Forming Alliances through Cheap Talk with Bounded Rational Agents

A Thesis submitted by

Yingting Fu

In partial fulfillment of the requirements for the degree of

Master of Science

in

Economics

Tufts University

May 2016

Advisor: Professor Enrico Spolaore

Abstract

I propose a model that is useful in the study of two themes: strategic misrepresentation of private information and formation of alliances. The game is a three-player dynamic game with incomplete information, in which coalition is impossible when all players are rational. I found multiple equilibria in which a perfectly rational sender can form a coalition with either a rational receiver or a bounded rational receiver. Moreover, it is easier for the sender to form a coalition with a rational receiver than with a bounded rational receiver in both pooling and separating equilibria. Lastly, I prove that in separating equilibria, it is possible for a perfectly rational player to deceive her perfectly rational opponents when she has a probability of being bounded rational. It is surprising that the symmetric separating equilibria exist if and only if the listeners do not play their strictly dominant strategies in the underlying game.

Acknowledgments

I am grateful to my thesis advisor, Prof. Enrico Spolaore, for his constant support, guide and encouragement throughout the course of my research. This thesis could not have been possible without him.

I would also like to thank my readers on the thesis committee, Prof. Lynne Pepall and Prof. Gilbert Metcalf, for their help and invaluable suggestions.

Finally, I am thankful to Linshuo Zhou for being a loyal reader and honest critic.

Contents

Chapter 1

Introduction

"We have no permanent friend. We have no permanent enemies.

We just have permanent interests."

–Benjamin Disraeli

"Politics makes strange bedfellows."

–Charles Dudley Warner

1.1 Cooperation and Coalitions

Cooperative activities constitute an important component of human interactions. Unlike the case of competition, which has been heavily studied and has become a foundation of equilibrium models in an economic tradition after Adam Smith, the understanding of cooperative behaviors required more careful and yet intriguing analysis: why do people engage in cooperations, and how do they choose whom to cooperate with?

Economic norms approach the first questions by specifying preferences and utilities of individuals. It is argued that by committing to cooperate, participants in an alliance get higher returns than they could get own their own. Such argument fit well into real-world examples: In international relations, countries form alliances to reduce trade tariffs, avoid military conflicts or strengthen their voices in global affairs; in politics, government, voters, and interest groups may collude to shape public policies that affect interests of their own groups; in domestic relationships, individuals form a household to enjoy matching surplus from marriage.

This essay seeks an answer to the second question, how alliances are formed. The answer would be straightforward if all parties in the potential alliance strictly prefer the coalition to the best alternatives for each of them under all circumstances, in which case an alliance can be imagined to be built without frictions. A more interesting view is to look at how decision-maker make choices with uncertainties and equally good options. For example, how did Soviet Union make the decision which side to ally with in WWII?

My contributions to this field of study are threefold. First, I propose an alternative way of understanding coalitions. Specifically, I show that when the all-inclusive coalition is not possible, an alliance can be formed due to potential allies' different ability to understand messages, and this is novel in the literature. Second, this finding can be used to explain a wide range of observations, such as belief manipulation through cheap talk in political campaigns. Third, it provides an extension of Crawford (2003) model to a three player game, which not only accommodates his major results, but also adds to the adaptability of the original model to analysis for different purposes.

This essay is organized as follows. The following part of Chapter 1 provides additional motivations and conducts literature review. Chapter 2 describes the model and results. Chapter 3 concludes and considers several extensions.

1.2 Electoral Campaigns

When candidates start out to run for office, they seek support through talks. They run campaigns, set agendas and send messages about what types of leaders they are going to become. More often than not, those messages are promises about future instead of highlightening the politician's credentials witnessed about the past. From Bill Clinton's "The economy, stupid" in his 1992 presidential campaign to the more recent "build that wall" in the wake of global terrorism, ethnic tension, and immigration policy predicament, political candidates communicate with potential voters to attract alliances. Those communications bare the characteristics of cheap talk: they are costless, non-verifiable and cannot be enforced by formal institutions.

Presumably, such communication would not have been important if the expectation on political leaders are one dimensional. For example, if all that voters want is "diplomacy", then every ambitious politician would get heavily engaged in exhibiting (or faking) such qualification. Agendas become irrelevant and the art of steering sensitive political issues will no longer be appreciated. Cheap talk gives way to the effective signaling of abilities.

However, even if desirable traits that make a great leader is multi-dimensional, the theoretical effectiveness of cheap talk in elections still remains doubtful. If a politician announces that he or she intends to implement a policy that is preferred by some voters to all other policies, will these voters set up their minds to elect this politician? Not necessarily. From words to outcomes, there are two requirements: competency and incentive. First of all, suppose that in the party system, decision making is a collective wisdom from a cabinet of political elites, then the competency of a politician him- or herself can be less a concern. In this case, the second condition can be even trickier: how can we ensure that politicians actually mean what they say?

Traditional wisdom answers this question by pointing out that “Politicians care about the next elections." Following this line of reasoning, with re-elections, our political system can motivate elected politicians to fulfill their promises during at least their first term in order to win the next election. This, in turn will restrain candidates from writing empty checks during campaigns. Such argument applies well to professional politicians, but less neatly to opportunists or players with little chance to win if following the old rule. For them, the option value of getting re-elected is much smaller, either because they do not have a long time horizon in playing politics, or because the chance of entering the current round is already so small. Elections are therefore more like a one-shot game for them. It is certainly wonderful to win a re-election four or five years later, but how can one care less about it when he or she can win the office this year by simply saying what those people want to hear?

Therefore, I argue that the incentive compatibility problem for candidates to truthfully announce agendas are not solved by the presence of re-election. In fact, as long as a democratic society opens door for political candidates from different backgrounds, and elected officials have some degree independent power, which is both enjoyable to them and cannot be easily withdrawn, some people will always have the incentive to run campaigns for the office by simply costless saying whatever their electorates would like to hear. As whether the candidate really means what he/she says becomes not verifiable, voters, in turn, should rationally ignore the message sent by such candidates. Cheap talk should never work in public elections.

Economists have attempted to address the puzzle of cheap talk effectiveness. Alesina (1988) was the first to show that if voters are rational and forward-looking, there will be dynamically inconsistent when politicians have a preference on policies, in contrast to being only motivated to win elections. More recently, Kartik and Van Weelden (2015) argued that cheap talk can be informative due to reputation concerns, but this is conditional on the credibility of policy-related promises.

I examine the function of cheap talk in broader contexts. Instead of being confined to voting, the central topic deals with alliance (or coalition) formation. Throughout this essay, a coalition is discussed by the common definition (as opposed to the definition in cooperative games): “a group of people, groups, or countries who have joined together for a common purpose." By this definition, in elections, politicians, and their constituents engage in at least some level of coalition building by steering society toward some (allegedly) shared goals. I discuss coalition in different levels. First, at cross-border level, coalitions may manifest itself as military alliances, trade partners or supporters in global issues. Second, such alliances can be observed when political parties try to form a majority in parliaments at the national level. Third, at the household level, marriage can also be viewed as a process of forming a cooperative union.

In this essay, I propose an alternative answer to the question why cheap talk can matter in coalition building. Considering the suspicious credibility of alleged agenda, I attribute the effectiveness of cheap talk to the presence of bounded rational parties. They hold beliefs easily manipulated by the talker, and therefore act as desired by the sender.

1.3 Military Alliances and Marriage

Dissimilar to political campaigns that inevitably involve collective choices, examples like military alliances and marriage is more suitable to analyzes games with a small number of players. In warfares, alliance formation involves a degree of uncertainty on the potential ally's commitment to the coalition. Good allies add to the chance of victory, while bad allies can become a stab at the back. The study of military alliances is particularly intriguing, partly because stakes at wars are usually tantamount and agreement between nations are not enforceable, and therefore incentives for deviation are ubiquitous. Therefore, it could be wondered that how those coalitions could possibly be formed, and how nations choose their allies *per se*. For example, what made the Soviet Union to sign the Molotov-Ribbentrop Pact with Nazi-Germany in 1939?

Marriage could pose a greater problem in this respect, because international relations can be regulated by repeated games (if regimes do not rise and fall apart so frequently), but getting married is a one-shot game (or at least as people hoped it to be). Economists since have long theorized marriage in terms of its structure, household production and searching (Becker 1973, 1974; Keeley 1977), but does this coalition point of view contribute to our understanding of such relationship?

1.4 Bounded Rationality

The central assumption in this paper is that people can be bounded rational in communication games. This idea is developed by Crawford (2003), who characterized bounded rational (Mortal) senders as truth-tellers or liars, and Mortal receivers as believers or inverters. The realization of such types is private information to bounded rational the agent. Crawford developed an attacker-defender model motivated by the military action in Normandy during WWII, when the Ally successfully deceived the Germans about which place to attack. In his paper, a sophisticated sender can deceive a sophisticated listener when the sender has a positive probability of being bounded rational.

Please note it is a very strong assumption that some people act in a way completely determined by their type. This treatment, however, reflects an essential characteristic of one type of bounded rational agents, who are not able to form a correct expectation about the structure of the game. Those bounded rational players, in general, can be efficient in maximizing their subjective payoffs, but their objective functions do not reflect reality. In the election case, voters behave bounded rationally when they change beliefs about the candidate's type in response to messages that do not reveal more information about the game. This can be caused due to mistaken priors and inaccurate world views. For example, a voter may think that being able to speak politically incorrectly signals candidness, which is a character that affects the true welfare of constituents. What they ignore, however, is that speaking politically incorrectly does not incur any cost to candidates, nor does it relate to the inherent personality of a candidate in any absolute way. This is a behavior that can be easily imitated by people without candidness. Limited by their experiences and knowledge, although people have their own good reason to believe that politically incorrect speeches and frankness are somehow positively correlated, choosing candidates according to this belief might nonetheless dangerous, since this presumption can be exploited by sophisticated politicians.

Furthermore, the strong form of naiveté is not necessary as long as there is room for manipulation of beliefs. In this respect, being perfectly rational becomes a greater practical challenge, as a huge subset of people's knowledge comes from repeatedly observed correlations rather than air-tight theories, and people are constantly making judgments based on those correlations without paying close attentions to factors that invalidate such application of knowledge.

1.5 Literature Review

Economic researches of bounded rationality explore one of the two approaches: limited capacity in optimization and incomplete understanding of the game structure. For the former approach, Stahl and Wilson (1995) noted that the self-reference problem made rational agent weak predictability of game. Sims (2003) also theorized rational inattention due to bounded capacities of processing information.

This paper follows more closely on the latter approach, which emphasizes inaccurate priors about the game. Jehiel (2005) examined manipulation of beliefs when bounded rational players only have a coarse understanding of other player's strategies. Bounded rational players form an analogy partition of the decision node of other players. It is proved that analogy-based equilibria require coarse knowledge available to agents to be correct. Eyster and Rabin (2005) proposed a formulation of bounded rationality that an agent underestimates the correlation between other people's private information and actions. Crawford's (2003) definition of bounded rationality also has precedents in the game theory literature: when Kreps and Wilson (1982) added trembling-hand equilibrium solutions to games of incomplete information, the idea that some player's having a tiny probability to hold inaccurate beliefs about the game structure would give rise to otherwise impossible outcomes.

Deception is another theme of this paper. As early as Von Neumann and Morgenstern (1944) formalized game theory, it has been known that mixed strategies in zero-sum games help the player from being detected. In the Spencian (1973) model of signaling, players can avoid revealing their types in pooling equilibria.

Studies more explicitly concerned with deceptions are like Sobel (1985), which draws a redefinition of lies by suggesting that when agents can only prove credibility through actions that directly influence their payoffs, an enemy will have the incentive to act as a friend in earlier stages to increase his future opportunities to exploit the decision maker. Crawford and Sobel (1982) developed a model of strategic communication, in which the receiver sends noisy signal to the receiver, and the welfare of both depends on the receiver's action. It is shown that in equilibrium, the sender always partitions the support of his type variable, and introduces noise into his signal only in partitions where he has observed elements. More recently, similar to the Crawford 2003 paper, Hendricks and McAfee (2006) used an attacker-defender model is proposed here when the attacker can choose to allocate investment between attack and misdirection. The authors found that the attacker invests fewer resources in the signaling technology when the technology is less revealing. When the opposite is true, the attacker is less likely to successfully feint the defender but invest in the signaling technology when he feints. Kartik et. al (2007) considered an alternative to the CS communication model, in which talk is costly. They showed that when the state space for the sender's types is unbounded above, the game admits a fully separating equilibrium in which messages are exaggerated and naïve listeners are deceived.

Chapter 2

A Model

2.1 Set Up

I propose a theoretic framework to examine the formation of coalitions, or alliances, in non-cooperative games. A three-player, zero-sum game is incorporated to facilitate the analysis of minimal winning coalitions. The structure of the game captures the following features.

First, it is favorable for a player to join a coalition under some circumstances, but detrimental for the player to do so under some other circumstances. The nature of those circumstances may be obscure when the player is required to take actions, but can be private information of the potential ally. In warfares, for example, forming an ally with a powerful and aggressive country might be to the benefit of a nation. Allying with a weak and appeasing country, however, might hamper the nation's chance to win. If the cost of investigating into the military power and willingness to fight is prohibitively high, it is reasonable to regard this information as private.

Second, an all-inclusive alliance is undesirable, if not nonexistent. This is a direct application of Riker's (1963) size principle in zero-sum games, in which he suggested that the winning coalition should be minimal. In addition to the convenience of such an assumption, the preference towards smaller alliance sizes is ubiquitous among strategic interactions. A coalition of all political parties existed shortly in post WWII western Germany, but due to the common fear of communist ideology invasion and subtle dissatisfaction against the US military control. The outside players made it difficult to argue the coalition is indeed all-inclusive. In the election context, for another example, a larger coalition size cannot only create a potential conflict of interest, but also a delusion of bargaining power for the major players compared to the alternative of being in a small cabinet that nonetheless secures the majority.

Third, the player with private information can communicate it with other players at no cost, but this communication can only be done in public. For simplicity, I assume that talks are cheap and in public, but this game can be examined when messages are costly or in private. Still, there can be real-life restrictions for unilateral cheap talk. A double agent will make efforts to keep documents classified or agreements secret in vain. Alternatively, for corporate strategies, anti-trust regulations can also mandate disclosure of negotiations. Despite practical issues, the credibility of unilateral cheap talk can also be a concern when there is downside in joining a coalition. Moreover, private cheap talks are often no more revealing than public cheap talks. When receiving a favorable message, the possibility of the cheap talker's being a "good" type and that she happen to choose to convince that particular listener should be balanced by the probability that the cheap talker is a "bad" type and tries to deceive both listeners. It would be even more suspicious when the talker reveals to you privately that she is a "bad" type, since she would only have the interest to convince you into believing so when she is in fact a "good" type and want to leave you out of the coalition.

2.1.1 Timeline

The game I study is a standard dynamic game with incomplete information, and its time line is illustrated as follows:

1. Nature chooses Agent 1's type, t_1, from the type space $\{a, n\}$, where a denotes that Agent 1 is of the aggressive type, meaning that she has innate belligerence and military power to initiate and gain victory in wars. n denotes that Agent 1 is not aggressive. Moreover, Agent 1 is aggressive with probability λ, and not aggressive with probability $1 - \lambda$. The probability distribution of Agent 1's type is a common prior to all nations.

2. Agent 1 learns her type, which is her private information.[1] She can then send a message, m, to both Agent 2 and 3 on her type. Her message $m \in \{A, N\}$, where A denotes a claim Agent 1 makes that she is the aggressive type, and N denotes Agent 1's claim that she is the non-aggressive type.

3. After receiving Agent 1's message, Agent 2 and 3 decides whether to join Agent 1's side. Each of them chooses from the action set $\{J, D\}$, where J denotes the decision of joining Agent 1, and D denotes the decision of not joining.

2.1.2 Payoff

The payoff matrices for the underlying game is shown in Table 1 and 2. The argument of payoff vectors denotes payoff to Agent 1, 2 and 3, respectively. Since Agent 1's messages do not have a direct bearing on the payoff outcome, her actions are omitted from the table.

If Agent 1 is type a, or "aggressive", then she would like to have a war with other

[1] I use feminine pronoun to refer to Agent 1, and masculine pronounce for Agent 2 and Agent 3 hereafter. I also use "Agent 1" and "the sender" interchangeably, and use "Agent 2 and 3", "the listeners" and "the receivers" interchangeably throughout the discussion.

countries, and it would be favorable to form a coalition with her. A *successful alliance* is said to be formed if Agent 1 can attract one and only one other nation to join her side and expropriate the nation that does not join them, which is defined formally at the end of this chapter after notations are introduced. Every member of the coalition gets β from the nation left out, where $\beta > 0$. If neither Agent 2 nor Agent 3 chooses to join, then no such coalition can be formed and all players get zero payoff. To the other extreme, if both Agent 2 and Agent 3 choose to join, then all nations would be on the same side and therefore loses the target for expropriation. A natural result of this zero-sum game is to have all players get zero payoff in this case.

If Agent 1 is type n, or "non-aggressive", then she would prefer to remain in peace with other countries. Every other player that joins the alliance would be exploited by Agent 1. A cost $\alpha (> 0)$ is incurred as a transfer from the party joined to Agent 1. The magnitude of α in relation to β carry different meanings. On one hand, when α is only a small fraction of β, then the threat of having an unfavorable ally posits only a minor problem compared to the benefit of successfully expropriating a third party in a desirable alliance. The downside of joining a wrong coalition can be interpreted as time and transaction cost incurred when making the deal, or some national defense assistance that a country may provide to a friend country to safeguard the alliance. On the other hand, if α is some significant multiple of β, the outcome of allying with a bad type becomes unbearable. The risk of being predated by one's own ally creates a major concern for a decision maker. An interesting parallel may take place in the marriage market, in where a pessimist may argue that the misery of marrying a bad person is so devastating that the gains from happy marriage seem not worth pursuing.

From the payoff matrices, it can be easily verified that if the game is static with complete information, Agent 2 and 3 both has a (strictly) dominant strategy of playing J if $t_1 = a$, while the dominant strategy is D if $t_1 = n$. This payoff structure gives rise to

		Agent 3	
		Join	Don't Join
Agent 2	Join	(0,0,0)	(β, β, -2β)
	Don't Join	(β, -2β, β)	(0, 0, 0)

Table 2.1: Payoff Matrix If $t_1 = a$ (Prob.$= \lambda$)

		Agent 3	
		Join	Don't Join
Agent 2	Join	(2α, -α, -α)	(α, -α, 0)
	Don't Join	(α, 0, -α)	(0, 0, 0)

Table 2.2: Payoff Matrix If $t_1 = n$ (Prob.$= 1 - \lambda$)

a unique Nash Equilibrium for the static game with complete information: Agent 2 and 3 would both play J in the upper game, and D in the lower game. Therefore, Agent 1's coalition would either have size 1 or 3, and winning coalition does not exist in equilibrium.

2.1.3 Notations

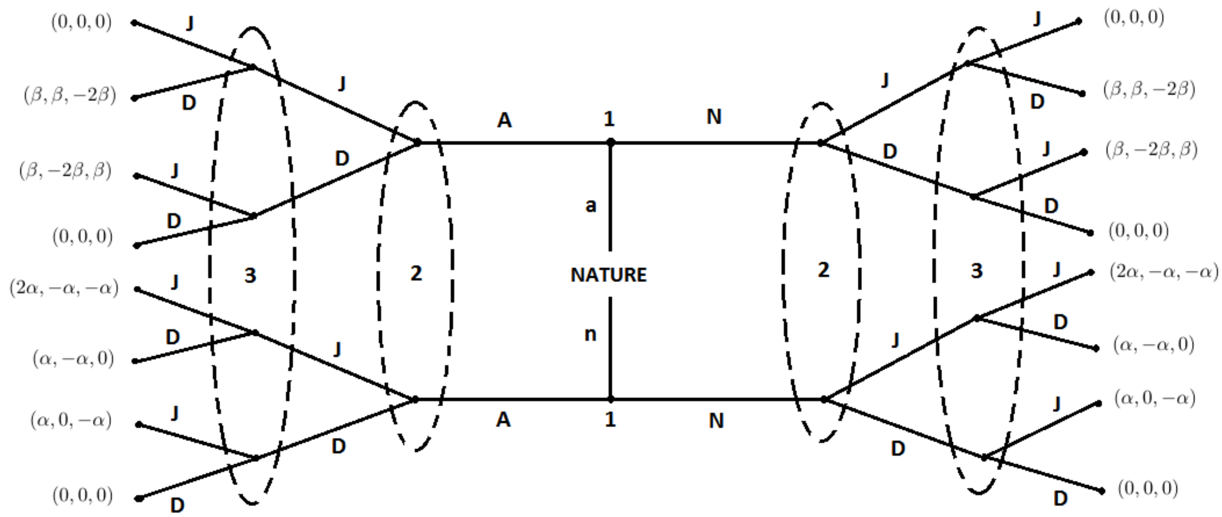
Let T be Agent 1's type space, where $T_1 = \{a, n\}$. $t_1 \in T_1$, then, is a random variable Nature picks form a Bernoulli distribution with parameter λ, which is common knowledge. Agent 1's message m maps her type, $t_1 \in T_1$, to an element in her message space, $M = \{A, N\}$. Hence, *ex ante*, Agent 1's pure strategy $s_1 : T \to M$, is defined by $s_1 = (m(a), m(n))$. Agent 1 can also have behavior strategy $\sigma_1 = (p(m = a|A), p(m = a|N))$. Agent 2 and 3, after receiving Agent 1's message, assign beliefs to their positions in their information space. Denote $b_i(m)$ as Agent i's belief that Agent 1 is type a, where $i = 2, 3$. Agent 2 and 3's pure strategies are denoted as order pairs $s_i = (r_i(A), r_i(N))$, where $r_i(x)$ denotes Agent i's (re)action when they receive a message x, $i = 2, 3$. Finally, let $S = (s_1, s_2, s_3)$ be the Cartesian product of strategies played by Agent 1, 2 and 3, respectively, and $s \in S$. Then $\pi_i(t_1, s)$ would be the payoff incurred to Agent i when Agent 1's type is t_1 and the strategy played jointly by the three players is s. Denote the payoff vector $\Pi = (\pi_1, \pi_2, \pi_3)$.

This chapter focuses on two different stories, contingent on the realized type of Agent 1. In the first case, when Agent 1 is the aggressive type, her interest is partly aligned with either one of the listeners. From now until Section 3.2, I explore what the aggressive type talker can do to achieve positive payoff. While the talker is the non-aggressive type, her role becomes a predator, or a free-rider, and I examine in Section 3.3 how she could misrepresent her private information and induce both listeners to her trap.

Definition 1. *A successful alliance is said to be formed when the "aggressive" type Agent 1 can attract exactly one other Agent to join and to jointly exploit the Agent that does not join. Formally, there is said to be a successful alliance if i.)* $t_1 = a$*,* $r_2(m(a)) = J$ *and* $r_3(m(a)) = D$*, or ii.)* $t_1 = a$*,* $r_2(m(a)) = D$ *and* $r_3(m(a)) = J$*.*

2.2 Baseline Case: Perfectly Rational Players

In this section, I consider the dynamic game with incomplete information when all players are perfectly rational. In particular, I focus on pure strategy Perfect Bayesian Equilibria (PBE) and semi-separating PBEs where Agent 2 and Agent 3 both play pure strategies. The extensive form game is illustrated below.



Proposition 1.A. *When all players are perfectly rational, any pure strategy PBE requires* $s_2 = s_3$.

Proof. Consider separating equilibria first. Suppose there exist a separating equilibrium of this game. Then on the equilibrium path, Agent 2 and Agent 3 will learn Agent 1's type from her message. Therefore, based on the dominant strategies for the static game with complete information, Agent 2 and 3 will both play J if $t_1 = a$, and play D if $t_1 = n$. Given Agent 2 and 3's actions and beliefs on the equilibrium path, Agent 1, however, would have an incentive to deviate and send the opposite message when her type is n, since her payoff would increase from 0 to 2α by doing so. That contradicts the definition of PBE. Hence, there does not exist a pure strategy separating PBE in this game.

For pooling equilibria, on the equilibrium path, Agent 2 and Agent 3's belief in their position of the game is consistent with the common prior. Therefore, on the equilibrium path, if Agent 2 plays J, then Agent 3 would play J if and only if $\lambda\beta + (1-\lambda)(-\alpha) \geq \lambda(-2\beta) + (1-\lambda)0$, or $\lambda \geq \frac{\alpha}{\alpha+2\beta}$. By symmetry, Agent 2 would not want to deviate under the same parametric configuration. There are two different categories of strategies to sustain this PBE. The first scenario is that Agent 2 and 3 both assign significant amount of belief that Agent 1's type is n off the equilibrium path. For example, let their beliefs, $b_2 = b_3 = 0$ when Agent 2 and 3 receive the off-equilibrium-path message. Then both nations would play D off equilibrium path, and Agent 1 would have no incentive to deviate and get zero payoff. Therefore, when $\lambda \geq \frac{\alpha}{\alpha+2\beta}$, there exist a pure strategy PBE where $S = ((A,A),(J,D),(J,D))$, $b_2 = b_3 = \lambda$ if $m = A$, and $b_2 = b_3 = 0$ if $m = N$. Equivalently, $S = ((N,N),(D,J),(D,J))$, $b_2 = b_3 = 0$ if $m = A$, and $b_2 = b_3 = \lambda$ if $m = n$, is also a PBE. In another case, Agent 2 and 3 can rationally ignore Agent 1's message, stick to their prior, and both play J off equilibrium path. If this is the case, Agent 1 would have no incentive to deviate since she gets the same payoff by sending either message. Her talk is cheap by all means. The associated PBE can therefore be $S = ((A,A),(J,J),(J,J))$, $b_2 = b_3 = \lambda$ for both m, or $S = ((N,N),(J,J),(J,J))$, $b_2 = b_3 = \lambda$ for both m.

Similarly, when Agent 2 plays D, then Agent 3 would play D if and only if $\lambda\beta + (1-\lambda)(-\alpha) \leq \lambda\cdot 0 + (1-\lambda)0$, or $\lambda \leq \frac{\alpha}{\alpha+\beta}$. By the same argument, when $\lambda \leq \frac{\alpha}{\alpha+\beta}$, there exist a pure strategy PBE where $S = ((A,A),(D,D),(D,D))$, $b_2 = b_3 = \lambda$ if $m = A$, and $b_2 = b_3 = 0$ if $m = N$. Or, equivalently, $S = ((N,N),(D,D),(D,D))$, $b_2 = b_3 = 0$ if $m = A$, and $b_2 = b_3 = \lambda$ if $m = n$. Moreover, $S = ((A,A),(D,J),(D,J))$ or $S = ((N,N),(J,D),(J,D))$ cannot be equilibrium under any belief, since type n nation would always want to deviate to the off-the-path message under those strategy profiles.

Those are the only pure strategy pooling PBEs. [2] To see this, it is obvious that every case when Agent 2 and 3 plays symmetric strategies has been discussed in above paragraphs.

[2] I follow the convention to discuss the uniqueness of equilibrium up to a strategy choice. Multiplicity in beliefs is treated as the same equilibrium as long as they as sustaining the same strategy.

Now suppose there exists a pure strategy pooling PBE that Agent 2 and 3 choose different actions on the equilibrium path, then it must be the case that one of them plays J, and the other plays D. By the analysis above, the parameters must satisfy $\lambda \leq \frac{\alpha}{\alpha+2\beta}$ and $\lambda \geq \frac{\alpha}{\alpha+\beta}$, which leads to a contradiction as $\beta > 0$ by assumption. Therefore, the only pure strategy pooling PBE are those when both listeners try to match the action of each other on the equilibrium path. Next, assume Agent 2 and 3 choose different off-the-path strategies in a pooling equilibrium. Without loss of generality, assume that Agent 2 plays J and Agent 3 plays D when the off-the-path message m' is sent. Let the on-the-path message be m, then the type a Agent 1 would always want to deviate since $\beta = \pi_1(m', J, D, a) > \pi_1(m, J, J, a) = 0$ and $\beta = \pi_1(m', J, D, a) > \pi_1(m, D, D, a) = 0$, which is a contradiction. □

In fact, a symmetry in the two listeners' strategies is not only required for PBEs where all player play pure strategies, but also for semi-separating equilibria where Agent 2 and 3 play pure strategies. I state this result in the following proposition.

Proposition 1.B. *When all players are perfectly rational, any PBE in which Agent 2 and 3 both play pure strategies (and Agent 1 can play mixed strategies) requires $s_2 = s_3$.*

Proof. For pooling and separating equilibria, the conclusion is proved in Proposition 1.

For semi-separating equilibria, it can only be that the listeners can learn that Agent 1's type is exactly a (or n) under one message, but does not know her type for sure under another message.

Consider Agent 1's strategy profile: play $m(a) = A$, and randomize if $t_1 = n$. Therefore, Agent 2 and 3 will have posterior belief $b_i(N) = p(t_1 = a|m = N) = 0, i = 2, 3$ when they receive N, and both will play D accordingly. For type n Agent 1 to be indifferent between playing A and N, those two actions must yield the same payoff to her in equilibrium. Therefore, $\pi_1(A, r_2(A), r_3(A), n) = \pi_1(N, D, D, n) = 0$, which implies $r_2(A) = r_3(A) = D$.

To justify Agent 2 and 3's responses, both players must have beliefs $b_i(A) = p(t_1 = a|m = A) \leq \frac{\alpha}{\alpha+\beta}, i = 2, 3$. Thus, for the listeners' beliefs to be consistent with sender's strategy, Agent 1's behavior strategy must satisfy $\frac{\lambda \cdot 1}{\lambda \cdot 1 + (1-\lambda)p(m=A|n)} = b_2(A) = b_3(A) \leq \frac{\alpha}{\alpha+\beta}$, or $p(m = A|n) \geq \frac{\lambda\beta}{(1-\lambda)\alpha}$. Hence, when $\frac{\lambda\beta}{(1-\lambda)\alpha} < 1$, there exist a semi-separating equilibrium that $(\sigma_1, s_2, s_3) = ((1, p(m = A|n)), (D, D), (D, D))$, where $p(m = A|n) \geq \frac{\lambda\beta}{(1-\lambda)\alpha}$, $b_2(A) = b_3(A) = \frac{\lambda \cdot 1}{\lambda \cdot 1 + (1-\lambda)p(m=A|n)}$ and $b_2(N) = b_3(N) = 0$. Another equivalent semi-separating PBE is that Agent 1 plays $m(a) = N$, and randomize if $t_1 = n$. Since her message does not directly affect the payoff vector, all Agent 2 and 3 need to do is to invert her message and to form beliefs that $p(t_1 = a|m = A) = 0$. The rest of the analysis is essentially the same as above.

Next, consider when Agent 1 only sends one message for certain if she is type n, and randomize if she is type a. Without loss of generality, assume $m(n) = A$. Suppose such an equilibrium exists, then Agent 2 and 3 will know immediately that $t_1 = a$ after receiving N, and they would respond by playing J. There are only two set payoff equivalent strategies for Agent 2 and 3 to play in order to make type a Agent 1 indifferent: $(s_2, s_3) = ((J, J), (J, J))$ or $((D, J), (D, J))$. The latter cannot be a PBE, since type n Agent 1 would then deviate to send message N and induced the receivers to play "join". When $\frac{(1-\lambda)\alpha}{2\lambda\beta} < 1$, it can be shown that the only semi-separating PBE in this case is that $(\sigma_1, s_2, s_3) = ((p(m = A|a), 1), (J, J), (J, J))$, where $p(m = A|a) \geq \frac{(1-\lambda)\alpha}{2\lambda\beta}$, $b_2(A) = b_3(A) = p(t_1 = a|A) = \frac{\lambda p(m=A|a)}{\lambda p(m=A|a)+(1-\lambda)}$, $b_2(N) = b_3(N) = 1$. Same analysis applies if $m(n) = N$.

Therefore, we know $s_2 = s_3$ in every semi-separating PBE when both listeners are playing pure strategies. □

The solution to the benchmark model suggests that Agent 2 and Agent 3 would always like to mimic the strategy of each other, not only on-the-path, but also off-the-path, in any PBE when both of them are playing pure strategies. Moreover, the Perfect Bayesian

Equilibrium usually renders unique predictions. For example, among pure strategy pooling PBEs, we know that both listener will play J upon receiving the on-the-path message if $\lambda > \frac{\alpha}{\alpha+\beta}$, and will play D upon receiving the on-the-path message $\lambda < \frac{\alpha}{\alpha+2\beta}$. Multiplicity of equilibria only exist when $\frac{\alpha}{\alpha+2\beta} \geq \lambda \geq \frac{\alpha}{\alpha+\beta}$, and the two receivers would response differently only due to coordination failures. I state an implication of the propositions in this section in the lemma below.

Lemma 1. *When all players are perfectly rational, there does not exist a Perfect Bayesian Equilibrium in which a successful alliance can be formed.*

Agent 1's incapability of forming a coalition, however, does not suggest that the equilibrium payoff would always be $(0, 0, 0)$. In fact, when $\lambda \geq \frac{\alpha}{\alpha+2\beta}$, the pooling equilibrium solution suggests that both listener will play J. Type n sender will get 2α, and Agent 2 and 3 both have expected payoff $(1-\lambda) \cdot (-\alpha) < 0$. Agent 1 would have an expected payoff $2(1-\lambda)\alpha$, *ex ante*. From a more symmetric point of view, Agent 2 and 3 can also form coalitions if they can commit to playing D whatever message they receive from Agent 1. By doing so, the coalition can always receive its maximum possible payoff sum, 0, regardless of Agent 1's type. Unfortunately, for both listeners, when the prior probability that $t_1 = a$ is high enough, or $\lambda \geq \frac{\alpha}{\alpha+2\beta}$, the expected benefit for of breaking the commitment and playing J, the dominant strategy when $t_1 = a$, would outweigh the risk of being expropriated by the type n ally. Therefore, Agent 2 and 3 cannot form such coalition in general. The game becomes a prisoner's dilemma for perfectly rational Agent 2 and 3 in a broader sense.

2.3 Bounded Rational Players

From this chapter on, I relax the assumption of perfect rationality to examine the possibility of Agent 1's forming a coalition with another player. In particular, I follow Crawford's

(2003) treatment of bounded rationality, and let the bounded rational message sender to have a probability of being a “truth-teller", who always sends a message about her true type, and a probability of being an “inverter", who always sends a false message about her type. Similarly, the bounded rational message receiver will have a probability of being a “believer", who always believes whatever the sender says, and a probability of being an “inverter", who always believes the opposite of whatever the sender says.

In this chapter, I allow both Agent 1 and Agent 2 to have a positive probability of being bounded rational, or $Mortal$, while letting Agent 3 to be rational, or *Sophisticated*. Therefore, the sender and one receiver have a probability of being $Mortal$, while at least one listener would always remain *Sophisticated*. I use this treatment to allow asymmetry between the two listeners, and am particularly interested in how a perfectly rational listener will respond if the other listener is $Mortal$.

This paper follows the notations in the Crawford paper: the behavior of Agent 1 (the “sender") population can be summarized by $s_l \equiv \Pr\{\text{Sender (Agent 1) is a } \textit{Liar}\}$, $s_t \equiv \Pr\{\text{Sender is a } \textit{Truth-teller}\}$, and $s_s \equiv \Pr\{\text{Sender is } \textit{Sophisticated}\}$, where $s_l + s_t + s_s = 1$. Similarly, Agent 2 population can be characterized by $r_b \equiv \Pr\{\text{Receiver (Agent 2) is a } \textit{Believer}\}$, $r_i \equiv \Pr\{\text{Receiver is an } Inverter\}$, and $r_s \equiv \Pr\{\text{Receiver is } \textit{Sophisticated}\}$, where $r_b + r_i + r_s = 1$.

Bounded rational players' presence can be reduced to outside structures of a game played by *Sophisticated* players. The behavioral parameters are not meant to exhaust all possible strategies played by those $Mortal$ listener and senders. Instead, it is more interesting to see how the presence of bounded rational players changes the outcome of the game played by perfectly rational players. This is very different from the Crawford paper, where the $Mortal$ player behavior can always be viewed as the best response to the behavior of his or her $Mortal$ opponent, the iteration process can be conducted on several levels and thus sustaining

Skeptical bounded rational players' "strategies". In my paper, a $Believer's$ behavior is indeed the best response when the sender is a $Truth-teller$, and an $Inverter's$ behavior is also the best response to $Liar's$ message. $Mortal$ sender's behavior, however, can be less accurately viewed as optimal strategies in response to the $Mortal$ receiver's behavioral parameters. [3]

As Crawford has noted in his paper, bounded rational player's payoff can be reduced to an outside structure of the game. The reduced normal form games played by *Sophisticated* players are summarized in Table 3 and 4, depending on the realized value of t_1.

Please note that some rows and columns are merged to produce the reduced normal form game, and the order how rows and columns are listed does not match in the two tables. To create a table for Agent 2 and 3's expected payoffs, therefore, one will need to weight the corresponding entries in Table 3 and Table 4, and the expected payoff matrix would have size 4×16, with 64 different entries. Fortunately, solving for PBEs does not require finding saddle point from normal form game, and using extensive form game would generate a clearer picture of the payoff structure.

Values in the normal form games are calculated using Bayesian updating of beliefs. For example, check the case when *Sophisticated* Agent 1 plays (A, A), *Sophisticated* Agent 2 plays (J, D) and Agent 3 plays (D, J). When $t_1 = a$, the equilibrium outcome will be that *Sophisticated* Agent 1 sends A, *Sophisticated* Agent 2 chooses J and Agent 3 chooses D. *Sophisticated* Agent 1 will get β if and only if Agent 2 joins her side, which will occur with probability r_b+r_s, when Agent 2 is a *Believer* and choose to play J after receiving A, or when Agent 2 is *Sophisticated* and plays the specified strategy profile. Hence, *Sophisticated* Agent 1 will have expected payoff $(r_b + r_s)\beta$. *Sophisticated* Agent 2's expected payoff, however,

[3] To avoid confusion in two different sources of private information, I use "type" to refer to t_1, Agent 1's private information of whether being *Aggressive* or *Non-aggressive*, and follow Crawford (2003) by using "behavior parametrics" refer to whether Agent 1 is *Sophisticated*, $Mortal$, *Liar* or *Truth-teller*, and Agent 2's being *Sophisticated*, $Mortal$, *Believer* or $Inverter$ instead.

will not only be a function of Agent 1's behavioral parameters, but also the her type, t_1. Upon receiving the message A and given that *Sophisticated* Agent 1 is playing (A, A), Agent 2 needs to consider three cases: 1.) Agent 1 is *Sophisticated*, which implies that $t_1 = a$; 2.) Agent 1 is a *Truth-teller* and therefore $t_1 = a$; 3.) Agent 1 is a *Liar* and therefore $t_1 = n$. Given that Agent 3 plays D upon receiving A, Agent 2 knows that by playing J, he will get β by playing J if $t_1 = a$, and gets $-\alpha$ if $t_1 = n$. Thus, *Sophisticated* Agent 2's expected payoff would be $\frac{P(t_1=a|m=A)\beta+P(t_1=n|m=A)(-\alpha)}{P(m=A)} = \frac{\lambda(s_s+s_t)\beta+(1-\lambda)s_l(-\alpha)}{\lambda(s_s+s_t)+(1-\lambda)s_l}$. Lastly, *Sophisticated* Agent 3's expected payoff is a function of all pieces of Agent 1 and 2's private information by the same argument. Therefore, he would not only need to consider the three cases the *Sophisticated* Agent 2 is concerned about, but also three possible cases about Agent 2's behavioral identity: 1.) Agent 2 is *Sophisticated* and plays J; 2.) Agent 2 is a *Believer* and plays J; 3.) Agent 2 is *Inverter* and plays D. When *Sophisticated* Agent 3 plays D, he would only get nonzero payoff when $t_1 = a$ and Agent 2 plays J, in which case he gets -2β. Therefore, Agent 3's expected payoff can be written as $P[t_1 = a \cap r_2(A) = J|m = A](-2\beta) = \frac{P[t_1=a \cap r_2(A)=J \cap m=A](-2\beta)}{P(m=A)} = \frac{\lambda(s_s+s_t)\cdot(r_s+r_b)\cdot(-2\beta)}{\lambda(s_s+s_t)+(1-\lambda)s_l}$. To summarize, when $S = ((A, A), (J, D)), (D, J))$ for *Sophisticated* players, their payoff vector is B_1.

The PBEs of the game can be solved using the same technique employed in the last section. An advantage of having such a general form game is that it allows many possible variations: by setting some behavioral parameters to be equal to zero, and others to be positive, one can have different combinations of perfectly rational and bounded rational players interacting with each other. The versatility of the general set up, however, can be a burden for analyzing and intuitively understanding parametric configuration in equilibrium. Hence, in the following chapters, I propose some special cases of the general game and provide solutions under these cases.

Table 2.3: Reduced Normal Form Game When $t_1 = a$

Player 1 \ Player 2	(J,D)	(J,D)	(J,J)	(J,J)	(D,J)	(D,J)	(D,D)	(D,D)
Player 3	(J,D) (J,J)	(D,J) (D,D)	(J,D) (J,J)	(D,J) (D,D)	(J,D) (J,J)	(D,J) (D,D)	(J,D) (J,J)	(D,J) (D,D)
(A,N) (A,A)	A_1	B_1	A_1	B_1	C_1	D_1	C_1	D_1
(N,A) (N,N)	E_1	F_1	E_1	F_1	G_1	H_1	G_1	H_1

Table 2.4: Reduced Normal Form Game When $t_1 = n$

Player 1 \ Player 2	(J,D)	(J,D)	(J,J)	(J,J)	(D,J)	(D,J)	(D,D)	(D,D)
Player 3	(J,D) (D,D)	(D,J) (J,J)	(J,D) (D,D)	(D,J) (J,J)	(J,D) (D,D)	(D,J) (J,J)	(J,D) (D,D)	(D,J) (J,J)
(A,N) (N,N)	H_2	G_2	F_2	E_2	F_2	E_2	H_2	G_2
(A,A) (N,A)	D_2	C_2	B_2	A_2	B_2	A_2	D_2	C_2

$$A_1 = \begin{pmatrix} r_i \cdot \beta \\ \frac{(1-\lambda)s_l(-\alpha)}{\lambda(s_s+s_t)+(1-\lambda)s_l} \\ \frac{r_i\lambda(s_s+s_t)\cdot\beta+(1-\lambda)s_l(-\alpha)}{\lambda(s_s+s_t)+(1-\lambda)s_l} \end{pmatrix} \quad B_1 = \begin{pmatrix} (r_b+r_s)\cdot\beta \\ \frac{\lambda(s_s+s_t)\beta+(1-\lambda)s_l(-\alpha)}{\lambda(s_s+s_t)+(1-\lambda)s_l} \\ \frac{(r_s+r_b)\lambda(s_s+s_t)(-2\beta)}{\lambda(s_s+s_t)+(1-\lambda)s_l} \end{pmatrix}$$

$$C_1 = \begin{pmatrix} (r_s+r_i)\beta \\ \frac{\lambda(s_s+s_t)(-2\beta)}{\lambda(s_s+s_t)+(1-\lambda)s_l} \\ \frac{[\lambda(r_s+r_i)(s_s+s_t)\beta+s_l(-\alpha)(1-\lambda)]}{\lambda(s_s+s_t)+(1-\lambda)s_l} \end{pmatrix} \quad D_1 = \begin{pmatrix} r_b\cdot\beta \\ 0 \\ \frac{r_b\cdot\lambda(s_s+s_t)(-2\beta)}{\lambda(s_s+s_t)+(1-\lambda)s_l} \end{pmatrix}$$

$$E_1 = \begin{pmatrix} r_b \cdot \beta \\ \frac{(1-\lambda)s_t(-\alpha)}{\lambda(s_s+s_l)+(1-\lambda)s_t} \\ \frac{r_b\lambda(s_s+s_l)\cdot\beta+(1-\lambda)s_t(-\alpha)}{\lambda(s_s+s_l)+(1-\lambda)s_t} \end{pmatrix} \quad F_1 = \begin{pmatrix} (r_i+r_s)\cdot\beta \\ \frac{(1-\lambda)s_t(-\alpha)+\lambda(s_s+s_l)\beta}{\lambda(s_s+s_l)+(1-\lambda)s_t} \\ \frac{(r_s+r_i)\lambda(s_s+s_l)(-2\beta)}{\lambda(s_s+s_l)+(1-\lambda)s_t} \end{pmatrix}$$

$$G_1 = \begin{pmatrix} (r_b+r_s)\cdot\beta \\ \frac{\lambda(s_s+s_l)(-2\beta)}{\lambda(s_s+s_l)+(1-\lambda)s_t} \\ \frac{[\lambda(r_b+r_s)(s_s+s_l)\beta+(1-\lambda)s_t(-\alpha)]}{\lambda(s_s+s_l)+(1-\lambda)s_t} \end{pmatrix} \quad H_1 = \begin{pmatrix} r_i\beta \\ 0 \\ \frac{r_i\cdot\lambda(s_l+s_s)(-2\beta)}{\lambda(s_s+s_l)+(1-\lambda)s_t} \end{pmatrix}$$

$$A_2 = \begin{pmatrix} (r_b+r_s+1)\alpha \\ \frac{(1-\lambda)(s_s+s_l)(-\alpha)}{\lambda s_t+(1-\lambda)(s_s+s_l)} \\ \frac{(1-\lambda)(s_s+s_l)(-\alpha)+\lambda s_t r_b\beta}{\lambda s_t+(1-\lambda)(s_s+s_l)} \end{pmatrix} \quad B_2 = \begin{pmatrix} (r_b+r_s)\alpha \\ \frac{(1-\lambda)(s_s+s_l)(-\alpha)+\lambda s_t\beta}{\lambda s_t+(1-\lambda)(s_s+s_l)} \\ \frac{\lambda s_t(r_b+r_s)(-2\beta)}{\lambda s_t+(1-\lambda)(s_s+s_l)} \end{pmatrix}$$

$$C_2 = \begin{pmatrix} (1+r_b)\alpha \\ \frac{\lambda s_t(-2\beta)}{\lambda s_t+(1-\lambda)(s_s+s_l)} \\ \frac{[\lambda s_t(r_s+r_i)\beta+(1-\lambda)(s_s+s_l)(-\alpha)]}{\lambda s_t+(1-\lambda)(s_s+s_l)} \end{pmatrix} \quad D_2 = \begin{pmatrix} r_b\alpha \\ 0 \\ \frac{\lambda s_t r_b(-2\beta)}{\lambda s_t+(1-\lambda)(s_s+s_l)} \end{pmatrix}$$

$$E_2 = \begin{pmatrix} (r_i+r_s+1)\alpha \\ \frac{(1-\lambda)(s_s+s_t)(-\alpha)}{\lambda s_l+(1-\lambda)(s_s+s_t)} \\ \frac{(1-\lambda)(s_s+s_t)(-\alpha)+\lambda s_l r_b\beta}{\lambda s_l+(1-\lambda)(s_s+s_t)} \end{pmatrix} \quad F_2 = \begin{pmatrix} (r_i+r_s)\alpha \\ \frac{(1-\lambda)(s_s+s_t)(-\alpha)+\lambda s_l\beta}{\lambda s_l+(1-\lambda)(s_s+s_t)} \\ \frac{\lambda s_l(r_s+r_i)(-2\beta)}{\lambda s_l+(1-\lambda)(s_s+s_t)} \end{pmatrix}$$

$$G_2 = \begin{pmatrix} (1+r_i)\alpha \\ \frac{\lambda s_l(-2\beta)}{\lambda s_l+(1-\lambda)(s_s+s_t)} \\ \frac{[(1-\lambda)(s_s+s_t)(-\alpha)+\lambda s_l(r_b+r_s)\beta]}{\lambda s_l+(1-\lambda)(s_s+s_t)} \end{pmatrix} \quad H_2 = \begin{pmatrix} r_i\alpha \\ 0 \\ \frac{\lambda s_l r_i(-2\beta)}{\lambda s_l+(1-\lambda)(s_s+s_t)} \end{pmatrix}$$

2.3.1 The Sender Is Sophisticated and One Listener Is Mortal

Let $s_s = 1$ and $r_b + r_i = 1$, then Agent 1 is *Sophisticated*, and Agent 2 is $Mortal$. This treatment can be linked to persuasion in real life. In political elections, for example, a candidate may want to ally with some group of voters. Suppose there is a group of bounded rational voter, who naively choose to believe or invert whatever the candidate says without carefully gauging the objective probabilities or considering monetary values of payoff incurred under every type, then both the *Sophisticated* sender and receiver would have to anticipate the behavior of such $Mortal$ players and adjust their strategies accordingly.

The payoff matrices with $s_s = 1$ and $r_s = 0$ are summarized in Table 5 and 6, which are special cases of Table 3 and 4. To reduce the number of unnecessary parameters, I substitute r_i with $1 - r_b$ throughout the analysis.

With $s_s = 1$ and $r_s = 0$, this game becomes a standard two-player signaling game. Since $Mortal$ Agent 2's action can be viewed as choosing his strategy from a singleton set predetermined by his behavioral parameters, he is reduced to an outside structure of the game. Also, as in the general formed game, Agent 1's message is no longer cheap talk. The presence of $Believers$ and $Inverters$ transforms her word directly into responses of some Agent 2 behavioral parameters, thus changing the expected payoff of the two *Sophisticated* players, Agent 1 and 3. Please note that Agent 3 still have (strictly) dominant strategies of playing J if $t_1 = a$, and playing D is $t_1 = n$. Like in the first section, I solve PBEs where Agent 3 plays pure strategies.

		Agent 3 J	Agent 3 D
Agent 1	A	$((1-r_b)\beta,\ (1-r_b)(\text{-}2\beta),\ (1-r_b)\beta)$	$(r_b\beta,\ r_b\beta,\ r_b(\text{-}2\beta))$
	N	$(r_b\beta,\ r_b(\text{-}2\beta),\ r_b\beta)$	$((1-r_b)\beta,\ (1-r_b)\beta,\ (1-r_b)(\text{-}2\beta))$

Table 2.5: expected payoff matrix when $t_1 = a$

		Agent 3 J	Agent 3 D
Agent 1	A	$((1+r_b)\alpha,\ -r_b\alpha,\ -\alpha)$	$(r_b\alpha,\ -r_b\alpha,\ 0)$
	N	$((2-r_b)\alpha,\ -(1-r_b)\alpha,\ -\alpha)$	$((1-r_b)\alpha,\ -(1-r_b)\alpha,\ 0)$

Table 2.6: expected payoff matrix when $t_1 = n$

Proposition 2. *When $s_s = 1$ and $r_s = 0$, there exist separating PBEs when Agent 1 and 3 can form a successful alliance. Moreover, Agent 3's expected payoff is weakly increasing with Agent 1's expected payoff in such equilibria.*

Proof. In separating equilibria, Agent 3 can always tell t_1 from the message sent by Agent 1. Given Agent 3's dominant strategies, Agent 1 can determine her optimal strategies through backward induction. Therefore, a PBE will sustain as long as Agent 1 is sending different messages given different types, and both have no incentive to deviate.

Knowing that Agent 3 must learn the true t_1 in separating PBE, Agent 1 understands that Agent 3 will play J if $t_1 = a$, and play D if $t_1 = n$. She chooses from two separating equilibria $S = (s_1, s_3) = ((A, N), (J, D))$ and $S = (s_1, s_3) = ((N, A), (D, J))$ conditional on her type t_1. Note that $\pi_1((A, N), (J, D)|t_1 = a) = (1 - r_b)\beta \geq r_b\beta = \pi_1((N, A), (D, J)|t_1 = a)$ if and only if $r_b \leq 0.5$, and $\pi_1((A, N), (J, D)|t_1 = n) = (1-r_b)\alpha > r_b\beta = \pi_1((N, A), (D, J)|t_1 = n)$ if and only if $r_b < 0.5$. Thus, Player 1 will play (A, N) if and only if $r_b \leq 0.5$, and will play (A, N) if and only if $r_b > 0.5$. These are all the separating equilibrium by definition of dominant strategy and Agent 3's beliefs.

i). If $r_b \leq 0.5$, then Agent 1 always plays (A, N) and Agent 3 always plays (J, D). Agent 1 gets $(1 - r_b)\beta$ when $t_1 = a$, and $(1 - r_b)\alpha$ when $t_1 = n$. Agent 3 gets gets $(1 - r_b)\beta$

when $t_1 = a$, and gets 0 when $t_1 = n$. Thus, $\frac{d\pi_3}{d\pi_1} = 1$ when $t_1 = a$, and $\frac{d\pi_3}{d\pi_1} = 0$ when $t_1 = n$.

ii). If $r_b > 0.5$, then Agent 1 will always plays (N, A) and Agent 3 always plays (D, J). Agent 1 gets $r_b\beta$ when $t_1 = a$, and $r_b\alpha$ when $t_1 = n$. Agent 3 gets $r_b\beta$ when $t_1 = a$, and gets 0 when $t_1 = n$. Thus, $\frac{d\pi_3}{d\pi_1} = 1$ when $t_1 = a$, and $\frac{d\pi_3}{d\pi_1} = 0$ when $t_1 = n$.

Therefore, in separating equilibria, when $t_1 = a$ and $m(a) = A$, Agent 3 will always join and the *Inverter* Agent 2 will not join, the successful alliance will be formed with probability r_i. Likewise, when $t_1 = a$ and $m(a) = N$, Agent 3 will still join and the *Believer* Agent 2 will not join, the successful alliance will be formed with probability r_b. Since Agent 3's expected payoff is weakly increasing with Agent 1's expected payoff under each type, it implies that, *ex ante*, Agent 3's expected payoff weighted by the probability of possible values of t_1 will also be weakly increasing with Agent 1's expected payoff by first order approximation. □

Proposition 3. *When $s_s = 1$ and $r_s = 0$, there does not exist a successful alliance of Agent 1 and 2.*

Proof. This proposition follows immediately from the fact that Agent 3 will always recognize her type and choose to join whenever $t_1 = a$. Hence, there is no way that Agent 1 and Agent 2 can exclude and expropriate Agent 3 when $t_1 = a$. Thus, there cannot be a successful alliance by definition. □

The analysis for separating equilibria with a *Sophisticated* sender, a *Mortal* receiver and a *Sophisticated* receiver is very intuitive: since the *Sophisticated* sender must reveal her true type to *Sophisticated* receivers in a separating equilibrium, she loses the opportunity to expropriate the *Sophisticated* receiver under any circumstance. This communication can be to the advantage of Agent 1, however, if she tries to communicate her type when she is indeed "aggressive", and Agent 3 would join her ally and they can bully Agent 2 together. Their interests are aligned through joint expropriation of Agent 2. Meanwhile, Agent 3 evades the downside of joining the coalition completely, because by definition of separating equilibrium, he is *Sophisticated* enough to tell when Agent 1 is not really “aggressive".

Another interesting observation from the separating equilibria is that the $Mortal$ receiver's expected payoff decreases when r_b deviates from 0.5 to a greater extent. That is, the $Mortal$ receiver is worse off when he is more “biased". When $r_b \leq 0.5$, $(s_1, s_3) = ((A, N), (J, D))$, Agent 2's expected payoff is $\lambda(1 - r_b)(-2\beta) + (1 - \lambda)(1 - r_b)(-\alpha)$. When $r_b > 0.5$, $(s_1, s_3) = ((N, A), (D, J))$, Agent 2's expected payoff is $\lambda \cdot r_b(-2\beta) + (1-\lambda)r_b(-\alpha)$. Since $\lambda \cdot r_b(-2\beta) + (1 - \lambda)r_b(-\alpha) < \lambda(1 - r_b)(-2\beta) + (1 - \lambda)(1 - r_b)(-\alpha)$ if and only if $r_b > 0.5$, Agent 2's expected payoff can be written as $min\{\lambda \cdot r_b(-2\beta) + (1-\lambda)r_b(-\alpha), \lambda(1-r_b)(-2\beta) + (1-\lambda)(1-r_b)(-\alpha)\} = -(|r_b - 0.5| + 0.5) \cdot (2\beta\lambda + \alpha(1-\lambda))$, which is decreasing in $|r_b - 0.5|$. Intuitively, this result suggest that the $Mortal$ listener is better off by remaining neutral between being a $Believer$ and an $Inverter$, if he can help at all. When there either one of the $Believer$ or $Inverter$ population is presented disproportionally, the $Sophisticated$ sender can then take advantage of the $Mortal$ listener's literal-mindedness and induce some of the $Mortal$ listeners to act against his own interests. In the following part of this chapter, I discuss pure strategy pooling PBEs.

Proposition 4. *When $s_s = 1$ and $r_s = 0$, there exist pure strategy pooling PBEs in which Agent 1 and 2 can form a successful alliance.*

Proof. By definition, every pooling PBEs can either one of the two cases: Agent 1 plays (A, A) or plays (N, N).

Case I: Suppose Agent 1 always sends A, then Agent 3 will play “Don't join” on-the-path if and only if $E[\pi_3((A, A), (D, r_3(N))] > E[\pi_3((A, A), (J, r_3(N))]$ for some $r_i(N)$. This condition can be written as $\lambda \cdot r_b(-2\beta) > \lambda \cdot (1 - r_b)\beta + (1 - \lambda) \cdot (-\alpha)$, or $r_b < \frac{(1-\lambda)\cdot\alpha}{\lambda\beta} - 1$. Now, check that Agent 1 has no incentive to deviate. Let Player 3 have belief $b_3 = \lambda$, he will play D regardless the message he receives given the parametric configuration. Then, Agent 1 will indeed send A if and only if $r_b\beta > (1-r_b)\beta$ and $r_b\alpha > (1-r_b)\alpha$, or $r_b > \frac{1}{2}$. Therefore, when $\frac{1}{2} < r_b < \frac{(1-\lambda)\cdot\alpha}{\lambda\beta} - 1$, there exists a pooling PBE $(s_1, s_3) = ((A, A), (D, D))$, $b_3(m) = \lambda$, $m = A, N$.

Case II: Suppose Agent 1 always sends N, then Agent 3 will play "Don't join" on-the-path if and only if $\lambda \cdot (1-r_b)(-2\beta) > \lambda \cdot r_b\beta + (1-\lambda)\cdot(-\alpha)$, or $r_b > 2 - \frac{(1-\lambda)\cdot\alpha}{\lambda\beta}$. Let Player 3 have belief $b_3 = \lambda$, he will play D regardless the message he receives given the parametric configuration. Then, Agent 1 will indeed send N if and only if $r_b < \frac{1}{2}$. Therefore, when $2 - \frac{(1-\lambda)\cdot\alpha}{\lambda\beta} < r_b < \frac{1}{2}$, there exists a pooling PBE $(s_1, s_3) = ((N, N), (D, D))$, $b_3(m) = \lambda$, $m = A, N$.

Thus, when $\frac{1}{2} < r_b < \frac{(1-\lambda)\cdot\alpha}{\lambda\beta} - 1$ or $\frac{1}{2} < r_b < \frac{(1-\lambda)\cdot\alpha}{\lambda\beta} - 1$, Agent 3 will always play D. When $t_1 = a$, Agent 1 have a chance to form coalition with the *Believer* or the *Inverter*, depending on the parametric configuration and the realized *Mortal* listener identity. □

Proposition 5. *When $s_s = 1$ and $r_s = 0$, there exist pure strategy pooling PBEs in which Agent 1 and 3 can form a successful alliance.*

Proof. For a pure strategy pooling PBE in which Agent 1 and 3 can form a successful alliance, Agent 3 must always play J on the path. Now consider the only possible cases: Agent 3 plays J off equilibrium path, or Agent 3 plays D off equilibrium path.

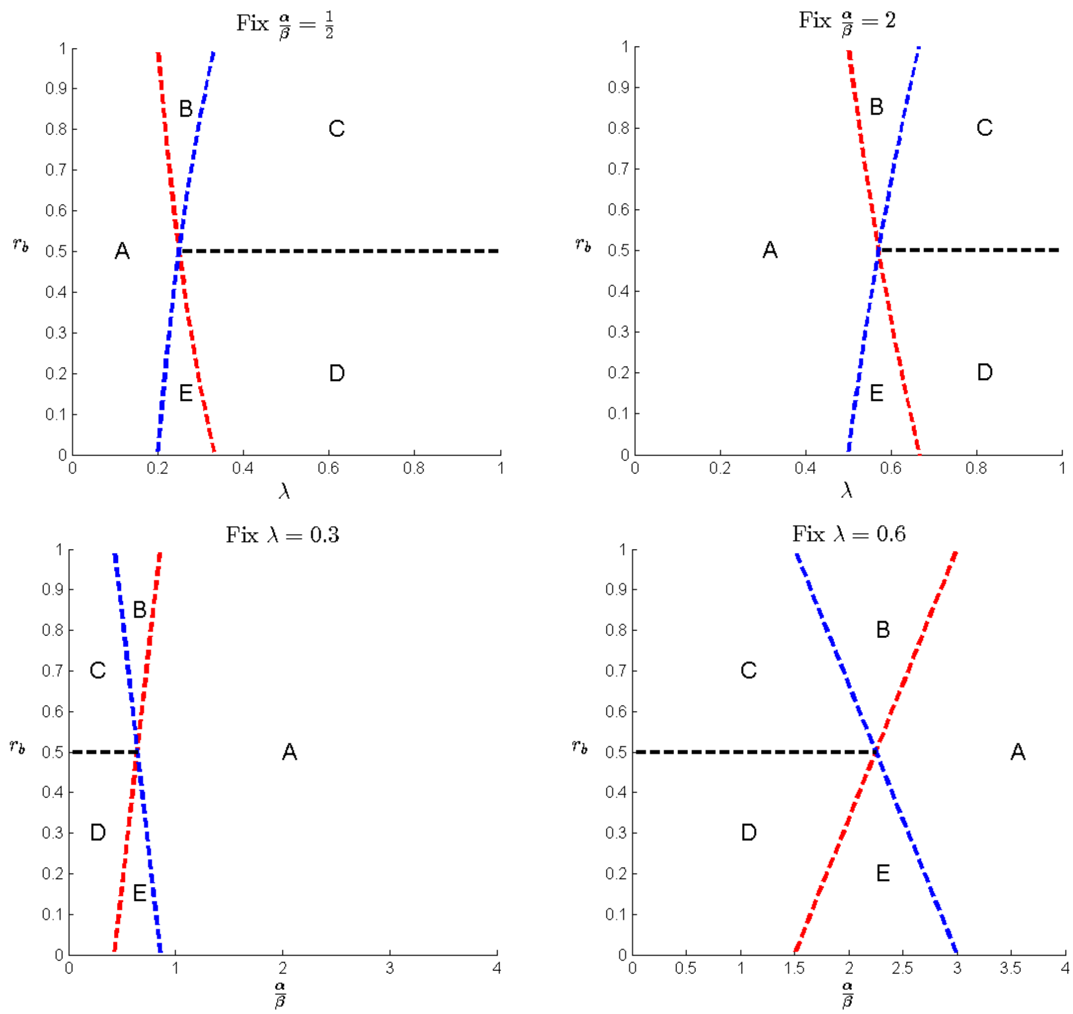
Suppose Agent 3 plays J off equilibrium path. If $r_b < (>)0.5$, then $\pi_1(m = A, t_1 = a) = (1-r_b)\beta > (<)\ r_b\beta = \pi_1(m = N, t_1 = a)$ but $\pi_1(m = A, t_1 = n) = (1+r_b)\beta < (>)\ (2-r_b)\beta = \pi_1(m = N, t_1 = n)$. Thus, type a and type n sender would always want to send different messages, which is incompatible with pooling equilibrium, which requires a player sends the same message regardless of her type.

Now let Agent 3 play D off equilibrium path. Suppose $s_1 = (A, A)$, then on one hand, the equilibrium can be sustained by Agent 3's off-the-path belief about the sender's type if and only if $b_3(N)\pi_3(D, t_1 = a) + (1 - b_3(N))\pi_3(D, t_1 = n) \geq b_3(N)\pi_3(J, t_1 = a) + (1 - b_3(N))\pi_3(J, t_1 = n)$, or $b_3(N) < \frac{\alpha}{(2-r_b)\beta+\alpha}$, which is possible given positive parameters. On the other hand, Agent 1 does not have incentives to deviate if and only if $\pi_1(m = A, r_3(A) = J|t_1) \geq \pi_1(m = N, r_3(N) = D|t_1)$ for both values of t_1. This implies that $(1-r_b)\beta \geq (1-r_b)\beta$ and $(1+r_b)\alpha \geq (1-r_b)\alpha$, which holds for all $r_b \geq 0$. On-the-path, Agent 3 will have no incentive to deviate if and only if $b_3(A)\pi_3(J, t_1 = a)+(1-b_3(A))\pi_3(J, t_1 = n) \geq b_3(A)\pi_3(D, t_1 = a) + (1-b_3(A))\pi_3(D, t_1 = n)$, where $b_3(A)$ must equal λ. This condition can be simplified as $r_b \geq \frac{(1-\lambda)\cdot\alpha}{\lambda\beta} - 1$. Therefore, the pooling equilibrium $s_1 = (A, A)$, $s_3 = (J, D)$, $b_3(N) < \frac{\alpha}{(2-r_b)\beta+\alpha}$ and $b_3(A) = \lambda$ exists if and only if $\frac{(1-\lambda)\cdot\alpha}{\lambda\beta} - 1 \leq r_b$.

Similarly, Suppose $s_1 = (N, N)$, then on one hand, the equilibrium can be sustained by Agent 3's off-the-path belief about the sender's type if and only if $b_3(A)\pi_3(D, t_1 = a)+(1-b_3(A))\pi_3(D, t_1 = n) \geq b_3(A)\pi_3(J, t_1 = a) + (1 - b_3(A))\pi_3(J, t_1 = n)$, or $b_3(A) > \frac{\alpha}{(1+r_b)\beta+\alpha}$, which is possible given positive parameters. On the other hand, Agent 1 does not have incentives to deviate if and only if $\pi_1(m = A, r_3(A) = D|t_1) \leq \pi_1(m = N, r_3(N) = J|t_1)$ for both values of t_1. This implies that $r_b\beta \geq r_b\beta$ and $r_b\alpha \geq (2 - r_b)\alpha$, which holds for all $r_b \leq 1$. On-the-path, Agent 3 will have no incentive to deviate if and only if $b_3(N)\pi_3(J, t_1 = a) + (1 - b_3(N))\pi_3(J, t_1 = n) \geq b_3(N)\pi_3(D, t_1 = a) + (1 - b_3(N))\pi_3(D, t_1 = n)$, where $b_3(N)$ must equal λ. This condition can be simplified as $r_b \leq 2 - \frac{(1-\lambda)\cdot\alpha}{\lambda\beta}$. Therefore, the pooling equilibrium $s_1 = (N, N)$, $s_3 = (D, J)$, $b_3(A) > \frac{\alpha}{(1+r_b)\beta+\alpha}$ and $b_3(n) = \lambda$ exists if and only if $r_b \leq 2 - \frac{(1-\lambda)\cdot\alpha}{\lambda\beta}$. □

The regions for pooling equilibria in which successful alliances can be formed are plotted below. I fix some of the parameters to show the relationship between other ones. Region C indicates where pooling equilibrium $(s_1, s_3) = ((A, A), (D, D))$ can be sustained, and region D indicates where pooling equilibrium $(s_1, s_3) = ((N, N), (D, D))$ can be formed.

Those are the two equilibria in which Agent 1 can form a successful alliance with some Agent 2. In region B+C+D, the pooling equilibrium $(s_1, s_3) = ((A, A), (J, J))$ can rise, and the pooling equilibrium $(s_1, s_3) = ((N, N), (J, J))$ is sustainable if parameters lie in region C+D+E. Those constitutes the pooling equilibria in which Agent 1 can form coalitions with Agent 3.



Please note that Agent 1 can form a coalition with Agent 3 under a wider range of parametric configurations than with Agent 2. Furthermore, whenever the parameters allow a pooling equilibrium in which Agent 1 can collude with Agent 2 to exist, there always exists another pooling equilibrium in which Agent 1 can collude with Agent 3 instead. This result is consistent with the case in pure strategy separating PBEs, in which Agent 1 can only form coalitions with Agent 3. This might help explain why it is more prevalent to observe *Sophisticated* people working with *Sophisticated* people and politicians colluding with multi-billionaires: intuitively, it is more difficult to exploit a *Sophisticated* counterpart than to exploit a simple-minded counterpart in equilibrium.

To summarize, when Agent 1 is *Sophisticated* and Agent 2 is $Mortal$, it is possible to form a successful alliance in both separating and pooling PBEs. Agent 1 can possibly form a successful alliance with a $Sophisticated$ listener under both pooling and separating equilibria, and she can only do so with a $Mortal$ sender in pooling equilibria. Moreover, in pooling equilibria, Agent 1 can always form an alliance with Agent 3 whenever the parameters allow her to do so with Agent 2.

Welfare and Discussions

It can be calculated that even in pooling equilibria where Agent 1 form a *successful coalition* with Agent 3, the relative scale of expected payoff incurred to Agent 2 and Agent 3 cannot be determined. It makes sense since Agent 3 is perfectly rational and therefore base his choice on more factors, but Agent 1 has decided to cooperate with some type of Agent 2. In the interim, the type of Agent 2 that enters the *successful coalition* has higher payoffs than Agent 3.

As can be seen from the previous analysis, conditions for crucial equilibria to exist depend heavily on parametric configurations. When α/β is small compared to λ, parameters satisfy region C and D, meaning that a *successful coalition* of Agent 1 with either Agent 2 or Agent 3 can be expected to rise. To interpret this intuitively, it suggests that if the potential harm of allying with a bad type is small relative to the gains of getting into a good alliance, and/or if the probability of the sender's being a good type is large enough, then both listeners can possibly ally with the sender. This can be associated with presidential elections, because the power of presidents is strictly balanced by the congress and courts, therefore leaving little room for devastating policies brought about by presidents. On the other hand, a great president can be a tremendous treasure to constituents. Especially when it comes to the United States, where Abraham Lincoln and Franklin D. Roosevelt brought about profound changes to the nation in history, the path could suggest that people do expect a risky reformist than a safe competent leader. Even in the presence of multiple candidates, the above argument suggests that a *successful coalition* with bounded rational voters, in which a bold candidate is preferred, can be a reasonable prediction.

2.3.2 Sender Is Possibly Mortal, Receivers Are Sophisticated

In this chapter, I choose parameters $s_s > 0$, $s_l + s_t > 0$ and $r_s = 1$ to produce qualitatively similar results as is in the Crawford model. In other words, Agent 1 will have positive probabilities of being *Sophisticated* and being $Mortal$, while Agent 2 and 3 are both *Sophisticated*. In the Crawford paper, the focus of the study is the sender's ability to represent her intention, while in this model, the interest becomes Agent 1's ability to misrepresent her private information. In particular, I study the conditions for Agent 1 to persuade Agent 2 and 3 to join her when she is actually the "non-aggressive" type.

For convenience, I let $\lambda = \frac{1}{2}$, meaning that Agent 1 is equally likely to be "aggressive" or "non-aggressive". It facilitates a closer comparison with Crawford (2003), since with identical priors assigned to both types of Agent 1, pretending to be another type becomes less suspicious. In the Crawford paper, the ability to deceive stems from the talker's flexibility to choose any combination of message and action. For this model, restricting the two types of the talker to be equally likely allows focuses to be shifted from Nature's assignment to the talker's effort to fake her types.

Unlike in the previous chapter, where the $Mortal$ receiver's role is emphasized in the game structure, I leave out $Mortal$ receivers but allow the sender to be potentially $Mortal$. The theme of "Feint" revisits in this chapter: how could the *Sophisticated* Sender fool an equally *Sophisticated* Receiver? The answer, as in Crawford's analysis, in the sender's probability of being $Mortal$.

The payoff matrix for the three nations are summarized in Table 7 and 8, where $r_s = 1$, $0 < s_s < 1$. Since the focal point is how Agent 1 can deceive Agent 2 and 3 when she is type n, and Agent 2 and 3's payoffs are independent in the underlying game when $t_1 = n$, the coalition formation is less a concern in this part of analysis. Therefore, I only discuss strategies that are symmetric between Agent 2 and 3. Another advantage of this treatment is that the game can be studied like a two player sender-receiver game, which is more close to the two-player game set up in Crawford model.

Table 2.7: Payoff matrices when $t_1 = a$

Agent 2 - J

Agent 1		Agent 3: J	Agent 3: D
	A	$(0,\, s_l(-\alpha),\, s_l(-\alpha))$	$(\beta,\, (s_s+s_t)\beta + s_l(-\alpha),\, (s_s+s_t)(-2\beta))$
	N	$(0,\, s_t(-\alpha),\, s_t(-\alpha))$	$(\beta,\, (s_s+s_l)\beta + s_t(-\alpha),\, (s_s+s_l)(-2\beta))$

Agent 2 - D

Agent 1		Agent 3: J	Agent 3: D
	A	$(\beta,\, (s_s+s_t)(-2\beta),\, (s_s+s_t)\beta + s_l(-\alpha))$	$(0,\, 0,\, 0)$
	N	$(\beta,\, (s_s+s_l)(-2\beta),\, (s_s+s_l)\beta + s_t(-\alpha))$	$(0,\, 0,\, 0)$

Proposition 6. *If* $r_s = 1$, $0 < s_s < 1$, *then there exist the following pure strategy separating PBEs where Agent 2 and 3 play symmetric strategies and the sophisticated Agent 1 tells the truth:*

1.) When $\frac{\alpha}{2\beta} < \frac{s_l}{1-s_l} < \frac{2\beta}{\alpha}$, *Agent 1 plays (A, N), Agent 2 and 3 plays (J, J);*

2.) When $\frac{\beta}{\alpha} < \frac{s_l}{1-s_l} < \frac{\alpha}{\beta}$, *Agent 1 plays (A, N), Agent 2 and 3 plays (D, D).*

Table 2.8: Payoff matrices when $t_1 = n$

Agent 2 - J

Agent 1 \ Agent 3	J	D
A	$(2\alpha,\ (s_s+s_l)(-\alpha),\ (s_s+s_l)(-\alpha))$	$(\alpha,\ s_t\beta+(s_s+s_l)(-\alpha),\ s_t(-2\beta))$
N	$(2\alpha,\ (s_s+s_t)(-\alpha),\ (s_s+s_t)(-\alpha))$	$(\alpha,\ s_l\beta+(s_s+s_t)(-\alpha),\ s_l(-2\beta))$

Agent 2 - D

Agent 1 \ Agent 3	J	D
A	$(\alpha,\ s_t(-2\beta),\ s_t\beta+(s_s+s_l)(-\alpha))$	$(0,\ 0,\ 0)$
N	$(\alpha,\ s_l(-2\beta),\ s_l\beta+(s_s+s_t)(-\alpha))$	$(0,\ 0,\ 0)$

3.) When $\frac{s_l}{1-s_l} > max\{\frac{\alpha}{2\beta}, \frac{\beta}{\alpha}\}$*, Agent 1 plays (A, N), Agent 2 and 3 plays (D, J).*

4.) When $\frac{\alpha}{2\beta} < \frac{s_t}{1-s_t} < \frac{2\beta}{\alpha}$*, Agent 1 plays (N, A), Agent 2 and 3 plays (J, J);*

5.) When $\frac{\beta}{\alpha} < \frac{s_t}{1-s_t} < \frac{\alpha}{\beta}$*, Agent 1 plays (N, A), Agent 2 and 3 plays (D, D).*

6.) When $\frac{s_t}{1-s_t} > max\{\frac{\alpha}{2\beta}, \frac{\beta}{\alpha}\}$*, Agent 1 plays (N, A), Agent 2 and 3 plays (J, D).*

Proof. I prove the first three cases in the proposition, since the rest statements follow immediately once one notice the symmetry in payoff functions.

For the first case, fix Agent 1's strategy (A, N), in equilibrium, Agent 2 and 3 must form the belief $b_2(A) = b_3(A) = 1$, and $b_2(N) = b_3(N) = 0$. Therefore, $s_2 = s_3 = (J, J)$ is an equilibrium if and only if $\pi_i(r_i = r_j = J|t_1) > \pi_i(r_i = D, r_j = J|t_1)$, where $i, j \in \{2, 3\}$, for $t_1 = a, n$. Thus, $s_2 = s_3 = (J, J)$ is an equilibrium if and only if $s_l(-\alpha) > (s_s + s_t)(-2\beta)$ and $(s_s + s_t)(-\alpha) > s_l(-2\beta)$, or $\frac{s_l}{s_s+s_t} < \frac{2\beta}{\alpha}$, $\frac{s_l}{s_s+s_t} > \frac{2\alpha}{\beta}$, which is equivalent to $\frac{2\alpha}{\beta} < \frac{s_l}{1-s_l} < \frac{2\beta}{\alpha}$ since $s_l + s_t + s_s = 1$. Given that Agent 2 and 3 play the same strategies regardless of Agent 1's message, *Sophisticated* Agent 1 will have no incentive to deviate since her own payoff does not depend directly on her message. This constitutes a PBE by definition.

For the second case, given the same s_1 and belief systems, $s_2 = s_3 = (D, D)$ is an equilibrium if and only if $\pi_i(r_i = r_j = D|t_1) > \pi_i(r_i = J, r_j = D|t_1)$, where $i, j \in \{2, 3\}$, for $t_1 = a, n$. Thus, $s_2 = s_3 = (D, D)$ is an equilibrium if and only if $0 > (s_s + s_t)\beta + s_l(-\alpha)$ and $0 > s_l\beta + (s_s + s_t)(-\alpha)$, or $\frac{\alpha}{\beta} < \frac{s_l}{1-s_l} < \frac{\beta}{\alpha}$. Similar to the discussion above,*Sophisticated* Agent 1 will have no incentive to deviate since her own payoff does not depend directly on her message. This also constitutes a PBE by definition.

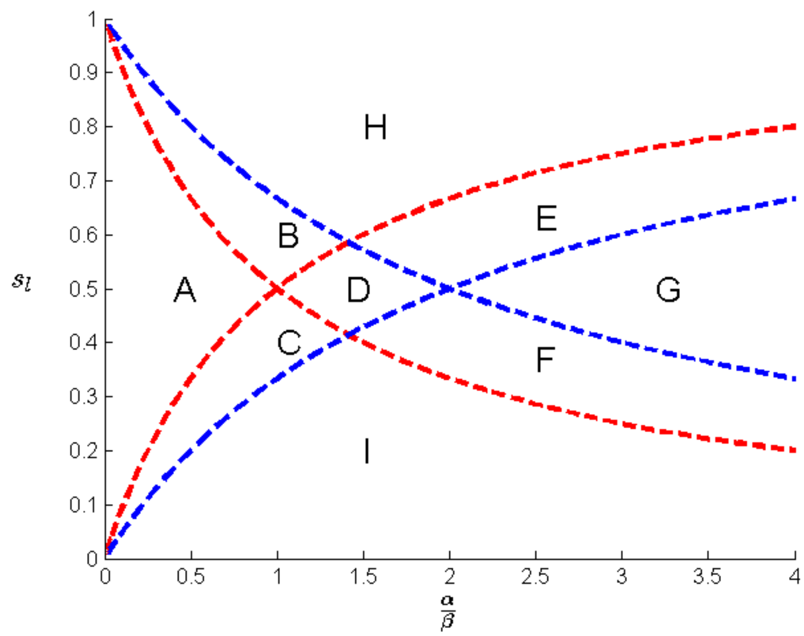
For the third case, given the same s_1 and belief systems, $s_2 = s_3 = (D, J)$ is an equilibrium if and only if $\pi_i(r_i = r_j = D|t_1 = a) > \pi_i(r_i = J, r_j = D|t_1 = a)$ and $\pi_i(r_i = r_j = J|t_1 = n) > \pi_i(r_i = D, r_j = J|t_1 = n)$, where $i, j \in \{2, 3\}$. Thus, $s_2 = s_3 = (D, J)$ is an equilibrium if and only if $0 > (s_s + s_t)\beta + s_l(-\alpha)$ and $(s_s + s_t)(-\alpha) > s_l(-2\beta)$, or $\frac{s_l}{1-s_l} > max\{\frac{\alpha}{2\beta}, \frac{\beta}{\alpha}\}$. □

Proposition 7. *If $r_s = 1$, $0 < s_s < 1$, the equilibria listed in the previous proposition are the only pure strategy separating PBEs where Agent 2 and 3 play the same strategy.*

Proof. The only potential pure strategy separating PBEs left are ((A, N), (J, D), (J, D)) and its almost equivalent counterpart, ((N, A), (D, J), (D, J)). In both cases, Agent 2 and 3 join if Agent 1' true type "aggressive", and choose not to join if Agent 1 is "non-aggressive". However, this cannot be the case in equilibrium, since the "non-aggressive" Agent 1 would have the incentive to mimic the "aggressive" type and induce both other players to join (and free ride them). □

The parametric configuration for equilibria to exist is shown in the graph below: Equilibrium 1 (or 4) can be sustained in region A+B+C+D; equilibrium 2 (or 5) can rise in region D+E+F+G; equilibrium 3 (or 6) can be maintained in region B+D+E+H.

The intuition, in general, is that when Agent 1 has some probability of being a *Liar* (*Truth-teller*), Agent 2 and 3 will not play the dominant strategy in the baseline model even if the *Sophisticated* Agent 1 is always telling the truth (a lie). This result is intuitive, as the probability of Agent 1's being $Mortal$ creates a concern for her listeners so that they cannot trust her valuable information in separating equilibria. Her ability to deceive in a revealing separating equilibrium crucially relies on her potential of being $Mortal$ and discussing her private information in an opposite manner.



This result echoes a main finding in the Crawford paper: when the sender has a probability of being $Mortal$, the *Sophisticated* sender can successfully feint and exploit a *Sophisticated* receiver. In addition, this paper produces a similar outcome that the behavioral parameters do not have to be high for the *Sophisticated* sender to successfully misrepresent her intention. In my analysis, those parameters do not have to be high for the type n Agent 1 to fully reveal her type but still exploit the *Sophisticated* receivers. In fact, s_l or s_t only needs to be greater than or equal to $\frac{\alpha}{2\beta}$ (region A+B+C+D+E+H) for type n Agent 1 to always induce Agent 2 and 3 to join her, and it can hold for infinitely small s_l (s_t) as long as $\frac{s_l}{1-s_l}$ (or $\frac{s_t}{1-s_t}$) $< \frac{\alpha}{2\beta}$. There is also a similar result in Crawford paper, while he suggested that in his paper, the reason that behavioral parameters can be very small in equilibrium is due to the restrictions on $Mortal$ players' equilibrium strategies, but in my model, $Mortal$ sender's strategies are determined by Nature when her type is randomly chosen.

What is somewhat counter-intuitive is that symmetric pure strategy separating equilibria exist, and only exist, when the listeners are not playing their dominate strategy in the static game, even if they have fully learned about the *Sophisticated* sender's type, t_1. Moreover, the deviation from static game dominant strategy gets more perverse if Agent 1's probability of being a *Liar* (*Truth-teller*) is higher, when $\frac{s_l}{1-s_l}$ (or $\frac{s_t}{1-s_t}$)$> max\{\frac{\alpha}{2\beta}, \frac{\beta}{\alpha}\}$. In that case, Agent 2 and 3 both plays (D, J), and the *Sophisticated* sender will get expected payoff $\frac{1}{2} \cdot 0 + (1 - \frac{1}{2})(2\alpha) = \alpha$. The necessary component of the statement is easier to understand: if the equilibrium is separating, given that type n Agent 1 would always have the incentive to fake her type when both listeners play dominant strategies in the static game, it cannot be an equilibrium. The other direction is more difficult to see, but the talker's another ego would always help justify some seemingly unnatural precautions taken by the listeners.

Numerical Examples

Consider the following cases: i.) $s_l = s_s = 0.5$, $s_t = 0.5$ and $\frac{\alpha}{\beta} = 1.5$. This is the case when the sender has half chance of being a liar, and half chance of being sophisticated. The danger posited by bad allies (α) is moderate compared to the stake in the coalition (β). According to the solutions above, separating equilibria 1.) - 3.) are the ones can be substantiated. In those cases, the *Sophisticated* sender always speaks her true profile, and for equilibria 1.) and 3.), even with regard to this, *Sophisticated* receivers choose to play $Join$ when they are told that the sender is $Non - aggressive$. Out of the fear that the sender is a liar (with the probability of a half) coupled with a moderate threat of bad ally compared to the gains in a *successful coalition* ($\frac{\alpha}{\beta} = 1.5$), the receivers choose to act in the opposite way upon the negative news.

ii.) $s_t = 0.2$, $s_l = 0.1$, $s_s = 0.7$ and $\frac{\alpha}{\beta} = 0.1$. This is the case when the sender has a low probability of being a truth-teller (0.2) and a liar (0.1), and high probability of being sophisticated (0.7). The cost of allying with an undesirable type is small relative the gain of joint expropriation, which resembles the free-ride case instead of the "predating the partner" story. This corresponds to region A in the s_t plot and region I in the s_l plot. The unique equilibrium 4.) suggests that if the *Sophisticated* type n sender reports A, rational receivers will play J regardless of the low probability of the information being true, since the downside of joining a bad coalition is minimal compared to the fortune one will make in a successful one.

Chapter 3

Conclusions

3.1 The Electoral Example Revisited

In this paper, I studied coalition formation when some agents are allowed to be naïve. My main finding is that the sender can collude with a bounded rational receiver and jointly exploit a sophisticated receiver in pooling equilibria. Furthermore, it applies to the case when a presidential candidate is babbling, and the *Sophisticated* receiver decides to rationally ignore his or her messages and not to vote for that person after careful calculation, but the cheap talk worked to naïve voters in a way the candidate had hoped.

The results allow us to think further about the role of rhetoric in politics. Not only does it appear in the Republican candidate Donald Trump's controversial speeches on race, gender, immigration and international trade, such policy-like agenda is also everywhere in Democratic senator Bernie Sanders' talks: free college education for everyone, free medical care to everyone and so on. Surveys and speculations have been made on what kind of people vote for those leaders, and the results seem to be no more than reinforcing our biased impression on the lower income class.

It would make more sense to ask why there would be such voters forming such political power with those candidates. Traditional theory would suggest that it is because those voters outnumber others, so that they become a group gaining more weight in politician's considerations. Simple special economy models also suggest that their extreme view points could manifest disequilibrium in candidates' choosing political positions. Insightful as they are, such researches are based on the assumption that those words carry weights.

My conclusion depart from those traditions and suggest that rhetoric is everything that matters. People have different mechanism of responding to rhetorics, which gives politicians to make strategies catered to some subgroup of voters. In this model, it requires there to be at least some proportion of $Mortal$ players that take the rhetoric literally, and those messages can have an effect on those people, but not the *Sophisticated* players. *Sophisticated* players have more thorough considerations on the Sender's motivation, uncertainty and possible outcome of the game under each scenario, therefore being less likely to respond to the Sender's messages. On the other hand, such deliberation can lead to different actions from naïve voters, subjecting *Sophisticated* players to the risk of expropriation.

This model characterizes a situation where politicians, Trump or Sanders alike, create a pooling equilibrium with voters. *Sophisticated* voters do not update any beliefs according to their messages, but $Mortal$ voters believe such message is linked to the true type of the candidate. It is ironic that supporters of the two candidates both agreed that they cannot believe anything Hilary Clinton says, since she "is such a politician and hypocrite". Although the model uses predetermined behavioral parameters for senders, in reality, a politician can invest in building a trust-worthy image to earn a population of "believers". In the end, however, it may suggest that in a game situation, strategies of rational cheap talkers there can only be a matter of how sincere a candidate may look.

3.2 Extensions

There are several ways to extend this model. First, it makes sense to extend it to an n-listener game to better account for the intricate interactions in public choices. There are at least two benefits of contemplating in a more generalized game. On one hand, bounded rational players can influence the campaign results indirectly through aggregated results. For example, it can be the case that the voting outcome from a small number of $Mortal$ voters can change people's expectations in early stages of elections. On the other hand, by allowing there to be different sizes in the populations of *Sophisticated* and $Mortal$ voters, it helps to analyze how the relative voting power of different voter groups affect the campaign strategy. In reality, this voting power can correspond to financial resources, political influences or simply count of votes.

Second, the model can be extended into dynamic games, so that both *Sophisticated* and $Mortal$ players can update their beliefs after observing the first stage outcome. Not only can it make the characterization $Mortal$ less rigid, this potential improvement can also make the model more interesting in a theoretical sense. Moreover, dynamic games are better suited to the modeling of our political reality. Adding another stage resembles the fact that most offices allow re-elections after one term, so as to help account for reputation effects in communication games.

Third, the definition of bounded rationality can be refined. To add sophistication to the naive players, one can specify a source of bias, or some cognition rules that those players rely upon. An advanced way of modeling bounded rationality can provide more insights into how those people make decisions in a supposedly strategic situation. However, it would not change the above results qualitatively. As long as they believe that they can infer something from a candidate's cheap talk, as opposed to treat it with strategic precaution, there is always room for manipulating their beliefs.

Lastly, it might be interesting to bring the predictions of this model into test. For example, one can use survey data about general public's perception about the trustworthiness of political candidates, along with election outcomes and policy implementation to study whether there is evidence for such belief manipulation.

Bibliography

[1] B Douglas Bernheim, Bezalel Peleg, and Michael D Whinston. Coalition-proof nash equilibria i. concepts. *Journal of Economic Theory*, 42(1):1–12, 1987.

[2] Archishman Chakraborty and Rick Harbaugh. Persuasion by cheap talk. *Available at SSRN 889190*, 2007.

[3] Vincent P Crawford. Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Economic Review*, pages 133–149, 2003.

[4] Vincent P Crawford and Nagore Iriberri. Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions? *Econometrica*, 75(6):1721–1770, 2007.

[5] Vincent P Crawford and Joel Sobel. Strategic information transmission. *Econometrica: Journal of the Econometric Society*, pages 1431–1451, 1982.

[6] David Ettinger and Philippe Jehiel. A theory of deception. *American Economic Journal: Microeconomics*, 2(1):1–20, 2010.

[7] Erik Eyster and Matthew Rabin. Cursed equilibrium. *Econometrica*, 73(5):1623–1672, 2005.

[8] Jacob Glazer and Ariel Rubinstein. A model of persuasion with boundedly rational agents. *Journal of Political Economy*, 120(6):1057–1082, 2012.

[9] Kenneth Hendricks and R Preston McAfee. Feints. *Journal of Economics & Management Strategy*, 15(2):431–456, 2006.

[10] Philippe Jehiel. Analogy-based expectation equilibrium. *Journal of Economic theory*, 123(2):81–104, 2005.

[11] Navin Kartik, Marco Ottaviani, and Francesco Squintani. Credulity, lies, and costly talk. *Journal of Economic theory*, 134(1):93–116, 2007.

[12] David M Kreps and Robert Wilson. Reputation and imperfect information. *Journal of economic theory*, 27(2):253–279, 1982.

[13] Christopher A Sims. Implications of rational inattention. *Journal of monetary Economics*, 50(3):665–690, 2003.

[14] Joel Sobel. A theory of credibility. *The Review of Economic Studies*, 52(4):557–573, 1985.

[15] Michael Spence. Job market signaling. *The quarterly journal of Economics*, pages 355–374, 1973.

[16] Dale O Stahl and Paul W Wilson. On players′ models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1):218–254, 1995.

[17] Matthias Sutter. Deception through telling the truth?! experimental evidence from individuals and teams*. *The Economic Journal*, 119(534):47–60, 2009.

[18] Dustin H Tingley and Barbara F Walter. Can cheap talk deter? an experimental analysis. *Journal of Conflict Resolution*, 55(6):996–1020, 2011.

[19] John Von Neumann and Oskar Morgenstern. *Theory of games and economic behavior*. Princeton university press, 2007.