n_i (m^n_i | \theta) \). Also denote by \( E(\tilde{\theta} | \tilde{m}^n_i (\theta)) \) the random variable which is the expectation of \( \tilde{\theta} \) on observing messages sent by sender \( i \)’s strategy given \( \theta \). Similarly define \( E(\tilde{\theta} | \tilde{m}^n_1 (\theta), \tilde{m}^n_2 (\theta)) \).

This completes the description of model and equilibrium concept for fixed \( n \). Denote the game by \( G^n \) and the set of WPBE by \( \mathcal{E}(G^n) \). Denote a generic element in \( \mathcal{E}(G^n) \) by \( e^n \). We assume that the information structure changes so that both senders observe the true state in the limiting game \( G^\infty \):

**Assumption 5.** \( \tilde{\theta}^n_i (\theta) \xrightarrow{p} \theta \) as \( n \to \infty \) for all \( \theta \) where \( \xrightarrow{p} \) denotes convergence in probability.

### 4.3 Robustness Criteria

In this section, we formally define a FRE and its robustness concepts. We define a FRE as follows:

**Definition 2.** A FRE is a WPBE in \( G^\infty \) such that \( E(\tilde{\theta} | \tilde{m}^{\text{fre}}_1 (\theta), \tilde{m}^{\text{fre}}_2 (\theta)) = \theta \) for all \( \theta \).

As shown by Battaglini (2002) and Ambrus and Takahashi (2008), the set of FRE’s in \( G^\infty \) is non-empty. For example, truth-telling by both senders is a FRE (supported by an appropriate out-of-equilibrium belief) in \( G^\infty \). Pointing out the implausibility of such FRE’s with ad hoc out-of-equilibrium beliefs, Battaglini (2002) proposed a FRE which does not depend on out-of-equilibrium beliefs. This FRE exploits the compromising dimension between the receiver and the sender, where each
sender tells the truth along their dimension of common interests with the receiver. 

To test the robustness of this FRE, LR proposed a robustness criterion we call strategy-robustness, which we present in Definition 3. Intuitively, their criterion requires that the information contents of nearby equilibrium strategies be similar to those of the FRE strategy. We define another robustness criterion, outcome-robustness, in Definition 4. Outcome-robustness requires that the outcomes of nearby equilibria be close to the FRE outcome.

**Definition 3.** A FRE is strategy-robust if there exists a sequence \( \{e^n\}_{n=1}^{\infty} \) that induces, for any \( \theta \), a sequence \( \{E(\tilde{\theta}|\tilde{m}_1^n(\theta))\}_{n=1}^{\infty} \) such that \( E(\tilde{\theta}|\tilde{m}_1^n(\theta)) \xrightarrow{d} E(\tilde{\theta}|\tilde{m}_1^{\text{fre}}(\theta)) \) for \( i = 1, 2 \) where \( \xrightarrow{d} \) denotes convergence in distribution.

**Definition 4.** A FRE is outcome-robust if there exists a sequence \( \{e^n\}_{n=1}^{\infty} \) that induces, for any \( \theta \), a sequence \( \{E(\tilde{\theta}|\tilde{m}_1^n(\theta),\tilde{m}_2^n(\theta))\}_{n=1}^{\infty} \) such that \( E(\tilde{\theta}|\tilde{m}_1^n(\theta),\tilde{m}_2^n(\theta)) \xrightarrow{m} E(\tilde{\theta}|\tilde{m}_1^{\text{fre}}(\theta),\tilde{m}_2^{\text{fre}}(\theta)) \) where \( \xrightarrow{m} \) denotes convergence in mean squared error.

What is the relationship between the two criteria? Does one criterion imply the other? Do we always have the same robustness result from both criteria? If the dimensions are independent, we have different results from these criteria: in this case, LR’s Lemma 2 implies that Battaglini’s FRE is strategy-robust while our Proposition 7 shows that it is not outcome-robust. Thus one can find at least one FRE such that the strategy-robustness does not imply the outcome-robustness. We leave the exact characterization of relationship between the two criteria for future research.

### 4.4 Result

**Proposition 7.** There exists no outcome-robust FRE in the model.

**Proof.** Let \( \tilde{m}_i^n(\theta) \) be given as previously defined. Note that we have,

\[
E(\tilde{\theta}_x|\tilde{m}_i^n(\theta)) = \mu_x \quad \text{for all } n \text{ and } \theta
\]  

(4.1)

by Lemma 2 from LR.\(^6\) Fix a FRE. Since \( E(\tilde{\theta}|\tilde{m}_1^{\text{fre}}(\theta),\tilde{m}_2^{\text{fre}}(\theta)) = \theta \) by definition, the

\(^6\)Note that LHS of (4.1) is a random variable. Thus LR’s Lemma 2 means that this random variable is degenerate at constant \( \mu_x \).
FRE is outcome-robust if and only if
\[ E(\tilde{\theta}|\tilde{m}_1^n(\theta), \tilde{m}_2^n(\theta)) \xrightarrow{m} \theta \quad \forall \theta. \tag{4.2} \]

Suppose (4.2) is true. Choose \( \theta' \) such that \( \theta' \neq \mu_x \). Letting \( \tilde{y}_n = E(\tilde{\theta}_x|\tilde{m}_1^n(\theta'), \tilde{m}_2^n(\theta')) \), we know \( \tilde{y}_n \xrightarrow{m} \theta'_x \). Thus we have,
\[ E(\tilde{y}_n - \theta'_x)^2 = (E(\tilde{y}_n - \theta'_x))^2 + V(\tilde{y}_n) \xrightarrow{} 0 \tag{4.3} \]
where \( V(\cdot) \) is the variance of random variable. This implies that both terms on RHS converge to 0. On the other hand,
\[ E(E(\tilde{y}_n|\tilde{m}_1^n(\theta')) - \theta'_x)^2 = (E(E(\tilde{y}_n|\tilde{m}_1^n(\theta')) - \theta'_x))^2 + V(E(\tilde{y}_n|\tilde{m}_1^n(\theta'))) \tag{4.4} \]
First term on RHS of (4.4) converges to 0 since
\[ (E(E(\tilde{y}_n|\tilde{m}_1^n(\theta')) - \theta'_x))^2 = (E(\tilde{y}_n - \theta'_x))^2 \xrightarrow{} 0 \]
where first line follows from the law of iterated expectation and second line follows from (4.3). Second term on RHS of (4.4) also converges to 0 since
\[ V(E(\tilde{y}_n|\tilde{m}_1^n(\theta'))) \leq V(\tilde{y}_n) \xrightarrow{} 0 \]
where second line follows from (4.3) and first line follows from the well-known inequality,
\[ V(\tilde{y}_n) = V(E(\tilde{y}_n|\tilde{m}_1^n(\theta'))) + E(V(\tilde{y}_n|\tilde{m}_1^n(\theta'))). \tag{4.5} \]
Hence, LHS of (4.4) converges to 0, which means,
\[ E(\tilde{y}_n|\tilde{m}_1^n(\theta')) \xrightarrow{m} \theta'_x. \tag{4.6} \]
Thus we have,
\[ E(\tilde{\theta}_x|\tilde{m}_1^n(\theta')) = E(E(\tilde{\theta}_x|\tilde{m}_1^n(\theta'), \tilde{m}_2^n(\theta'))|\tilde{m}_1^n(\theta')) \]
\[ = E(\tilde{y}_n|\tilde{m}_1^n(\theta')) \xrightarrow{m} \theta'_x \]
which contradicts (4.1). Thus (4.2) is not true, and therefore the FRE is not outcome-robust. \qed
4.5 Discussion

We briefly discuss the generality of our results. In Proposition 7, we show that there is a contradiction to condition (4.1) by assuming that the FRE under consideration is outcome-robust. This condition, however, only depends on sender 1’s preference over lotteries in the x-dimension, and not on other assumptions. Thus the Proposition holds for any $\mathbb{R}^l, l \geq 1$, as long as there is a sender who is risk-neutral over his most preferred dimension.\footnote{More precisely, assuming $\theta \in \mathbb{R}^l$, Proposition 7 holds if there is a sender whose preference is such that $\tilde{a}' \succ \tilde{a}$ if (i) $E(\tilde{a}'_x) > E(\tilde{a}_x)$ for some dimension $x$ or (ii) $E(\tilde{a}'_x) = E(\tilde{a}_x)$ and $\tilde{a}'_{-x} \succ_{-x} \tilde{a}_{-x}$ where $\succ_{-x}$ is his preference over lotteries in other dimensions.}

Proposition 7 also holds under the bounded state space as well. In fact, if the state space is bounded, we can use a weaker convergence concept in Definition 4: instead of convergence in mean squared error, we can use convergence in probability and leave the proof for Proposition 7 intact.\footnote{Suppose the state space is bounded. Then convergence in probability is equivalent to convergence in mean squared error. (See Grimmett and Stirzaker (1992) for reference.) Thus the proof for Proposition 7 does not change.}


