On the Economics of Information: Three Essays

by

Jong Jae Lee

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

July, 2017

Abstract

In this dissertation, we study whether individuals with differing interests are able to achieve a socially efficient outcome in the presence of incomplete information about the others. Unlike the case of complete information, an individual's decision may reveal his private information, thereby impinging on the others' decisions. This signaling aspect of one's decision would force a decision maker to take account of what others would come to know about his private information. Studying this feature leads us to a rigorous examination, first of all, of how the notion of information ought to be understood and thus to be mathematically formulated; and secondly, of how this signaling aspect reduces the range of achievable efficient decision rules relative to the case of complete information.

In the first chapter titled "Formalization of Information: Knowledge and Belief", we engage in the first task by studying the issue Billingsley (1995) and Dubra and Echenique (2004) raise about the use of σ-algebra to model information. They provide an example to show that the formalization of information by σ-algebras and by partitions need not be equivalent. Although Hervés-Beloso and Monteiro (2013) provide a method to generate a σ-algebra from a partition and another method for going in the opposite direction, we show that their two methods are in fact based on two *different* notions of information: (i) information as belief, (ii) information as knowledge. If information is conceived to allow for falsehood, case (i) above, the equivalence between σ-algebras and partitions holds after applying the notion of posterior-completion suggested by Brandenburger and Dekel (1987). If information is conceived not

to allow for falsehood, case (ii) above, the equivalence holds only for measurable partitions and countably-generated σ-algebras.

In the second chapter titled "Common Knowledge and Efficiency with Incomplete Information", we engage in the second task. Holmström and Myerson (1983) show that we need only check for efficiency on common knowledge events to determine that an incentive compatible decision rule is efficient. By a sharper notion of common knowledge, based on the notion of posterior-completion described in the first chapter, we show that we need only check for efficiency in a *strict* subset of common knowledge events known as self-evident events and furthermore, that this is the *minimal* class of events that one needs to check.

In the third chapter titled "Mediator Selection in International Conflict: Bias, Effectiveness, and Incidence", we adapt the question of achieving efficiency to the context of international conflicts and mediation. As war incurs a cost, an efficient outcome is thus a peaceful one in this context. We allow for disputants to make a joint decision whether to accept a potentially biased mediator who would communicate with them and propose a decision rule on their behalf. This extends the mechanism design problem of Hörner et al. (2015) to allow for mediator bias and its endogenous determination. Our main finding is that *both* disputants would accept a biased mediator if war is highly likely to occur in a conflict and the mediator's bias is moderate. More importantly, once a mediator has been accepted, the probability of attaining peace is *independent* of the intensity of her bias: because war is inefficient, the interest of the mediator's favored disputant is best served by promoting peace.

Advisors: Ying Chen, M. Ali Khan

Acknowledgement

This dissertation could not have been completed without help from countless people in my life. I am deeply grateful to my primary supervisors Ying Chen and M. Ali Khan for their unconditional support and guidance. It has been an honor to be their student, and I am lucky to have them as my advisers. Getting to know M. Ali Khan was a serendipity of my life. I still remember when he wrote "Mathematics is a language" on the blackboard on the first day of his lecture on theory. The commonalities I share with him about the importance of philosophy, history of economic thoughts, and mathematics made me feel much closer to him than any other one in my life, and he is truly a role model to me. I came to know Ying Chen when she joined Hopkins in 2014. Her remarks regarding my works have always been to the point and thus very helpful in overcoming numerous obstacles I have been facing through my research. I would also like to express my deep appreciation to John Quah for his careful reading of my chapters and his insightful comments about them. I am greatly indebted to him for completing my third chapter.

I wish to express my deep appreciation to Hülya Eraslan for her encouragement in my decision to do economic theory. In my second year, I wished to do economic theory, but I was not at all confident about whether I would be able to do it. It was her remarks in the email that gave confidence to me, in which she express her trust in my ability to do theory. Also, I wish to mention that my second chapter would not be possible without her. The theory reading group she organized got me involved in the world of economic theory, and my second chapter came out of the discussion about Hömstrom-Myerson in the reading

group.

I am also grateful to my undergraduate supervisors, Gyu Ho Wang and E Young Song. Gyu Ho Wang was the first one who got me interested in economic theory, and with a bit harsh training, he taught me what a mathematical rigor is. E Young Song enlightened me with his insightful remarks about what economic models can say about the world outside, which made it a habit for me to think of the implication of theory to the economic and social phenomena.

I wish to say special thanks to Jaiung Jun, a mathematician I was lucky to know during my days in Hopkins. From my conversations with him regarding what can be said for sure from economic theory, I got to come across category theory and tropical geometry which opened my eyes to the deeper world of mathematics and their potential applications to economic theory. Although my collaboration with him on the economic applications of tropical geometry to economics has not proceeded well enough to be included into this dissertation, I believe we shall make a good progress in the near future.

I would like to thank my fellow doctoral students, to name a few, Liuchun Deng, Sohini Mahapatra, Alanna Bjorklund-Young, Emmanuel Garcia-Morales, Victor Ronda, Osama Khan, and Kyungmin Kang for their feedback, cooperation and of course friendship. In addition, I would like to express my gratitude to Jeongho Park, Seunghun Lee, and Cheonwoo Kim for their support.

A special gratitude goes out to my two life-long friends, Dongkyu Chang and Jae Ho Lee. Dongkyu Chang and I have known each other since 2004 and, as a close friend who wanted to do economic theory, he and I have had about a decade of discussions about economic theory. I wish my current collaboration with him would soon turn out to be a meaningful work. Jae Ho Lee has been

with me since my high school days, and whenever I lose confidence in myself and wish to give up, he gave his trust in me and encouraged me. In addition, I wish to mention Geunhye Kim, Seungmin Kook, and Jonghyek Jeong for their friendship. My life would not be as bearable as it has been without them.

Finally, but by no means least, I would like to thank my parents, Yeonkyu Lee and Seonhyun Yoo, and my brother Jongsang Lee. They have given me the gift of sweet childhood memories, and the gift of dreams: the ones they have made come true, the ones I achieved myself because of their encouragement and support, and the ones I would like to achieve in the future. My mother Seonhyun Yoo have always given me the gift of her love, and it has given me strength for the hard times. Because of her, I have been able to see the good in people and understand what really matters in life.

This dissertation is dedicated to my late father Yeonkyu Lee. He was a historian who worked on the transition from feudalism to capitalism in Britain. His scholastic ideals and values became a part of me a long time ago, and they have been with me every step of the way as I have searched for my own scholastic journey. During my undergraduate years, he often asked me what I learnt and whether I enjoyed it. Also, he enjoyed listening to my own view about the subjects I learnt. He was not merely there as my father, but also as my life-long academic fellow and supporter. When I decided to pursue my doctorate study, he said he really expected to read my dissertation and to discuss it together. Unfortunately, however, he suddenly passed away before I joined Hopkins. I really miss him and our interesting and long-lasting discussions.

Table of Contents

List of Figures

Chapter 1

Formalization of Information: Knowledge and Belief

1.1 Introduction

In any model that deals with a decision maker (henceforth DM) facing uncertainty, the DM's information is often described by either a signal (equivalently, a random variable), a partition or a σ-algebra. Specifically, one signal is more informative than another if it is sufficient in Blackwell's sense for another; one partition is more informative than another if it is finer; a σ-algebra is more informative than another if it is larger.[1] A natural question is whether all these three orderings can be equivalently used to represent information. In other words, it is to ask whether there is a mapping from one category of representation to another while preserving the ordering in the two categories that are being used. The answer to this question had been understood to be positive. Nevertheless, as we shall see below, the understanding is far from complete.

[1] For a pair of partitions, the strictly finer partition distinguishes more elements, implying that a DM can say more accurately about the true state (the state in which she lies). For a pair of σ-algebras, the larger one contains more sets. For larger number of sets, a DM is able to say whether it contains the true state or not, thus having more information.

Billingsley (1995) raises concerns that partitions and σ-algebras may not always be equivalently used by presenting a simple but powerful example: a unit interval is given as the state space equipped with the Lebesgue measure. A partition that consists of every singleton indicates that the DM knows exactly in which state she lies. On the contrary, the smallest σ-algebra generated by the partition implies that the DM is totally ignorant, for it contains countable or co-countable sets that are of Lebesgue measure zero.[2] In addition, Dubra and Echenique (2004) highlight Billingsley's concern by embedding his example in the context of a decision problem. They consider another partition that contains only two cells. This partition is obviously less informative than the partition in Billingsley's example. However, if one compares the expected utility values conditional on the smallest σ-algebras generated by those partitions, the value based on the two-cell partition is larger. That is, the σ-algebra generated by the two-cell partition is more informative.

In response to these cautionary warnings, Hervés-Beloso and Monteiro (2013) (henceforth HM) argue that one may disregard them. By taking a partition as a primitive representation of information, they introduce a notion of an *informed set* which corresponds to a (possibly uncountable) union of partition cells. The collection of all informed sets is indeed a σ-algebra. If one generates σ-algebras in this way, a strictly finer partition always yields a larger σ-algebra. Arguably, the collection of informed sets represents the informational content of a given partition. To establish the equivalence between partitions and σ-algebras, they also suggest another method of deriving a partition from a given σ-algebra.

[2] By the smallest σ-algebra generated by the partition, we mean that the σ-algebra contains all the complements and the countable unions of partition cells.

Given a measure space equipped with a strongly-Blackwell σ-algebra, HM suggest to form a partition by collecting atoms of a σ-algebra if it is countably-generated.[3] If not, they suggest to consider a countably-generated σ-algebra that differs from the given σ-algebra by null sets.[4] This implies, although HM do not so explicitly argue, that the informational content of a σ-algebra is captured by the corresponding countably-generated σ-algebra.

In this paper, our primary goal is to show that HM leave unsettled the following question, "What is the information (equivalently, the informational content) preserved when one generates a σ-algebra from a partition or when one goes in the opposite direction?" HM claim that it is the collection of informed sets, and they interpret the notion of an informed set to denote the set of which occurrence (or non-occurrence) a DM knows. This naturally leads one to ask a question about the difference between what one merely knows and what one is informed of. Unfortunately, however, HM are silent on this question. In addition, the informational content, as HM claim, is also captured by the countably-generated σ-algebra that differs from a σ-algebra by null sets. This implies that if both a partition and a σ-algebra contain the same informational content, the collection of informed sets of the partition must be countably-generated. We present a counterexample in which this is not the case (Example 5). Furthermore, the collection of informed sets of a partition may contain non-measurable sets, because the informed sets do not depend on a given measurable space. We show that

[3] A countably generated σ-algebra is the smallest σ-algebra generated by a collection of countably many subsets of the state space.

[4] A null set is a set to which a DM ascribes zero probability. HM refers to it as a negligible set of states.

this indeed happens in Billingsley's example (Example 4). This poses a technical impossibility of defining a probability measure on the non-measurable set when computing the expected utility value as in Dubra and Echenique (2004), not to mention a conceptual difficulty of how to understand that a DM is informed about a set lying outside the event space.[5] More importantly, HM's treatment provides a contradictory answer about whether a probability measure conveys any informational content or not. As noted, informed sets of a partition is invariant to any choice of a measure space and a measure defined on it. This suggests that a probability measure does not convey any information. Contradictorily, a probability measure conveys information if one considers the informational content embodied in a σ-algebra, as it is unique up to null sets.

The secondary goal of this paper is to tackle these issues and to establish an equivalence relationship between a partition and a σ-algebra in representing information. Our innovation is to bring out with especial salience the two distinct notions of information, knowledge and belief, that are well-recognized among researchers working in epistemic logic and game theory.[6] The distinction lies in whether information is conceived to be factual or not. To elaborate, if one insists that information cannot be false in order to distinguish it from a rumor, then he conceives information to arise from knowledge. On the contrary, if one allows for the possibility that information may turn out to be false, then he conceives information to arise from belief.

[5]The existence of non-measurable sets can be addressed by Theorem 4 and the following Remark 3 in HM. However, the notion of an informed set, as it is defined in HM, fails to accommodate this: the collection of informed sets in Billingsley's example, according to HM, is the power set even when the underlying σ-algebra is strongly Blackwell (See Example 4 in HM). In fact, we propose the notion of an informed event to accommodate Theorem 4 and Remark 3 in HM.

[6]See, for example, Aumann (1999a,b), Maschler et al. (2013), and Meyer (2003).

The advantage of bringing out these two notions of information is that each notion, either knowledge or belief, is formally defined as an operator from a measurable space (or, equivalently, an event space) to itself that satisfies a certain set of axioms (Definition 4, 5). One can thus see easily whether a mathematical object such as a partition and a σ-algebra qualifies for being a formalization of information (as knowledge/belief), by inspecting the relationship between a knowledge/belief operator to the mathematical object of one's interest. By taking advantage of the two notions, we resolve the issues that HM leave open. Firstly, we show that the notion of an informed event imposes a counterfactual restriction on that of knowledge/belief. To be specific, we define a K-informed event for information as knowledge, and a B-informed event for information as belief. An informed event requires that if one knows/believes whether an event occurs or not at one state, then he must know/believe it even in a hypothetical situation that he lies at other states (Example 3). Secondly, we show that the collection of K-informed events is the restriction of the collection of informed sets (defined by HM) to a measurable space, thereby resolving the issue regarding the presence of non-measurable set (Lemma 3). This is immediate from the definition of knowledge/belief being an operator from a measurable space to itself.

Turning to the remaining issues, we show that if one conceives information as knowledge, measurable partitions and countably-generated σ-algebras can be used interchangeably to formalize information (Theorem 4). This implies that the preserved informational contents are the K-informed events of measurable partitions. Moreover, it also reveals that probability does not convey any information, for the K-informed event is invariant to a specific choice of a probability

measure.

A further question is whether we need restrict the use of partitions or σ-algebras only to the case where partitions are measurable or σ-algebras are countably-generated. We argue that if one conceives information as belief, we do not need such a restriction. By adopting the technique of *posterior-completion*[7] proposed by Brandenburger and Dekel (1987), we show that if the posterior-completion of a σ-algebra is larger then the posterior-completion of a partition is strictly finer, and vice versa. Then, what is the informational content in this case? We argue that the informational content is indeed the collection of B-informed events, and it depends on a specific choice of a probability measure. Specifically, a proper regular conditional probability (either directly from a σ-algebra or from the smallest σ-algebra generated by a partition) captures the notion of belief. More importantly, the collection of B-informed events is the posterior-completion of a given σ-algebra. Since B-informed events are defined in relation to a given probability measure, probability conveys information.

The paper is structured as follows: we present preliminary definitions including the notions of knowledge and belief in Section 2. In Section 3, under the conception of information as knowledge, we establish an equivalence between measurable partitions and countably-generated σ-algebras in formalizing information. Moreover, we discuss the issues regarding the notion of informed sets as formalized by HM. Section 4 consists of an equivalence result under the conception of information as belief. Then, we conclude in Section 5.

[7]The posterior-completion of a σ-algebra is to create the smallest σ-algebra by adding events that are either measure zero or one with a proper regular conditional probability measure, into a given σ-algebra. The posterior-completion of a partition is to add in those events to the partition.

1.2 Preliminaries

Partitions and σ-algebras Let $(\Omega, \mathcal{F})$ be a measurable space, where Ω is a nonempty set of states endowed with a σ-algebra $\mathcal{F}$, so-called the event space. Measurable sets of the σ-algebra $\mathcal{F}$ are called events. We assume that Ω is a complete separable metric space, and the event space $\mathcal{F}$ is a strongly Blackwell σ-algebra.[8] The complement of an event E is denoted by $\neg E$.

Definition 1. Let X and Y be partially ordered sets (posets) with the partial orderings $\succeq^X$ and $\succeq^Y$. A mapping $\Phi : X \to Y$ is an order isomorphism if Φ is bijective and preserves order in the following sense: $x \succeq^X x' \iff \Phi(x) \succeq^Y \Phi(x')$. If such an order-isomorphism exists, X and Y are said to be order-isomorphic.

Definition 2 (Partition)**.** Let $(\Omega, \mathcal{F})$ be given. A collection of nonempty events is called a *partition* and denoted by Π if it satisfies the following:

(1) $\cup\{E|E \in \Pi\} = \Omega$;

(2) If $E, F \in \Pi$ and $E \neq F$, then $E \cap F = \emptyset$.

Note that we define a partition to be a collection of events (or, equivalently, measurable sets). Let Π_ω denote an element of Π containing a state ω, and it is unique. For two partitions Π and Π', we say that Π is *finer* than Π', denoted by $\Pi \succeq^P \Pi'$, if for each $\omega \in \Omega$, $\Pi_\omega \subseteq \Pi'_\omega$. Let P be a collection of all partitions of Ω. Then, $\succeq^P$ is a partial ordering on P and $(P, \succeq^P)$ is a partially ordered set (poset).

[8] A σ-algebra is a strongly Blackwell σ-algebra if it is separable and every two countably generated sub-σ-algebras with the same atom coincide.

Definition 3 (Sub-σ-algebra)**.** Let $(\Omega, \mathcal{F})$ be given. A sub-σ-algebra $\mathcal{G}$ is a sub-collection of events satisfying the following two properties:

(1) Closed under complements: for any $E \in \mathcal{G}$, $\neg E \in \mathcal{G}$.

(2) Closed under countable unions: for any countable number of events $\{E_i\}_{i \in I}$ with $E_i \in \mathcal{G}$, $\cup_{i \in I} E_i \in \mathcal{G}$.

For a σ-algebra $\mathcal{G}$ and a state $\omega \in \Omega$, an atom $\mathcal{A}(\omega, \mathcal{G}) = \cap\{G \in \mathcal{G} | \omega \in G\}$ is the smallest set containing ω in a σ-algebra $\mathcal{G}$. Whenever $\mathcal{G}$ is obvious, we simply denote it by $\mathcal{A}_\omega$.

Let Σ be a collection of all sub-σ-algebras of Ω. A sub-σ-algebra $\mathcal{G}$ is larger than $\mathcal{H}$ if for every $E \in \mathcal{H}$, $E \in \mathcal{G}$. This naturally defines a partial ordering $\succeq^\sigma$ on Σ such that for two sub-σ-algebras $\mathcal{G}$ and $\mathcal{H}$, $\mathcal{G} \succeq^\sigma \mathcal{H}$ if $\mathcal{G}$ is larger than $\mathcal{H}$. Then, $(\Sigma, \succeq^\sigma)$ is a poset.

For the two posets $(P, \succeq^P)$ and $(\Sigma, \succeq^\sigma)$, define a mapping $F : (P, \succeq^P) \to (\Sigma, \succeq^\sigma)$ such that for $\Pi \in P$, $F(\Pi)$ is the smallest σ-algebra generated by the partition cells of Π. Then, as the following example from Billingsley (1995) shows, F is not an (order) isomorphism.

Example 1 (Billingsley)**.** Let $\Omega = [0, 1] \subset \mathbb{R}$ endowed with a Borel σ-algebra $\mathcal{F}$. Let $\Pi = \{\{\omega\} | \omega \in \Omega\}$ and $\Pi' = \{\left[0, \frac{1}{2}\right), \left[\frac{1}{2}, 1\right]\}$. Then, $F(\Pi) = \{E \in \mathcal{F} |$ either E or $\neg E$ is countable$\}$ and $F(\Pi') = \{\emptyset, \left[0, \frac{1}{2}\right), \left[\frac{1}{2}, 1\right], \Omega\}$. Clearly, Π is finer than Π' ($\Pi \succeq^P \Pi'$). However, neither σ-algebra is larger than the other: neither $F(\Pi) \succeq^\sigma F(\Pi')$ nor $F(\Pi') \succeq^\sigma F(\Pi)$.

Belief and Knowledge The following definitions are standard in the literature on epistemic logic and game theory. For example, see Aumann (1999a,b),

Maschler et al. (2013), and Meyer (2003).

Definition 4 (Belief)**.** Let $(\Omega, \mathcal{F})$ be given. An operator $B : \mathcal{F} \longrightarrow \mathcal{F}$ is said to be a *belief* if B satisfies the following axioms:

A1 Conjunction: For any countable index set I and events $\{E_i\}_{i \in I}$ with $\cap_{i \in I} E_i \in \mathcal{F}$, $\cap_{i \in I} B(E_i) = B(\cap_{i \in I} E_i)$.

A2 Consistency: $B(E) \cap B(\neg E) = \emptyset$.

A3 Positive introspection: $B(E) \subseteq B(B(E))$ for $E \in \mathcal{F}$.

A4 Negative introspection: $\neg B(E) \subseteq B(\neg B(E))$ for $E \in \mathcal{F}$.

For $\omega \in \Omega$ and $E \in \mathcal{F}$, $\omega \in B(E)$ is read as "A DM *believes* an event E at a state ω." Therefore, for an event E, $B(E)$ is an event that whenever it occurs, the DM believes that the event E occurs. In this sense, $B(E)$ is the event that is an evidence based on which the DM believes E.

Definition 5 (Knowledge)**.** Let $(\Omega, \mathcal{F})$ be given. An operator $K : \mathcal{F} \longrightarrow \mathcal{F}$ is said to be *knowledge* if it satisfies the axioms of a belief operator and the following additional axiom:

A5 Non-delusion: $K(E) \subseteq E$ for $E \in \mathcal{F}$.

Note that a knowledge operator K is also a belief operator, but the converse does not hold in general. In what follows, we shall use B to denote a belief operator and K a knowledge operator. Similarly to the case of belief, we say that the DM *knows* at ω that the event E occurs, or simply that the DM knows E at ω if $\omega \in K(E)$.

Any belief operator B satisfies the following properties:

A6 Necessitation: $B(\emptyset) = \emptyset$.

A7 Monotonicity: $E \subseteq F$ implies $B(E) \subseteq B(F)$.

The proof is easy, so we omit it.[9] Given a belief operator, one can completely describe what the DM believes at each state, or his *doxastic* status. Similarly, a knowledge operator specifies what the DM knows at each state, or his *epistemic* status. If one chooses a different belief (or knowledge) operator, it indicates a different doxastic (or epistemic) status as it is illustrated in the following example.

Example 2. Let $\Omega = \{\omega_1, \omega_2\}$ and $\mathcal{F} = 2^\Omega$. Consider two knowledge operators, K and K' such that $K(\{\omega\}) = \{\omega\}$ for $\omega \in \Omega$, and $K'(\{\omega\}) = \emptyset$. Let ω be the true state. For any event E with $\omega \in E$, $\omega \in K(E)$ but $\omega \notin K'(E)$ unless $E = \Omega$. The knowledge operator K thus implies that a DM knows all the events that actually occur at the true state. On the contrary, K' indicates that the DM does not know any event that occurs, except for that the state space Ω itself occurs.

Note that the notion of belief and thus of knowledge rely on the event space $\mathcal{F}$. Although the definitions given in this paper are standard in the literature on epistemic logic and game theory, this reliance may raise an issue about why some sets of states, if they lie outside the event space, are precluded from being the subjects of belief and knowledge. This issue becomes trivial if the event space is given as the powerset. Hence, we shall focus on the case where the event space is strictly smaller than the powerset. Then, a natural question arises. What is the meaning of an event if it does not merely mean a set

[9] Interested readers may see, for example, Bacharach (1985).

of states? Before answering this, one cannot understand why the set of states being an event is essential in defining the notion of belief and thus of knowledge. Unfortunately, however, there is no consensus about why some sets of states are not events. Savage (1972) thus insists the event space to be the powerset, but for a technical need to define a countably additive probability measure, the event space is required to be smaller as in Arrow (1966). Shafer (1986) interprets this restriction as *complexity of describing states, thus of comparing acts*. Villegas (1964), implicit though, takes this point by taking events to be a primitive of uncertainty. Taking Shafer's point of view, we interpret the event space to be the collection of sets of states which the DM is able to recognize.[10] Accordingly, sets of states lying outside the event space are not recognizable to the DM. As the DM cannot believe/know those that he cannot recognize, we may preclude those sets of states from being the subjects of belief and thus of knowledge.

Now, we define an informed event.

Definition 6 (Informed event)**.** For a belief operator $B : \mathcal{F} \longrightarrow \mathcal{F}$, an event $E \in \mathcal{F}$ is an *B-informed event* if $B(E) \cup B(\neg E) = \Omega$. Similarly, for a knowledge operator K, an event $E \in \mathcal{F}$ is said to be *K-informed event*. A DM is said to be *B-informed (K-informed, resp.) about an event E at ω* if E is an B-informed (K-informed, resp.) event and $\omega \in B(E)$ ($\omega \in K(E)$, resp.).

The above definition draws a distinction between what one knows/believes and what one is informed about. Although he knows/believes the event, he may not be informed about it. For him to be informed, he must know either the

[10] This interpretation is similar to the view in Heifetz et al. (2006). They consider events to be "those that can be "known" or be the object of awareness." For more discussion about the conception of an event, see Al-Najjar (2009).

event occurs or not at any state. This requires that the DM has *counterfactual* knowledge/belief about the event. To illustrate this possibilities, consider a variant of Example 2.3 in Halpern (1999).

Example 3. Bob is in a room with the light on. The door is painted either red or blue, and he can tell which color. However, he might not have distinguished the colors, had the room been dark. Formally, there are four states, $\{(red, off), (blue, off), (red, on), (blue, on)\}$, where (red, off) denotes a state in which the door is red and the light is off, and the other states can be similarly interpreted. Let RED, $BLUE$, ON, and OFF be the events that the door is red, the door is blue, the light is on, and the light is off. Let K be the knowledge operator describing Bob's knowledge. Then, $K(ON) = ON$, $K(OFF) = OFF$, $K(RED) = \{(red, on)\}$, and $K(BLUE) = \{(blue, on)\}$. Suppose that only the event RED is of an agent's interest, and the realized state is (red, on). As a consequence, Bob *knows* that the event RED occurs. Were the realized state to be (red, off), however, he would have not known that RED occurs, nor does $BLUE = \neg RED$ occur. For $K(RED) \cup K(\neg RED) = ON \neq \Omega$, RED is not an informed event. Therefore, Bob is *not* informed of the event RED.

In this example, an event ON is a K-informed event. At the state (red, on), an agent knows that the light is on. In addition, he would know whether the light is on or off, even in his imagination that any other state might have occurred.

The following lemma shows that a K-informed event is sufficient for a DM to know itself. In this sense, a K-informed event represents information.

Lemma 1. *Let E be a K-informed event. Then, E is self-evident*[11]*: $E = K(E)$.*

[11] This term originates in Aumann (1999a). Whenever a self-evident event occurs, it informs the DM of its occurrence. The self-evident event, therefore, *is* the knowledge about itself.

Proof. Suppose that E is a K-informed event, i.e. $K(E) \cup K(\neg E) = \Omega$. By **A5**, $K(E) \subseteq E$, so it suffices to show that $E \subseteq K(E)$. By **A2**, $K(E) \cap K(\neg E) = \emptyset$ and thus $\neg K(E) = K(\neg E)$. Again by **A5**, $\neg K(E) = K(\neg E) \subseteq \neg E$. Thus, $E \subseteq K(E)$. □

By definition of knowledge and belief, it is easy to see that a K-informed event is a B-informed event, but not every B-informed event is a K-informed event. Moreover, a B-informed event is not necessarily self-evident.

1.3 Representation of Information as Knowledge

We first present a well-known result on the relationship between a partition and a knowledge operator.

Lemma 2. *For a partition $\Pi \in P$, define $K_\Pi(E) = \{\omega | \Pi(\omega) \subseteq E\}$ for each $E \in \mathcal{F}$. Then, K_Π satisfies **A1-A5**. For an operator $K : \mathcal{F} \to \mathcal{F}$ satisfying **A1-A5**, define a partition $\Pi_K = \{\Pi_K(\omega) | \omega \in \Omega\}$, where $\Pi_K(\omega) = \cap\{E \in \mathcal{F} | \omega \in K(E)\}$. Then, $\Pi = \Pi_{K_\Pi}$.*

For the proof, see Aumann (1999a). According to the above lemma, a K-informed event can be defined with respect to a partition in the following way: E is a K-informed event with respect to a partition Π if $E = K_\Pi(E)$. By adapting the notion of a K-informed event to a partition, we can compare our notion of a K-informed event directly with HM's notion of an informed set. For comparison, we present HM's notion of an informed set.

Definition 7 (Informed set in HM)**.** A set $E \subseteq \Omega$ is an informed set defined

by a partition Π if for every $F \in \Pi$, either $F \subseteq E$ or $F \subseteq \neg E$. The collection of informed sets of Π is denoted by $\mathcal{I}_\Pi$.

The definition of an informed set by HM is related to ours by the following lemma. Let $\mathcal{F}_\Pi$ denote the collection of K-informed events adapted to a partition Π.

Lemma 3. *Let $(\Omega, \mathcal{F})$ be given. For a partition Π, let $\mathcal{I}_\Pi$ denote a collection of its informed sets defined by HM, and let $\mathcal{F}_\Pi$ denote a collection of its K-informed events. Then, $\mathcal{F}_\Pi = \mathcal{I}_\Pi \cap \mathcal{F}$. Moreover, $\mathcal{F}_\Pi$ is a sub-σ-algebra of $\mathcal{F}$.*

The collection of informed sets by HM does not have to be a sub-σ-algebra. That is, there may exist an informed set that is non-measurable.

Example 4. Let $\Omega = [0, 1]$ equipped with a Borel σ-algebra, and let μ be the Borel measure defined on it. Let $\Pi = \{\{\omega\} | \omega \in \Omega\}$ be a partition that contains all singletons. Then, the collection of its informed sets $\mathcal{I}_\Pi$ is the powerset. Obviously, this is larger than the Borel σ-algebra, and contains a well-known non-measurable set, so-called Vitali set. See Royden (1988) for its definition.

Now, we investigate the relationship between a knowledge operator and a σ-algebra. From the discussion on partitions, one can easily see that a knowledge operator defines a σ-algebra. What is not clear is whether a σ-algebra may define a knowledge operator. For our purpose, we need the following definition.

Definition 8 (Countably generated σ-algebra)**.** A sub-σ-algebra $\mathcal{G}$ is *countably generated* if there is a collection of countably many events $\mathcal{U} = \{E_i | i \in \mathbb{N}\}$ such that $\mathcal{G}$ is the smallest σ-algebra containing $\mathcal{U}$.

We show that a countably-generated σ-algebra also represents information as knowledge.

Lemma 4. *Let $\mathcal{G}$ be a countably-generated sub-σ-algebra. Define for an event $E \in \mathcal{F}$,*

$$K(E) = \cup\{G \in \mathcal{G} | G \subseteq E\}.$$

Then, K is indeed a knowledge operator. Moreover, every event in $\mathcal{G}$ is a K-informed event, i.e. $K(G) = G$ for every $G \in \mathcal{G}$.

Proof. To show that K is a knowledge operator, it suffices to show **A1**,**A4** and **A5**, because they implies the rest (Bacharach, 1985). For **A1**, let $(E_i)_{i \in I}$ be given for a countable index set I. Then, $\cap_{i \in I} K(E_i) = \cup\{\cap_{i \in I} G_i \in \mathcal{G} | G_i \subseteq E_i, \forall i \in I\} = \cup\{\cap_{i \in I} G_i \in \mathcal{G} | \cap_{i \in I} G_i \subseteq \cap_{i \in I} E_i\} = K(\cap_{i \in I} E_i)$. For **A4**, since a countably-generated σ-algebra $\mathcal{G}$ is a sub-σ-algebra of a strongly Blackwell σ-algebra $\mathcal{F}$, it is closed under complements and arbitrary unions, and thus $\neg K(E) \in \mathcal{G}$ holds. Then, $K(\neg K(E)) = \cup\{G \in \mathcal{G} | G \subseteq \neg K(E)\} = \neg K(E)$. Lastly, **A5** and the last claim that $K(G) = G$ for $G \in \mathcal{G}$ trivially follow from the definition of K. □

As both partitions and countably-generated σ-algebras represent information as knowledge, one may wonder whether they can be always equivalently used. Unfortunately, however, this is not true.

Example 5. Let $\Omega = [0, 1]$ endowed with a Borel σ-algebra $\mathcal{F}$. Let μ be the Borel measure. Define a mapping $\phi : [0, 1] \to [0, 1]$ such that for $\omega \in [0, 1]$, $\phi(\omega) = \omega + \alpha$ if $\omega + \alpha \leq 1$ and $\phi(\omega) = \omega + \alpha - 1$ if $\omega + \alpha > 1$, where α is an irrational number. Let $\omega \sim \omega'$ be an equivalence relation on $[0, 1]$ so that $\omega \sim \omega'$

if and only if $\phi^n(\omega) = \omega'$ for some $n \in \mathbb{N}$. Then, $\Pi(\omega) = \{\omega' | \omega' \sim \omega\}$ is countable and dense in $[0,1]$. Moreover, the collection of these subsets $\Pi = \{\Pi(\omega) | \omega \in [0,1]\}$ is a partition of Ω. The informed events of this partition are well-known to be ϕ-invariant measurable subsets of Ω and they have either measure 0 or measure 1 (Cornfeld et al., 2012).[12] Then, the collection of informed events $\mathcal{F}_\Pi$ contains an atom of measure 1, which cannot be an element of Π, and thus it is not countably-generated. Moreover, a partition Π' generated by $\mathcal{F}_\Pi$ is not the same as the partition Π.

The above example illustrates that if the collection of K-informed events from a partition is not countably-generated, the partition generated by such a σ-algebra does not preserve K-informed events when one goes from a σ-algebra to a partition. Therefore, we restrict our attention to partitions whose collections of K-informed events are countably-generated σ-algebras.

Definition 9. A partition Π is said to be *measurable* if $\mathcal{F}_\Pi$ is countably-generated.

Let Σ^c be a sub-collection of Σ such that it contains all countably-generated sub-σ-algebras. We naturally endow Σ^c with the partial ordering $\succeq^\sigma$ restricted to Σ^c. With a slight abuse of notations, write it also as $\succeq^\sigma$. Then, $(\Sigma^c, \succeq^\sigma)$ is a poset. Let P^M denote a collection of all measurable partitions of Ω, endowed with a partial ordering $\succeq^P$ restricted to P^M. Then, $(P^M, \succeq^P)$ is a poset. Now, we have our first main result as follows:

[12] The collection of informed sets suggested by HM consists of ϕ-invariant subsets of Ω. The collection includes non-measurable subsets, and the collection of informed events excludes those non-measurable subsets as it is obvious from Lemma 3.

Theorem 1. *The collection of measurable partitions* $(P^M, \succeq^P)$ *and the collection of countably-generated sub-*σ*-algebras* $(\Sigma^c, \succeq^\sigma)$ *are order-isomorphic: Define* $\Phi : (P^M, \succeq^P) \to (\Sigma^c, \succeq^\sigma)$ *such that for* $\Pi \in P$, $\Phi(\Pi) = \mathcal{F}_\Pi$ *is a collection of informed events. Define* $\Psi : (\Sigma^c, \succeq^\sigma) \to (P^M, \succeq^P)$ *such that for a countably-generated sub-*σ*-algebra* $\mathcal{G} \in \Sigma^c$, $\Psi(\mathcal{G}) = \{\mathcal{A}(\omega, \mathcal{G}) | \omega \in \Omega\}$ *is a partition that contains atoms of* $\mathcal{G}$*. Then, the following properties hold.*

(1) Φ *is injective and order-preserving.*

(2) Ψ *is injective and order-preserving.*

(3) $\Phi \circ \Psi = I_{\Sigma^c}$ *and* $\Psi \circ \Phi = I_{P^M}$*, where* I_{Σ^c} *and* I_{P^M} *are the identity functions defined on* Σ^c *and* P^M*, respectively.*

Moreover, the informational content of a measurable partition Π *or a countably-generated sub-*σ*-algebra* $\mathcal{G}$ *is the collection of* K*-informed events, and a* K*-informed set is defined by a knowledge operator* K *deriving from* Π *or* $\mathcal{G}$*.*

Remark 1. *Note that an atom* $\mathcal{A}(\omega, \mathcal{G})$ *of a countably-generated sub-*σ*-algebra is an event (a measurable set) because a countably-generated sub-*σ*-algebra of a strongly Blackwell* σ*-algebra* $\mathcal{F}$ *is closed under arbitrary unions as long as it is measurable with respect to a larger* σ*-algebra. See Remark 3 of HM.*

For comparison, we restate the result of HM in the following.

Lemma 5. *Let* $(P, \succeq^P)$*,* $(\Sigma, \succeq^\sigma)$*, and* $(\Sigma^c, \succeq^\sigma)$ *be given. Define* $\Phi : (P, \succeq^P) \to (\Sigma, \succeq^\sigma)$ *such that for* $\Pi \in P$, $\Phi(\Pi) = \mathcal{F}_\Pi$ *is a collection of informed events. Define* $\Psi : (\Sigma^c, \succeq^\sigma) \to (P, \succeq^P)$ *such that for a countably-generated sub-*σ*-algebra* $\mathcal{G}$, $\Psi(\mathcal{G}) = \{\mathcal{A}(\omega, \mathcal{G}) | \omega \in \Omega\}$ *is a partition that contains atoms of* $\mathcal{G}$*. Then, the following holds.*

(1) Φ *is injective and order-preserving.*

(2) Ψ *is injective and order-preserving.*

(3) For $\mathcal{G} \in \Sigma^c$, $(\Phi \circ \Psi)(\mathcal{G}) = \mathcal{G}$, *i.e.* $\Phi \circ \Psi = I_{\Sigma^c}$, *where* I_{Σ^c} *is the identity function defined on* Σ^c.

For proof of Theorem 4, see HM.

Remark 2. *Note that the codomain of* Φ *is* Σ, *not* Σ^c. *Due to the existence of non-measurable partition, as we show in Example 5,* $\Psi \circ \Phi = I_P$ *does not hold. That is,* Φ *cannot have* Ψ *as its inverse, thus* $(P, \succeq^P)$ *and* $(\Sigma, \succeq^\sigma)$ *are not order-isomorphic. The proof of Theorem 4 follows naturally from the above lemma and the definition of a measurable partition.*

We are concluding this section by showing how our result addresses the problem identified in Billingsley's example.

Example 6. Recall that in Billingsley's example, $\Omega = [0, 1]$ endowed with a Borel σ-algebra $\mathcal{F}$. Let Π be the partition that contains every singleton. Then, the collection of K-informed events corresponding to Π consists of every event in $\mathcal{F}$. As the measurable space $(\Omega, \mathcal{F})$ is assumed to be a complete separable metric space, $\mathcal{F}$ is countably-generated. Therefore, the partition Π' generated from $\mathcal{F}$ by collecting all of its atoms is indeed the same as Π.

1.4 Representation of Information as Belief

In this section, we fix $(\Omega, \mathcal{F}, \mu)$, and we additionally assume that $\mathcal{F}$ is a Borel σ-algebra. We first argue that the generical equivalence of σ-algebras as it is

defined in HM indeed represents information as belief, not as knowledge. For this purpose, we present some definitions.

Definition 10 (Generical Equivalence of σ-algebra)**.** Any two sub-σ-algebras $\mathcal{G}$ and $\mathcal{H}$ are *generically equivalent* with respect to a probability measure μ if

(1) for every $G \in \mathcal{G}$, there is $H \in \mathcal{H}$ such that $\mu(G \triangle H) = 0$, and

(2) for every $H \in \mathcal{H}$, there is $G \in \mathcal{G}$ such that $\mu(G \triangle H) = 0$.

Definition 11 (Proper Regular Conditional Probability)**.** Let $(\Omega, \mathcal{F}, \mu)$ and let $\mathcal{G}$ be a sub-σ-algebra. Then, a *regular conditional probability* is a function $Q : \mathcal{F} \times \Omega \to [0, 1]$ satisfying the following:

(1) for each $\omega \in \Omega$, $Q(\cdot, \omega)$ is a probability measure on $\mathcal{F}$.

(2) for each $E \in \mathcal{F}$, $Q(E, \cdot)$ is a version of $p(E|\mathcal{G})$ such that $p(E|\mathcal{G})$ is $\mathcal{G}$-measurable and integrable, and $\int_G p(F|\mathcal{G})d\mu = \mu(F \cap G)$ for all $G \in \mathcal{G}$.

Moreover, the regular conditional probability Q is said to be *proper* if $Q(E, \omega) = \mathbb{1}_E(\omega)$ for each $E \in \mathcal{G}$, where $\mathbb{1}_E(\omega) = 1$ if $\omega \in E$, and 0 otherwise.

By our assumption on the measurable space $(\Omega, \mathcal{F})$, a proper regular conditional probability exists (Blackwell and Ryll-Nardzewski, 1963).[13] Now, we show that one can define a belief operator by a proper regular conditional probability.

Lemma 6. *Let $\mathcal{G}$ be a sub-σ-algebra, and let $Q(E, \omega)$ be a proper regular conditional probability derived from the probability space $(\Omega, \mathcal{F}, \mu)$ and $\mathcal{G}$. Define an*

[13] This reveals why we need to restrict $\mathcal{F}$ to be a Borel σ-algebra, instead of being a strongly Blackwell σ-algebra in this subsection. If $\mathcal{F}$ is not a Borel σ-algebra, a proper regular conditional probability may not exist. See Shortt (1984).

operator $B : \mathcal{F} \to \mathcal{F}$ *such that for each event* $E \in \mathcal{F}$,

$$B(E) = \{\omega \in \Omega | Q_\omega(E) = 1\}.$$

Then, B *satisfies* ***A1-A4****. That is,* B *is a belief operator.*

For the proof, see Brandenburger and Dekel (1987). In the above lemma, $\mathcal{G}$ can be any σ-algebra which, for example, can be the smallest σ-algebra generated by a partition. Therefore, one can always define a belief operator regardless of whether one starts from a partition or from a σ-algebra. Similarly to the case of knowledge, we consider a collection of all B-informed events and denote it by $\mathcal{F}_Q$

In general, B does not satisfy **A5**, i.e. $B(E) \subseteq E$ does not necessarily hold. Therefore, B is not a knowledge operator. Moreover, note that for each $\omega \in \Omega$, Q_ω is not a complete measure on $\mathcal{G}$ as the following example illustrates.

Example 7. Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$, $\mathcal{F} = 2^\Omega$, and a sub-σ-algebra $\mathcal{G} = \{\emptyset, \{\omega_1, \omega_2\}, \{\omega_3\}, \Omega\}$. The probability measure μ is given as $\mu(\{\omega_1\}) = \mu(\{\omega_3\}) = 0.5$. Let $E = \{\omega_2\}$ and $F = \{\omega_1, \omega_2\}$. The posterior beliefs for E and F at ω_3 can be calculated as $Q(F, \omega_3) = Q(E, \omega_3) = 0$. On the measurable space $(\Omega, \mathcal{G})$, Q_{ω_3} is not a complete measure, for $E \notin \mathcal{G}$.

Motivated by this observation, Brandenburger and Dekel (1987) propose the following:

Definition 12 (Posterior Completion)**.** The posterior completion of a σ-algebra $\mathcal{G}$ is the σ-algebra $\hat{\mathcal{G}}$ generated by $\mathcal{G}$ and the class of sets $\{G \in \mathcal{G} | Q(G, \omega) = 0 \text{ for every } \omega \in \Omega\}$. That is, $\hat{\mathcal{G}} = \{G \in \mathcal{G} | Q(G, \omega) = 0 \text{ or } 1 \text{ for every } \omega \in \Omega\}$ and it is said to be the posterior-completed σ-algebra.

Although the definition takes a sub-σ-algebra as primitive, one can take a partition as primitive as well by the following procedure: For a given partition Π, generate the smallest σ-algebra containing the partition cells, say $\mathcal{H}$, and then apply the procedure described in the above definition to obtain the posterior-completed σ-algebra $\widehat{\mathcal{H}}$. Then, the posterior-completed partition $\hat{\Pi}$ is the collection of the atoms of $\widehat{\mathcal{H}}$. As a matter of fact, the posterior completion of a partition is to add in B-informed events. All these imply that the posterior-completed σ-algebra is indeed a collection of all B-informed events. .

Lemma 7. *Let $\mathcal{G}$ be a sub-σ-algebra, and let B be the resulting belief operator (by Lemma 6). The posterior-completed σ-algebra of $\mathcal{G}$ is indeed a collection of B-informed events:*

$$\hat{\mathcal{G}} = \{E \in \mathcal{F} | B(E) \cup B(\neg E) = \Omega\}.$$

By definition of the posterior-completed σ-algebra, the proof is obvious. In Example 7, the posterior-completion leads to the powerset.

Define a binary relation $\sim$ such that for all two sub-σ-algebras $\mathcal{G}$ and $\mathcal{H}$, $\mathcal{G} \sim \mathcal{H}$ if $\hat{\mathcal{G}} = \hat{\mathcal{H}}$. It is not hard to see that this relation is an equivalence relation. That is, the two sub-σ-algebras are considered to be equivalent if their posterior-completed σ-algebras are identical. Now, we connect the notion of generical equivalence to the notion of a posterior-completion.

Lemma 8. *Any sub-σ-algebra is generically equivalent to its posterior-completion with respect to the proper regular conditional probability measure Q.*

Proof. Let $\mathcal{G}$ be a sub-σ-algebra, and let $\hat{\mathcal{G}}$ be its posterior-completed σ-algebra with respect to a proper regular conditional probability Q. Clearly, $\mathcal{G} \subseteq \hat{\mathcal{G}}$. Take

any event $E \in \hat{\mathcal{G}}$. If $E \in \mathcal{G}$, it is trivial. Suppose that $E \notin \mathcal{G}$. Then, for any $\omega \in E$, either $Q(E, \omega) = 0$ or 1. If $Q(E, \omega) = 0$, trivially there exists an empty set in $\mathcal{G}$ satisfying $Q(E \triangle \emptyset, \omega) = 0$. Otherwise if $Q(E, \omega) = 1$, there exists an event $F \in \mathcal{G}$ such that $E \subset F$ and thus $Q(F, \omega) = 1$. Hence, $Q(E \triangle F, \omega) = Q(F \setminus E, \omega) = 0$. □

We are concluding this section by presenting our second main result that establishes an equivalence between partitions and σ-algebras for representing information as belief. Let $P^{pc} = P/\sim$ denote a collection of all posterior-completion of partitions of Ω, endowed with a partial ordering $\succeq^P$ restricted to P^{pc}.[14] Then, $(P^{pc}, \succeq^P)$ is a poset. Similarly, let $\Sigma^{pc} = \Sigma/\sim$ denote a collection of all posterior-completion of sub-σ-algebras, endowed with a partial ordering $\succeq^\sigma$ restricted to Σ^{pc}. Then, $(\Sigma^{pc}, \succeq^\sigma)$ is a poset.

Theorem 2. *The collection of all posterior-completed partitions $(P^{pc}, \succeq^P)$ and the collection of all posterior-completed sub-σ-algebras $(\Sigma^{pc}, \succeq^\sigma)$ are order-isomorphic: Define $\Phi : (P^{pc}, \succeq^P) \to (\Sigma^{pc}, \succeq^\sigma)$ such that for $\Pi \in P$, $\Phi(\Pi) = \mathcal{F}_\Pi$ is a collection of B-informed events. Define $\Psi : (\Sigma^{pc}, \succeq^\sigma) \to (P^{pc}, \succeq^P)$ such that for a posterior-completed sub-σ-algebra $\mathcal{G} \in \Sigma^{pc}$, $\Psi(\mathcal{G}) = \{\mathcal{A}(\omega, \mathcal{G}) | \omega \in \Omega\}$ is a partition that contains atoms of $\mathcal{G}$. Then, the following properties hold.*

(1) Φ is injective and order-preserving.

(2) Ψ is injective and order-preserving.

(3) $\Phi \circ \Psi = I_{\Sigma^{pc}}$ and $\Psi \circ \Phi = I_{P^{pc}}$, where $I_{\Sigma^{pc}}$ and $I_{P^{pc}}$ are the identity

[14] The equivalence relation $\sim$ between any two partitions Π and Π' is defined so that the smallest σ-algebras generated by these partitions, denoted by $\sigma(\Pi)$ and $\sigma(\Pi')$, have the same posterior-completed σ-algebra, i.e. $\sigma(\Pi) \sim \sigma(\Pi')$.

functions defined on Σ^{pc} and P^{pc}, respectively.

Moreover, the informational content of a posterior-completed partition Π or a posterior-completed sub-σ-algebra $\mathcal{G}$ is the collection of B-informed events, and a B-informed set is defined by a belief operator B deriving from Π or $\mathcal{G}$ through a proper regular conditional probability.

Proof. (1) is trivial, for the posterior-completed σ-algebra is the collection of all B-informed events of the posterior-completed partition. For (2), suppose that $\mathcal{G}$ and $\mathcal{G}'$ are two different posterior-completed σ-algebras such that $\mathcal{G} \subseteq \mathcal{G}'$. The corresponding partitions are $\Pi = \{\mathcal{A}(\omega, \mathcal{G}) | \omega \in \Omega\}$ and $\Pi' = \{\mathcal{A}(\omega, \mathcal{G}') | \omega \in \Omega\}$. Take any $\omega \in \Omega$. Then, $\mathcal{A}(\omega, \mathcal{G}') = \cap\{G \in \mathcal{G}' | \omega \in G\} = \cap\{G \in \mathcal{G} \cup \mathcal{H} | \omega \in G\} \subseteq \cap\{G \in \mathcal{G} | \omega \in G\} = \mathcal{A}(\omega, \mathcal{G})$. As to (3), it is easy to see that two different posterior-completed partitions cannot yield the same σ-algebra. Therefore, it suffices to show that two different posterior-completed σ-algebras generate two different partitions. Suppose that $\mathcal{G}$ and $\mathcal{G}'$ are two different posterior-completed σ-algebras. Assume without loss of generality that there exists an event $E \in \mathcal{G}$ but $E \notin \mathcal{G}'$. Suppose to the contrary that the corresponding partitions are the same, i.e. $\Pi = \{\mathcal{A}(\omega, \mathcal{G}) | \omega \in \Omega\} = \{\mathcal{A}(\omega, \mathcal{G}') | \omega \in \Omega\}$. Since Π is the posterior-completed partition, there exists $\omega' \in \Omega$ and $\Pi'(\omega') \subseteq E$ such that $Q'(\Pi'(\omega'), \omega') = 1$ where Q' is the proper regular conditional probability measure defined by $\mathcal{G}'$ together with μ. Then, $Q'(E, \omega') = 1$ because $\Pi'(\omega') \subseteq E$. This implies that $E \in \mathcal{G}'$, for $\mathcal{G}'$ contains every event F such that $Q'(F, \omega') = 1$. This contradicts to the assumption that $E \notin \mathcal{G}'$. □

The above theorem shows that after completing each σ-algebra $\mathcal{G}$ with respect to a proper regular conditional probability measure Q (defined jointly by $\mathcal{G}$ and

μ), the σ-algebra $\mathcal{G}$ uniquely determines a partition Π.

Our result, which is based on the technique of posterior-completion, provide a different result from HM regarding what is a partition that preserves the informational content of the sub-σ-algebra in Billingsley's example.

Example 8. Consider the following σ-algebra $\mathcal{G}$ in Billingsley's example:

$$\mathcal{G} = \{E \in \mathcal{F} | \text{ either } E \text{ or } \neg E \text{ is countable}\}.$$

The posterior-completion of $\mathcal{G}$ is thus $\mathcal{F}$ which is the Borel σ-algebra. The partition generated from this posterior-completed σ-algebra $\mathcal{F}$ is the partition that contains every singleton. This is, in fact, the partition that generates $\mathcal{G}$.

Remark 3. *Recall that in HM, the partition claimed to have the same informational content as $\mathcal{G}$ is the coarsest partition $\Pi' = \{\Omega\}$. Notice that $\mathcal{G}$ is the smallest σ-algebra generated by the finest partition $\Pi = \{\{\omega\} | \omega \in \Omega\}$. As $\mathcal{G}$ contains every singleton, a DM can distinguish each state from the other. This is the information that $\mathcal{G}$ inherits from the partition Π. However, HM's treatment of $\mathcal{G}$ ignores this information, while focusing solely on the information provided by the uniform probability distribution. On the other hand, our treatment requires the informational content of $\mathcal{G}$ to come from both the partition Π* and *the uniform probability distribution conditioned on Π, as one usually defines a conditional probability. The information contained in Π is not lost, thus implying that the DM is fully informed of which state occurs. Hence, the informational content of $\mathcal{G}$ must be equal to the underlying event space, which is the Borel σ-algebra $\mathcal{F}$.*

1.5 Conclusion

In this paper, we establish an equivalent relationship between partitions and σ-algebras as formalizations of information, and equip the notion of an informational content with a precise and intuitive meaning by viewing it through the two different but related notions of knowledge and belief. Although both a partition and a σ-algebra have been prevalently used to formally represent information, there has only been a vague understanding about the relationship between the two. However, Billingsley (1995) and Dubra and Echenique (2004) raise a concern about the use of σ-algebra by coming up with an example in which a partition and the σ-algebra generated by it fail to contain the same informational content.

Hervés-Beloso and Monteiro (2013) engage this example and elaborate on the meaning of information. They provide a notion of an informed set, and suggest the two alternative methods: one for generating a σ-algebra from a partition and the other for going in the opposite direction. However, we find out that their suggestion still leaves the meaning of information ambiguous. When it comes to a partition, the information content captured by the notion of an informed set depends neither on a given measurable space nor on a probability measure. On the other hand, for a given σ-algebra, the informational content, in general, relies on a specific choice of a probability measure. Even when information content is captured by a countably-generated σ-algebra, HM are silent about whether or not it is a collection of all informed sets for some partition.

By separating the notion of information into the two notions of knowledge

and belief, we elaborate on the meaning of information in relation to a probability measure. The two notions are distinct regarding whether the concept of information is required to satisfy the truthfulness or not. If one allows for falsity, the notion one works with is that of belief. We show that a proper regular conditional probability, and the posterior completion of a partition/a σ-algebra correspond to this conception of information. Specifically, the presence of null events captures the possible falsity of information. Based on the conception of information as belief, we show that partitions and σ-algebras can be equivalently used after applying the technique of the posterior-completion proposed by Brandenburger and Dekel (1987). The idea behind posterior completion is to add in null events to a partition (or a σ-algebra) to generate a new partition (a new σ-algebra) that allows a DM to incorporate the possibility of falsity in his information. On the other hand, if the concept of information is based on knowledge, information must be independent of one's belief (which is captured by a probability measure). In this case, we show that only measurable partitions and countably-generated σ-algebras can be equivalently used.

We conclude that although the distinction between knowledge and belief matters for the equivalence between partitions and σ-algebras when formalizing information either by a partition or by a σ-algebra, one can safely assume information as belief in a practical sense. In almost all economic models, a partition or a σ-algebra is equipped with a probability measure to formalize information of a DM. Therefore, the only thing one needs to make sure is to apply posterior completion before using a partition or a σ-algebra to analyze the problem in his hand.

Chapter 2

Common Knowledge and Efficiency with Incomplete Information

2.1 Introduction

The question of whether individuals are able to achieve an efficient allocation has been central in economics. With complete information, the answer is likely to be positive. If a currently given allocation, the so-called status quo allocation, is inefficient, some individual may propose another allocation that would make him better off without making others worse off. As Coase (1960) argues, other individuals would accept the proposal unless the cost of bargaining is substantial.

In an economy with incomplete information, the conclusion is less likely to be true. Even when the proposed allocation leads to a Pareto improvement, individuals may reject the proposal. Behind this seemingly puzzling argument lies the possibility that the act of proposing itself reveals the proposer's private information, and thus reverses the preferences of individuals. Noticing this

possibility, Wilson (1978) proposes two notions of efficiency in economies with incomplete information by requiring both notions to satisfy no revelation of private information. That is, an allocation is efficient unless there exists a common knowledge event on which another allocation Pareto-dominates it.

Holmström and Myerson (1983) (henceforth, HM) elaborate on Wilson's notions further by identifying three different issues embedded in them: one about defining the notion of efficiency in the presence of privation information, another about whether each individual would follow a decision rule sincerely, so-called incentive compatibility, and the last one about the condition under which an incentive compatible and efficient (in short, incentive efficient) decision rule[1] would be implemented without revealing any private information of individuals.

Using this tripartite separation, HM define the notion of efficiency analogously to the case with complete information. A decision rule is efficient if there is no other decision rule that every individual, conditional on her private information, prefers to it. Accordingly, to ask whether a decision rule is incentive efficient is not the same as to ask whether such a decision rule can be implemented without the possibility of information revelation. In other words, there may exist an incentive efficient decision rule that is implementable through some information revelation.

Surprisingly, however, HM show that *an incentive compatible decision rule is incentive efficient if and only if there does not exist any common knowledge event that such a decision rule is dominated by another incentive compatible*

[1]In the interim stage, each individual does not know the others' information, thus what an individual proposes is not merely an allocation, rather a decision rule that specifies an allocation for each state of information that all individuals might have privately.

decision rule[2]. This implies that checking for efficiency on common knowledge events is sufficient to determine an incentive efficient decision rule. This saves one the effort of considering all possible events. Due to this advantage, it has been widely utilized in various contexts. For example, Vohra (1999) states the following:

> "...it is enough for the objecting coalition to be able to improve upon the status-quo over a discernible event[3]. For the grand coalition,...The argument follows from Theorem 1 of Holmström and Myerson (1983).[4]"

In this paper, we find that the definition of common knowledge event, which HM use to prove their result, is unnecessarily restrictive in light of the standard definition originating in Aumann (1976). Specifically, by applying Brandenburger and Dekel (1987) (henceforth BD)'s definition of common knowledge events[5], we show that there are more common knowledge events that are not accounted for in HM's definition. The class of common knowledge events is larger with BD's definition than with HM's. This naturally leads to a question whether the assertion in Theorem 1 of HM still holds if BD's definition of common knowledge events applies. We argue that the answer is positive.

Replacing HM's definition of common knowledge events by BD's may raise

[2]This is the statement of Theorem 1 of HM.

[3]A discernible event in Vohra (1999) is equivalent to a common knowledge event in Holmström and Myerson (1983).

[4]Vohra (1999), p.130

[5]Our specific choice of BD's definition comes out of our concerns that Aumann's definition does not admit a direct comparison with HM's definition. The latter depends on a probability measure, while the former does not. BD extend Aumann's definition to accounts for a probability measure, thus addressing our concerns. Due to its dependence on a probability measure, BD's definition is often referred to as *common belief with probability* 1 to differentiate it from Aumann's definition.

concerns that it actually weakens HM's result by increasing the burden of checking for efficiency. However, we argue that such apparent burdens can be safely disregarded. We need only check for efficiency in a *strict* subset of common knowledge events known as self-evident events. Furthermore, the class of self-evident events is the *minimal* class of events that one needs to check. When applying BD's definition to HM's model of an economy with incomplete information (which we refer to as HM economy), a self-evident event is the smallest event among all the common knowledge events containing a set of non-null states. This implies that any common knowledge event larger than a self-evident event necessarily contains a null state. If one individual proposes a change from the status quo decision rule to an incentive compatible decision rule that makes himself better off without hurting the others at such a null state, then the other individuals would come to know the proposer's type immediately.

This paper is organized as follows. In Section 2, we begin with a description of the economy with incomplete information as suggested by HM. We also present BD's definition of common knowledge events, and in Section 3, by applying it to HM economy, we compare HM's definition with BD's. In Section 4, we present our main result. Finally in Section 5, we conclude.

2.2 Preliminaries

2.2.1 Environment

In this section, we present the description of an economy with incomplete information and the notion of an incentive efficient decision rule with the relevant definitions by closely following HM.

Economy Let $I = \{1, 2, ..., N\}$ be a nonempty finite set of agents. Each agent i has private information, *type* which takes a value from a finite set T_i. An information state (or simply a state) is thus a type profile $t \in T = \prod_{i \in I} T_i$. Let T_{-i} be a set defined by $T_{-i} = \prod_{j \neq i} T_j$. Let $\mathcal{F}$ be a σ-algebra on T, which we refer to as an event space, and let p_i be a prior probability measure associated with agent i. Then, $p_i(E)$ denotes the prior belief of agent i about an event $E \in \mathcal{F}$. For notational convenience, we shall write $p_i(t)$ to mean $p_i(\{t\})$. We assume that all the agents agree on events with zero prior probability: For every agent i and an event $E \in \mathcal{F}$, $p_i(E) = 0$ implies that $p_j(E) = 0$ for all $j \neq i$. Therefore, we shall refer to an event that occurs with zero probability as a null event.

For a type profile $t = (t_i, t_{-i}) \in T$, agent i cannot distinguish type profiles $\hat{t} = (t_i, \hat{t}_{-i})$. We thus define agent i's information partition $\mathcal{P}^i = \{P^i(t) | t \in T\}$ such that $P^i(t) = \{\hat{t} \in T : \hat{t} = (t_i, \hat{t}_{-i})\}$. The partition cell $P^i(t)$ is the set of states indistinguishable to agent i. In other words, the agent at least knows that the true state does not lie outside $P^i(t)$.

Let $\mathcal{F}^i$ be the smallest σ-algebra generated by $\mathcal{P}^i$, and $q_i : \mathcal{F} \times T \to [0, 1]$ be a conditional probability measure. Then, $q_i(E, t)$ denote agent i's interim belief about how likely an event E is to occur at a state t, and this can be calculated by Bayes' rule whenever applicable: $q_i(E, t) = \frac{p_i(P^i(t) \cap E)}{p_i(P^i(t))}$ if $p_i(P^i(t)) \neq 0$.

Note that $q_i(E, t)$ can be an arbitrary number in $[0, 1]$ if $p_i(P^i(t)) = 0$, i.e. prior belief about agent i's type being t_i is zero. However, there are cases where assigning an arbitrary number is somewhat counterintuitive. Consider a case where $P^i(t) \subseteq E$. Once agent i's type is realized to be t_i, the agent knows for sure that the event E occurs. Naturally, this intuition tells that $q^i(E, t) = 1$.

Moreover, if $P^i(t)$ lies outside E, then it seems that we must specify $q^i(E,t)=0$ because if the state t were to realize, the agent would know for sure that E does not occur. Therefore, we formally impose the following property on $q^i(E,t)$:

Assumption 1. A conditional probability q^i of an agent i is **proper**: for each $t \in T$, $q^i(E,t) = 1_E(t)$ for each $E \in \mathcal{F}^i$, where $1_E(t) = 1$ if $t \in E$ and 0 otherwise.

For notational convenience, we denote $p_i(\hat{t}_{-i}|t_i)$ to be the proper conditional probability that i would assign to a singleton event $\hat{t} = (t_i, \hat{t}_{-i})$ if her own type is t_i and the realized state is t, i.e. $p_i(\hat{t}_{-i}|t_i) = q^i(\{\hat{t}\}, t)$[6].

Let D_0 be a finite set of feasible decisions, and let D be the set of probability distributions over D_0. The preference of each agent $i \in I$ is given by von Neumann-Morgenstein utility function $u_i(\cdot, t) : D \longrightarrow \mathbb{R}$. Then, the economy is completely specified by a list $(I, D_0, \{T_i\}_{i\in I}, \{p_i\}_{i\in I}, \{u_i\}_{i\in I})$.

Incentive efficient decision rule Let $\delta : T \longrightarrow D$ be a decision rule, and let Δ be a collection of decision rule. Then, the payoff of an agent i of type t_i under a decision rule δ is defined as $U_i(\delta|t_i) = \sum_{t_{-i}\in T_{-i}} p_i(t_{-i}|t_i)u_i(\delta(t), t)$. We first introduce the following: A decision rule γ dominates δ at t if $U_i(\gamma|t_i) \geq U_i(\gamma|t_i)$ for all $i \in I$ and $U_j(\gamma|t_j) \geq U_j(\gamma|t_j)$ for some j at t. Moreover, for a nonempty event $R \subseteq T$, γ dominates δ within R if it dominates at every $t \in R$. If $R = T$, we say simply that γ dominates δ. Given the notion of dominance, one may define interim efficiency simply by a undominated decision rule as in the case of complete information. For convenience, we shall drop the term 'interim' unless it is necessary in all what follows.

[6]One may take $p_i(\hat{t}_{-i}|t_i)$ as primitive, and define $q_i(E,t) = \sum_{\hat{t}_{-i}\in E\cap P^i(t)} p_i(\hat{t}_{-i}|t_i)$.

Definition 13 (Efficiency)**.** A decision rule δ is *efficient* (in the interim sense) if there is no decision rule γ that dominates δ.

Definition 14 (Incentive Compatibility)**.** A decision rule δ is said to be incentive compatible for i if

$$U_i(\delta|t_i) \geq U_i(\delta, \hat{t}_i|t_i) \equiv \sum_{t_{-i} \in T_{-i}} p_i(t_{-i}|t_i) u_i(\delta(t_{-i}, \hat{t}_i), t) \text{ for } \forall t_i \in T_i, \forall \hat{t}_i \in T_i.$$

Moreover, a decision rule δ is incentive compatible if δ is incentive compatible for all $i \in I$.

For later use, we shall denote the set of incentive compatible decision rules by $\Delta^* \subset \Delta$.

Definition 15 (Interim Efficient Decision Rule)**.** A decision rule δ is incentive efficient if δ is incentive compatible and efficient.

2.2.2 Knowledge, Common Knowledge, and Self-Evident Event

In this subsection, we introduce essential concepts regarding a common knowledge event following mainly Brandenburger and Dekel (1987). Before we begin, recall that we are given a measurable space $(T, \mathcal{F})$ equipped with each agent i's (prior) probability measure p_i. The information structure of an agent i is given by a partition $\mathcal{P}^i$ or, equivalently, by the smallest σ-algebra $\mathcal{F}^i$ by the partition. Moreover, we denote a proper regular conditional probability by q_i.

Consider an event $E \in \mathcal{F}$ and a state $t \in T$. We shall formalize a sentence like "An agent i knows E at t" by $q_i(E, t) = 1$. We wish to emphasize that the presence of null events may cause a trouble in appropriately defining the notion

of an common knowledge event. This is exactly the concern Brandenburger-Dekel address. They argue that one need to add in events that are null in the sense of proper regular conditional probability, to the information partition or the corresponding σ-algebra. This requirement is often called as posterior completion. We formally state it as follows:

Definition 16 (posterior completion)**.** *The posterior completion* of a σ-algebra $\mathcal{F}^i$ is a σ-algebra generated by $\mathcal{F}^i$ and the class of events $\{G \in \mathcal{F} | q_i(G,t) = 0 \text{ for every } t \in T\}$. The resulting σ-algebra is said to be *the posterior-completed σ-algebra*, and denoted by $\hat{\mathcal{F}}^i$.

One may define the posterior completion of a partition as follows: Let $\mathcal{P}^i$ be a partition, which is the collection of atoms of a σ-algebra $\mathcal{F}^i$. Then, the posterior completion of $\mathcal{P}^i$ is simply the collection of atoms of the posterior completed σ-algebra $\hat{\mathcal{F}}^i$, and denoted by $\hat{\mathcal{P}}^i$.

Definition 17 (Knowledge)**.** For a probability space $(T, \mathcal{F}, p_i)$, define a function $K_i : \mathcal{F} \longrightarrow \hat{\mathcal{F}}^i$ such that for every event $E \in \mathcal{F}$,

$$K_i(E) = \{t | q_i(E,t) = 1\}.$$

Then, K_i is said to be *a knowledge function.* Moreover, an agent $i \in I$ is said to *know* that an event E occurs at a state t if $t \in K_i(E)$.

We say that an event $F \in \mathcal{F}$ is non-null in a posterior sense if for an agent i, $q_i(F,t) > 0$ for every $t \in T$.

Definition 18 (self-evident event and common knowledge event)**.** An event $F \in \mathcal{F}$ is said to be *self-evident* if $K_i(F) = F$ for all $i \in I$. That is, F is a

non-null (in a posterior sense) member of $\cap_i \hat{\mathcal{F}}^i$. For $t \in T$, an event E is *a common knowledge event* at a state t if there is a self-evident event F such that $t \in F$ and $F \subseteq E$[7].

We may state the above definition in terms of partitions. To do this, we define the following: A partition $\mathcal{P}$ is a coarsening of a partition $\mathcal{P}'$ (or $\mathcal{P}'$ is a refinement of $\mathcal{P}$) if for each $P_k \in \mathcal{P}$, there exists a set $\kappa \subseteq \{1, 2, ...\}$ such that $\{P'_m\}_{m \in \kappa}$ constitutes a partition of P_k. The *join* is the coarsest common refinement of partitions $\{\mathcal{P}^i\}_{i \in I}$, and denoted by $\vee_{i \in I} \mathcal{P}^i$. The *meet* is the finest common coarsening of partitions $\{\mathcal{P}^i\}_{i \in I}$, and denoted by $\wedge_{i \in I} \mathcal{P}^i$.

Lemma 9. *An event $F \in \mathcal{F}$ is self-evident if and only if it is a non-null (in a posterior sense) member of the meet $\wedge_{i \in I} \hat{\mathcal{P}}^i$.*

The proof is trivial by the relationship between a posterior-completed partition and a posterior-completed σ-algebra.

2.3 Characterization of Common Knowledge Event

In an economy $\mathcal{E}$, characterizing self-evident events and common knowledge events according to the definitions in the previous section is cumbersome. We thus characterize them in terms of $p_i(\hat{t}_{-i}|t_i)$ for direct comparison with HM.

Lemma 10. *An event F is a self-evident event if and only if for any $t = (t_i, t_{-i}) \in F$ and any $i \in I$, F satisfies the following:*

[7] In the original definition by BD that also allows for an infinite (possibly uncountable) number of states, the self-evident event F needs to be contained in the event E almost surely, i.e. $q_i(F \setminus E, t) = 0$. However, in the HM economy where the state space is finite, this condition is equivalent to the condition that the self-evident F is a subset of E.

1. *Posterior beliefs are zero outside* F *:* $p_i(\hat{t}_{-i}|t_i) = 0$ *for* $\forall \hat{t} = (t_i, \hat{t}_{-i}) \notin F$*, and*

2. *Posterior beliefs are non-zero within* F *:* $p_i(\tilde{t}_{-i}|t_i) > 0$ *for* $\forall \tilde{t} = (t_i, \tilde{t}_{-i}) \in F$*.*

E is a common knowledge event at a state t if and only if there is a self-evident event F such that $t \in F$ and $F \subseteq E$. Moreover, we say that E is a common knowledge event *if there exists $t \in E$ at which it is a common knowledge event.*

Proof. If there exists no null events, it is trivial. Hence, suppose that there exists a null event, i.e. there exists a state $t \in T$ such that $p_i(t) = 0$ for all i. Take any $t \in F$ and any $i \in I$. Suppose that both conditions hold. Then, $q_i(F, t) > 0$ by the second condition. Moreover, $q_i(F, t) = 1$ by the first condition. Hence, F is a non-null member of the meet,i.e., self-evident event. For the other direction, suppose that F is a self-evident event. Take any $t \in F$ and any $i \in I$. Then, $q_i(F, t) = 1$. Take any $\tilde{t} = (t_i, \tilde{t}_{-i}) \in F$. Then, $\tilde{t} \in \hat{P}^i(t)$, which implies the second condition. Take any $\hat{t} = (t_i, \hat{t}_{-i}) \notin F$. $\hat{t} \in P^i(t)$ implies that $q_i(\{\hat{t}\}, t) = 0$. This satisfies the first condition. The characterization of a common knowledge event is obvious by its definition. □

One should be cautious that a common knowledge event E may not be a common knowledge event at some state $t \in E$.

Corollary 1. *Let $t \in T$ be given, and let E be a common knowledge event. Then, the event E is a common knowledge event at t if and only if $t \in F$ for some self-evident event $F \subseteq E$. That is, if the state lies outside any self-evident events, the event E is not a common knowledge event at such a state t.*

Proof. By definition of a common knowledge event, "*If*" part is trivial. For the other direction, suppose that the state t does not belong to any self-evident event F contained in E, i.e. $t \notin F$ for all $F \subseteq E$. Suppose further to the contrary that E is a common knowledge event at t. Then, by Lemma 10 above, there exists a self-evident event F' such that $t \in F'$ and $F' \subseteq E$. This leads to a contradiction that t does not belong to any self-evident event. □

The above corollary implies that a self-evident event contains all the information required to determine a common knowledge event.

Now, we compare our characterization with the one proposed by HM. For this purpose, we present HM's characterization of common knowledge events (Lemma 1 of HM) in the name of HM common knowledge to avoid the confusion with ours.

Definition 19 (HM Common Knowledge Event)**.** An E is a *common knowledge event in the sense of HM*, or simply *HM common knowledge event*, if and only if E is of the form $E = \prod_{i \in I} E_i$, where each $E_i \subseteq T_i$, and

$$p_i(\hat{t}_{-i}|t_i) = 0 \text{ for } \forall t_i, \forall \hat{t} = (t_i, \hat{t}_{-i}) \notin E, \forall i.$$

The above condition implies that a HM common knowledge event E is a rectangular event satisfying that the conditional probability is degenerate on for every event F such that $F \cap E = \emptyset$ and $F_i = E_i$ for some $i \in I$. Notice that it requires no condition inside the event.

It is immediate that if an event is a HM common knowledge event, then it is a common knowledge event. However, the converse does not hold. We provide an example showing that there exists a common knowledge event that is not rectangular.

Example 9. Consider the case where there are two agents and two types for each agent. Then, $T = \{11, 12, 21, 22\}$ where $mn = (t_1^m, t_2^n)$. Let the prior probabilities for each agent be $p_i(\{22\}) = 0$ for all $i = 1, 2$. Then, the partitions of agents are given as $\mathcal{P}^1 = \{\{11, 12\}, \{21, 22\}\}$ and $\mathcal{P}^2 = \{\{11, 21\}, \{12, 22\}\}$. By posterior completion, $\{22\}$ should be added in to both agents' partitions. Then, the posterior-completed partitions and their meet are the followings:

$$\hat{\mathcal{P}}^1 = \{\{11, 12\}, \{21\}, \{22\}\}, \quad \hat{\mathcal{P}}^2 = \{\{11, 21\}, \{12\}, \{22\}\}, \text{ and } \hat{\mathcal{P}}^1 \wedge \hat{\mathcal{P}}^2 = \{\{11, 12, 21\}, \{22\}\}$$

$\{11, 12, 21\}$ is a non-null (in a posterior sense) member of the meet and thus a self-evident event. Denote this event by $E = \{11, 12, 21\}$. Then, E is a common knowledge event at $t = 11$. However, it is not a HM common knowledge event: There is no $E_i \subseteq T_i$ for $i = 1, 2$ such that $E = E_1 \times E_2$.

In relation to a self-evident event, we can also see easily that if an event is a self-evident event, then it is a HM common knowledge event, but the converse does not hold.

Example 10. Consider the same setting as Example 9 except that $p_i(\{t\}) = 0$ for all $t = 12, 21, 22$ and for all $i = 1, 2$. The properness of the posterior probability requires $q_1(\{21, 22\}, t) = 1$ for $t = 21, 22$ and $q_2(\{12, 22\}, t) = 1$ for $t = 12, 22$. Note that neither agent 1's posterior beliefs for any singleton event in $\{21, 22\}$ nor agent 2's posterior beliefs for any singleton event in $\{12, 22\}$ can be calculated by Bayes' rule. Hence, those posterior beliefs can be determined in an arbitrary manner. Assume that $q_1(\{21\}, t) = 0.5$ for $t = 21, 22$ and $q_2(\{12\}, t) = 0.5$ for $t = 12, 22$. In other words, $p_1(t_2 = 1 | t_1 = 2) = 0.5$ and $p_2(t_1 = 1 | t_2 = 2) = 0.5$. Then, $\{12\}$ should be added in to agent 1's partition

while $\{21\}$ should be added in to agent 2's partition. Then, the posterior-completed partitions and their meet are the followings:

$$\hat{\mathcal{P}}^1 = \{\{11\}, \{12\}, \{21, 22\}\}, \quad \hat{\mathcal{P}}^2 = \{\{11\}, \{21\}, \{12, 22\}\}, \text{ and } \hat{\mathcal{P}}^1 \wedge \hat{\mathcal{P}}^2 = \{\{11\}, \{12, 21, 22\}\}$$

$\{11\}$ is a non-null (in a posterior sense) member of the meet and then a self-evident event. Let $E = \{11, 12\}$. E is clearly a HM common knowledge event. However, it is not a self-evident event.

We shall simply summarize the result by the following lemma.

Lemma 11. *Let $\mathcal{S}$, $\mathcal{HM}$, and $\mathcal{C}$ be the collection of all self-evident events, all HM common knowledge events, and all common knowledge events, respectively. Then, we have the following:*

(a) $\mathcal{S} \subset \mathcal{HM} \subset \mathcal{C}$

(b) $\mathcal{S} = \mathcal{HM} = \mathcal{C} = \mathcal{F}$ *if and only if* $p_i(t) > 0$ *for all* $i \in I$ *and all* $t \in T$*, i.e. there exists no null event.*

The proof is obvious by Lemma 10 and Definition 19.

2.4 Common Knowledge and Incentive Efficient Decision Rule

HM in Theorem 1 of their work shows that one needs to inspect HM common knowledge events to find out an incentive efficient decision rule. This is indeed a powerful result for it actually reduces the number of incentive compatible decision rules to consider. Specifically, suppose that we are considering an incentive compatible decision rule δ as a candidate for an incentive efficient decision rule.

By the result proved by HM, it is not necessary to consider an incentive compatible decision rule γ that dominates δ outside common knowledge events. By Lemma 11, we see that this does not work if there exists no null event. In what follows, we thus assume the following:

Assumption 2. There exists a null event $E \in \mathcal{F}$ such that $p_i(E) = 0$ for all i.

Now, we recast Theorem 1 of HM to see how it reduces the number of events to consider to find out an incentive efficient decision rule.

For simplicity of presentation, we shall define the following notation: For an incentive compatible decision rule $\delta \in \Delta^*$ and an event $E \in \mathcal{F}$, let $\Delta^*(\delta, E) \subset \Delta^*$ to denote the set of all incentive compatible decision rules that dominates δ within the event E:

$$\Delta^*(\delta, E) = \{\gamma \in \Delta^* | \gamma \neq \delta \text{ and } \gamma \text{ dominates } \delta \text{ within } E\}$$

Then, it satisfies the following properties: For $E, F \in \mathcal{F}$

- $\Delta^*(\delta, E \cup F) = \Delta^*(\delta, E) \cap \Delta^*(\delta, F)$
- $E \subseteq F$ implies $\Delta^*(\delta, F) \subseteq \Delta^*(\delta, E)$

Theorem 3 (HM Theorem 1)**.** *An incentive compatible decision rule δ is interim incentive efficient if and only if there does not exist any HM common knowledge event E such that δ is interim dominated within E by another incentive-compatible decision rule:*

$$\Delta^*(\delta, T) = \emptyset \iff \cap\{\Delta^*(\delta, E) | E \in \mathcal{HM}\} = \emptyset$$

One can clearly see that $\cap\{\Delta^*(\delta, E)|E \in \mathcal{HM}\} \subset \Delta^*(\delta, T)$ because $\mathcal{HM} \subset \mathcal{F}$. As we argued in the previous section, HM's result is based on a somewhat arbitrarily restrictive definition of common knowledge events. A natural question is whether the same result holds if we replace HM definition with our definition in Lemma 10. The answer is positive.

Theorem 4. *An incentive compatible decision rule δ is interim incentive efficient if and only if there does not exist any common knowledge event E such that δ is interim dominated within E by another incentive-compatible decision rule:*

$$\Delta^*(\delta, T) = \emptyset \iff \cap\{\Delta^*(\delta, E)|E \in \mathcal{C}\} = \emptyset$$

The proof is trivial, for T itself is a common knowledge event. Therefore, the above theorem does not reduce the number of events we need to check for efficiency to determine an incentive efficient decision rule.

Is there any way to reduce the number of events further than to consider HM common knowledge events? The answer is positive as illustrated by the following example:

Example 11. Consider the same setting as in Example 10. Let $E = \{11, 12\}$. Then, E is a HM common knowledge event. Moreover, it is a common knowledge event only at $t = 11$. Suppose that there are only two incentive compatible decision rules, $\Delta^* = \{\delta, \gamma\}$, such that $U_i(\gamma(11), 11) = U_i(\delta(11), 11)$ for all $i = 1, 2$, $U_1(\gamma|t_1 = 2) < U_1(\delta|t_1 = 2)$, $U_2(\gamma|t_2 = 2) > U_2(\delta|t_2 = 2)$, $U_2(\gamma(12), 12) < U_2(\delta(12), 12)$ and $U_2(\gamma(22), 22) > U_2(\delta(22), 22)$. We first argue that within a

HM common knowledge event E, γ dominates δ because

$$U_1(\gamma|t_1 = 1) = U_1(\delta|t_1 = 1), \quad U_1(\gamma|t_1 = 2) < U_1(\delta|t_1 = 2),$$

$$U_2(\gamma|t_2 = 1) = U_1(\delta|t_2 = 1), \quad U_2(\gamma|t_2 = 2) > U_2(\delta|t_2 = 2).$$

However, it suffices to consider $F = \{11\}$. At state $t = 12$, it is not common knowledge that γ dominates δ. Specifically, suppose that the two agents are considering a change from δ to γ. By the specification of the utility function, both agents would gain by agree with the change. Note, however, that E is not a common knowledge event at $t = 12$. Then, if both agents were to agree with the change, agent 2 would know that the agent 1's type is $t_1 = 1$ because if agent 1 with $t_1 = 2$ would have objected the change. Now, $p_2(t_1 = 1|t_2 = 2) = 1$ and agent 2 would want to repeal her consent to the change. Hence, in fear of this, the agent 1 would object to the change when asked for a consent. That is, γ will not be chosen over δ even though the former dominates the latter on E. To reiterate, it suffices to consider $F = \{11\}$.

In the above example, even when E is a HM common knowledge event, it suffices to check for efficiency on a smaller event F to see that γ is incentive efficient. This shows that if there exists a null state $t \in E$ such that E is not a common knowledge event at t, then there is always a room for the possibility that an agent's proposal would reveal his information.

We thus argue that one may determine an incentive efficient decision rule only by checking for efficiency on the self-evident events. Moreover, this is the maximum extent to which one may reduce the number of events for the job of finding out an incentive efficient decision rule.

Theorem 5. *An incentive compatible decision rule δ is interim incentive efficient if and only if there does not exist any self-evident event E such that δ is interim dominated within E by another incentive-compatible decision rule:*

$$\Delta^*(\delta, T) = \emptyset \iff \cap\{\Delta^*(\delta, E) | E \in \mathcal{S}\} = \emptyset.$$

Proof. By $\cap\{\Delta^*(\delta, E) | E \in \mathcal{S}\} \subset \Delta^*(\delta, T)$, "*only if*" part is trivial. For the other direction, suppose that $\cap\{\Delta^*(\delta, E) | E \in \mathcal{S}\} = \emptyset$ but $\Delta^*(\delta, T) \neq \emptyset$. Then, there exists $\gamma \in \Delta^*(\delta, T)$, i.e γ dominates δ. For any self-evident event E, $\Delta^*(\delta, E) = \emptyset$: there exists no incentive compatible decision rule that dominates δ within every self-evident event E. Hence, γ does not dominate δ within $\cup\{E | E \in \mathcal{S}\}$. Then, there exists an agent $j \in I$ and a state $t \notin \cup\{E | E \in \mathcal{S}\}$ such that $U_j(\gamma | t_j) > U_j(\delta | t_j)$. Since the state t lies outside every self-evident event, the posterior probability must be degenerate at t. That is, $p_j(t_{-j} | t_j) = 0$. This implies that $U_j(\gamma | t_j) = p_j(t_{-j} | t_j) u_j(\gamma(t), t) = 0 = U_j(\delta | t_j) = p_j(t_{-j} | t_j) u_j(\delta(t), t)$, which is a contradiction. □

Corollary 2. *The class of self-evident events $\mathcal{S}$ is the minimal class of events among the classes of events $\mathcal{G}$ satisfying the following condition:*

$$\Delta^*(\delta, T) = \emptyset \iff \cap\{\Delta^*(\delta, E) | E \in \mathcal{G}\} = \emptyset.$$

Proof. Suppose that there exists a class of events $\mathcal{G} \subset \mathcal{S}$ satisfying $\Delta^*(\delta, T) = \emptyset \iff \cap\{\Delta^*(\delta, E) | E \in \mathcal{G}\} = \emptyset$, i.e. δ is incentive efficient if and only if there exists no other incentive compatible decision rule that dominates δ within any event $E \in \mathcal{G}$. Then, there exists a self-evident event $F \in \mathcal{S}$ such that $F \notin \mathcal{G}$. Suppose that $\Delta^*(\delta, F) \neq \emptyset$. This implies that there exists an incentive compatible decision rule γ that dominates δ in F. Since F is a self-evident

event, it is also a common knowledge event that γ dominates δ at every $t \in F$. This leads to a contradiction, because δ is not incentive efficient. □

2.5 Conclusion

In this paper, we investigate the idea originating in Wilson (1978) that in order to determine whether an incentive compatible decision rule is efficient or not, one need only check whether it is common knowledge that there exists another incentive compatible decision rule that dominates it. HM show that this idea is indeed valid. By giving a close examination, however, we find that their definition of common knowledge is arbitrarily restrictive, when comparing it with the standard definition of Brandenburger and Dekel (1987). There are more common knowledge events that are not accounted for in HM's definition. This weakens HM's result by increasing the number of events on which we need to check for efficiency in order to determine an incentive compatible and efficient decision rule. However, we argue that HM's result can actually be strengthened. We argue that it is sufficient to consider a strict subset of a common knowledge event, known as self-evident events. Moreover, this is the minimal class of events one need to check. As every self-evident event consists of non-null states, our result suggests that one may safely assume that every state is non-null in a finite state space model (like HM economy), when working with the purpose of studying an incentive compatible and efficient decision rule.

Chapter 3

Mediator Selection in International Conflict: Bias, Effectiveness, and Incidence

3.1 Introduction

Third-party mediation is one of the most commonly used technique for resolving international conflicts (Bercovitch and Gartner, 2008, p.5). In particular, it has become commonplace since the end of World War II (Frazier and Dixon, 2006, p.395)[1]. Growing reliance on mediation as a mean for conflict resolution naturally raises a question of what makes for a successful mediation. Particularly, third party's impartiality, or "even-handedness" has been emphasized to

[1] Although the number varies across the databases, the incidence of the third-party mediation (or simply mediation) accounts for about a 30 to 40 percent rate. The variation depends on the definition of conflicts as well as the time periods considered in databases. For example, the International Crisis Behavior(ICB) database defines a conflict broadly as a situation in which there exists only some perceived threat of increased hostilities. In the ICB database, out of the 434 conflicts that occurred between 1918 and 2001, only 128 conflicts (30 percents) experienced the third-party mediation. The International Conflict Management(ICM) database, however, defines a conflict in a more restrictive sense: it must involve a significant use of force and/or some fatalities. According to the ICM database that identifies 104 bilateral interstate conflicts between 1965 and 1995, the mediation occurred in 40 conflicts.

be crucial by scholars and practitioners[2]: A disputant, facing a mediator who is biased against him, would be less willing to accept the mediator's recommendation. Even worse, anticipating such a circumstance, he would not agree to initiate mediation in the first place. Nonetheless, mediators are often biased: The United States in the Falkland island war and the Soviet Union in the Vietnam war are just two of many available examples.

The literature on this subject provides an explanation by arguing either that a biased mediator may resolve conflicts better than a unbiased one (Kydd, 2003), or that there is a shortage of unbiased third parties (Beardsley, 2006; Beber, 2012). Nevertheless, the literature is silent as to why a disputant accepts a biased mediator in the first place, and behind this silence lies a naive understanding that a disputant would accept mediation if it is likely to be effective. However, peace is not the end itself for a disputant, but merely a mean to increase his own welfare.

The purpose of this paper is to address this issue by answering the following questions: "Why, and under which circumstances would disputants accept a biased mediator? If accepted, is such a mediator as effective in promoting peace as an unbiased one?" To this end, we build a simple model of mediator selection where each disputant, facing a potentially biased mediator, makes a decision to accept mediation or not. If both disputants agree, such a mediator would make a recommendation, as a mechanism designer, to disputants about

[2]For example, the United Nations single out impartiality as a cornerstone of mediation, without which any meaningful resolution of the conflict is hampered. See United Nations, Guidance for Effective Mediation (2012), http://www.un.org/wcm/webdav/site/undpa/shared/undpa/pdf/UN%20Guidance%20for%20Effective%20Mediation.pdf. Moreover, even when individual nations are involved, they usually issue public avowals of impartiality as the United States in the Middle East and in the Falkland Islands Crisis of 1982 (Smith, 1985).

which action to take and how to divide the contested resources between the two parties. Otherwise, disputants are engaged in a situation in which each disputant chooses whether to start a war or not. In either case, each disputant's decision may reveal his private information.

The novelty of our model is to introduce mediator bias. Unlike the case of mechanism design problems that assume a unbiased mediator whose sole purpose is to promote peace (Fey and Ramsay, 2010; Hörner et al., 2015) (henceforth HMS), an optimal mechanism does not treat disputants symmetrically. Therefore, one cannot simply restrict his attention to the type-dependent constraints by ignoring the identities of each disputant. Specifically, when it comes to a disputant favored by a mediator, it is not easy to figure out whether a participation constraint or an incentive compatibility constraint does or does not bind at the optimum.

We begin our analysis by revealing how an optimal mechanism, if proposed by a biased mediator, differs from the one by an unbiased mediator. We find that a biased mediator allocates more resources to her ally, while giving the opponent more chances to enjoy a peaceful outcome (Corollary 3). This differential treatment arises, for a biased mediator must give an information rent to the disfavored party. There are two options: to raise the peace probabilities or to raise the share for the disfavored party. These two options affect the welfare of the favored party differently. The former benefits both parties by saving the resources that might have been wasted under war. The latter, however, does harm to the favored party. The biased mediator would then choose the cheaper way of allocating more peace probabilities to the disfavored party.

When turning to the occurrence of mediation, we find that if the likelihood

of peace is low in a conflict, a biased mediator is accepted by disputants as long as her bias is moderate (Theorem 7). As a conflict is less likely to end up with a peaceful outcome, disputants are willing to accept a mediator with a more extreme level of bias. This is because a weak disputant has a stronger incentive to pretend to be strong in order to induce his opponent not to attack him. This, in turn, increases the amount of an informational rent a biased mediator must provide to the disfavored party under mediation. As mediation is now more attractive, the disfavored party would accept mediation even when the mediator's bias is more extreme.

More importantly, we argue that a biased mediator, accepted by both disputants, is equally effective as an unbiased one (Theorem 6). An immediate implication is that if accepted, the peace probability attained by such a mediator is independent of the intensity of her bias. This is striking because a biased mediator is not interested in the peaceful resolution of a conflict per se unlike an unbiased one. Then, how does a biased mediator end up with achieving peace as effectively as unbiased one? A biased mediator may gain from promoting peace. By saving the resources that might have been wasted under war, she can allocate more resources to the favored party under peace. However, the gains from promoting peace is necessarily followed by the cost of providing a larger amount of the informational rent to the disfavored party. If moderately biased, a mediator finds the cost negligible. Consequently, she ends up with promoting peace as an unbiased mediator does.

We contribute to the literature on mediator bias by showing that mediator bias is not harmful in achieving peace, if one considers the endogenous selection of a mediator by disputants. Although Kydd (2003) reach the similar conclusion,

his result requires that the mediator possesses information that is not available to the disputants[3]. More importantly, Kydd considers a model in which a mediator is exogenously given. Contrarily, our result holds without assuming the private information that a mediator possesses about the disputants. Moreover, as it is highlighted, we contribute to the literature by studying the demand side of mediation that has been largely conceived to play a little role in explaining why a biased third party acts as a mediator.

In relation to the literature that adopts a mechanism-design theoretic approach to mediation, we contribute by investigating how an optimal mechanism changes when the mediator (or the mechanism designer) is biased in favor of one disputant. The literature assumes a unbiased mediator who seeks to maximize the peace probability. As we discussed, the optimal mechanism offered by a biased mediator is qualitatively different from the one by an unbiased mediator. Moreover, we contribute by dealing with the technical challenge as to whether a participation constraint or an incentive compatibility constraint does or does not bind at the optimum, when it comes to a disputant favored by a mediator.

To be specific, the difficulty of solving for an optimal mechanism in our model can be easily seen by comparing the technique used in HMS under the assumption of an unbiased mediator. Our model extends their mechanism design problem by allowing for mediator bias, and thus admits a direct comparison with HMS. By utilizing the unbiasedness of a mediator, HMS imposes symmetry

[3]Kydd challenge the negative view about mediator bias by arguing that if a mediator has a privileged access to the information that one disputant is weaker than his opponent, she may persuade the weaker disputant to make a concession to the opponent by providing such information more credibly when she is biased in favor of the weaker disputant than when she is unbiased. This argument initially comes from Touval (1975), and Kydd formalize the argument by highlighting the role of a mediator in providing information. For more discussions, see Smith (1985), Touval (1975), and Favretto (2009)

on the choice variables while specifying a value to some choice variables through an educated guess. They thus simplify the problem into a linear programming problem in one variable. This technique does not work for our model.

We thus proceed by simplifying the problem into a linear programming problem in four variables that are related only to the peace probabilities. We then treat the problem as if one disputant's type is known, and solve for the two variables as expressions of the other two variables. The expressions are piecewise linear in the two probability variables, and we solve for an optimal mechanism by depicting them geometrically on the space of the two probability variables. Although our model concerns a mediation problem in the context of international conflict, it may also apply to a various bargaining problem with incomplete information that arises in, for example, trade disputes and litigation cases. Therefore, our technique may apply to analyze these problems when the mediator is potentially biased toward one party.

This paper is organized in the following order. Section 2 formally introduces a model of mediator selection. Section 3 presents an optimal mechanism, revealing how differently a biased mediator mediates from an unbiased one. Section 4 , concludes the paper by discussing implications of the main result and potential directions of the future research. Most of the proofs can be found in the Appendix.

3.2 A Model

3.2.1 A Model of Conflict: War-and-Peace game

This subsection presents a simple model of international conflicts, so-called War-and-Peace game, following the setup suggested by HMS(2015). Two countries or disputants $i = 1, 2$ are in conflict with each other over a pie of which size is normalized to one. In peace, each disputant owns the half of the pie. Each disputant may take an action among the two alternatives, attacking his opponent, *'Attack'* and staying in peace, *'Stay'*. Let $a_i \in \{S, A\}$ denote an action taken by disputant i, where A denotes 'Attack' and S denotes 'Stay'. If either disputant chooses to attack his opponent, war breaks out between the two disputants and the size of pie shrinks to $\theta < 1$. Otherwise, both disputants stay in peace thus the size of the pie remains the same.

When war breaks out, its outcome (and how the two disputants share the pie) depends on both disputants' overall strength that reflects comprehensively their military powers, their diplomatic ability, or the aggressiveness of their leaders and citizens. We assume that each disputant i's overall strength is his private information, which we capture by his *type*, $\tau_i \in \{H, L\}$ where H (high) means that disputant i's overall strength is high, and L (low) means that disputant i's overall strength is low. Each disputant is likely to be of high (H) type with probability $q \in (0, 1)$. If both disputants are of the same type, then the war ends in a tie and the countries share the pie of size θ equally, i.e. $\theta/2$. When types are asymmetric, high-type disputant wins and low-type disputant loses. The winner gets the larger share $p > 1/2$ of the pie, while L-type obtains the rest. The corresponding payoffs are $p\theta$ and $(1 - p)\theta$. We assume that H-type

has an incentive to wage a war against low-type: $p\theta > 1/2$. To sum up, the situation is described by the following:

	S	A
S	$1/2, 1/2$	$\theta/2, \theta/2$
A	$\theta/2, \theta/2$	$\theta/2, \theta/2$

(a) Identical Types

	S	A
S	$1/2, 1/2$	$p\theta, (1-p)\theta$
A	$p\theta, (1-p)\theta$	$p\theta, (1-p)\theta$

(b) Different Types: 1 is H, 2 is L

Figure 3.1: Conflict Situation: War-and-Peace Game

Let $\gamma = \frac{q}{1-q}$ denote the likelihood of a disputant believes his opponent to be H-type, and let $\delta = \frac{p\theta - 1/2}{(1-\theta)/2}$. As it is common in the simultaneous-move games, this War-and-Peace game also has multiple equilibria. We assume that both disputants play according to the following equilibrium strategy profile: L-type chooses to stay, and H-type chooses to attack if $\gamma < \delta$ and to stay otherwise[4]. In other words, L-type disputant always prefers peace to war because he has no chance to win. H-type disputant, on the other hand, prefers peace to war if and only if he is more likely to meet the same high type opponent ($\gamma < \delta$). Specifically, the condition implies that the expected payoff under war is larger than the expected payoff under peace: $\gamma < \delta \iff q(\theta/2) + (1-q)p\theta > 1/2$.

The rationale behind our choice of a specific equilibrium is that it is a weakly dominant strategy for L-type to stay, and that an equilibrium strategy profile in which H-type chooses to stay if $\gamma < \delta$ Pareto-dominates all the other equilibria. To make the conflict situation non-trivial, we assume that $\gamma < \delta$ in what follows. Then, war breaks out with probability $1-(1-q)^2$ and peace is attained with the remaining probability. The resulting equilibrium payoff for H-type disputant is his expected payoff under war, $q\left(\frac{\theta}{2}\right) + (1-q)p\theta$. For L-type, his payoff is

[4]For the detailed analysis of War-and-Peace game under arbitrary (possibly asymmetric) beliefs, see Appendix.

$q(1-p)\theta + (1-q)\left(\frac{1}{2}\right)$. Notice that the equilibrium payoff for L-type disputant is different from his expected payoff under war, for he enjoys the peaceful payoff $(1/2)$ when facing the same L-type with probability $1-q$.

3.2.2 A Model with Mediator Selection

To improve on a given conflict situation, both parties may initiate a mediation. A mediator has no direct interest in a specific allocation of the pie and thus wants to maximize the overall utilitarian welfare. However, the mediator may not be impartial: she is interested in conferring advantages to one disputant. Let $\lambda \in [0,1]$ measure the degree by which the mediator is biased towards disputant 1. For example, $\lambda = \frac{1}{2}$ indicates an unbiased mediator. As λ increases, a mediator's bias toward disputant 1 gets more extreme. We thus define the payoff of a mediator whose bias is λ as

$$w_\lambda = \lambda U_1 + (1-\lambda)U_2$$

where U_1 and U_2 are payoffs for disputant 1 and 2. That is, a mediator is identified by her bias toward disputant 1, and the unit interval $[0,1]$ from which λ takes its value is thus the set of (potential) mediators.

Our model of mediator selection consists of the two stages: the *selection* stage and the *mediation* stage.

Selection Stage Nature chooses randomly a potential mediator λ from $[0,1]$. This potential mediator makes a mediation offer to both disputants. Given the offer, both disputants (after learning their types) simultaneously makes a decision of whether to accept the mediation offer or not. Formally, an acceptance

strategy of disputant i of type $\tau_i \in \{H, L\}$ is $v^i : \{H, L\} \to \{0, 1\}$, where 1 denotes "accept" and 0 denotes "reject". If both disputants agree to accept the offer ($v^1_{\tau_1} = v^2_{\tau_2} = 1$ for some $(\tau_1, \tau_2) \in \{H, L\}^2$, then the selection stage is said to be successful, and the mediation stage begins. Otherwise, both disputants play the War-and-Peace game.

In either case, both disputants as well as the mediator observe the choice made by each disputant when the selection stage is over. As each disputant chooses a strategy *after* learning his type, his private information may be revealed to both his opponent and the mediator. Let q_i denote the posterior belief about disputant $i = 1, 2$ being H-type, and it becomes a common knowledge.

Mediation Stage Once the mediator with bias λ is accepted, each disputant privately sends a report $m \in \{H, L\}$ to the mediator. Given the report m, the mediator makes a recommendation (or a mechanism) in order to maximize her expected payoff. We assume that the mediator commits herself to the mechanism. By applying the revelation principle, we consider only the direct mechanisms. The mechanism thus consists of a type-dependent recommendation to each disputant about which action to take, either "Stay" or "Attack", and about how to split the pie conditional on the event that peace is achieved. Without loss of generality, a direct mechanism proposed by the mediator with her bias λ and her beliefs about disputants q_1 and q_2, can be summarized as the tuple $M \equiv M(\lambda, q_1, q_2) = (p_\tau, b_\tau)_{\tau \in \{H,L\}^2}$ where p_τ is the probability of peace and b_τ is the share of the pie allocated to disputant 1 when the reported type profile is $\tau \in \{H, L\}^2$[5]. Given the mechanism $M(\lambda, q_1, q_2)$ proposed by

[5]This formulation of a direct mechanism utilizes the following facts: (i) the sum of the shares is one under peace, and (ii) a war breaks out unilaterally (i.e. if one disputant is

the mediator, each disputant makes a decision whether to accept the proposed mechanism or not. We assume that if either disputant rejects the mechanism, war breaks out surely. Moreover, again by invoking the revelation principle, we consider only the direct mechanisms in which each disputant reports his type truthfully. To be specific, the incentive compatibility constraints for disputant 1 of each type are stated as follows:

$$
\begin{aligned}
(IC_{1H}) \quad & q_2 \left[p_{HH} b_{HH} + (1 - p_{HH}) \left(\frac{\theta}{2} \right) \right] + (1 - q_2) \left[p_{HL} b_{HL} + (1 - p_{HL}) p\theta \right] \\
& \geq q_2 \left[p_{LH} b_{LH} + (1 - p_{LH}) \left(\frac{\theta}{2} \right) \right] + (1 - q_2) \left[p_{LL} b_{LL} + (1 - p_{LL}) p\theta \right] \\
(IC_{1L}) \quad & q_2 \left[p_{LH} b_{LH} + (1 - p_{LH})(1 - p)\theta \right] + (1 - q_2) \left[p_{LL} b_{LL} + (1 - p_{LL}) \left(\frac{\theta}{2} \right) \right] \\
& \geq q_2 \left[p_{HH} b_{HH} + (1 - p_{HH})(1 - p)\theta \right] + (1 - q_2) \left[p_{HL} b_{HL} + (1 - p_{HL}) \left(\frac{\theta}{2} \right) \right]
\end{aligned}
$$

The left-hand side of (IC_{1H}) is the interim payoff of disputant 1 of H-type when he truthfully reports his type. When facing H-type with probability q_2, peace is achieved with probability p_{HH} and the share b_{HH} is allocated. With the remaining probability $1 - p_{HH}$, the mediation fails and war thus breaks out. However, H-type does not win the war and obtain $\frac{\theta}{2}$. Facing L-type with probability $1 - q_2$, peace is achieved with probability p_{HL} and the share b_{HL} is allocated. When the mediation fails, war breaks out. Then, H-type wins and thus obtains $\frac{\theta}{2} + (1 - \theta)$. The right-hand side is the interim payoff when H-type lies by reporting his type as L-type. In addition, the incentive compatibility constraint for disputant 1 of L-type can also be interpreted in a similar way.

recommended to attack, then war breaks out regardless of the other disputant's action). For details, see Appendix.

The participation constraints for disputant 1 of high type and low type are:

$$(PC_{1H}) \quad q_2 \left[p_{HH} b_{HH} + (1 - p_{HH}) \left(\frac{\theta}{2} \right) \right] + (1 - q_2) \left[p_{HL} b_{HL} + (1 - p_{HL}) p\theta \right]$$

$$\geq q_2 \left(\frac{\theta}{2} \right) + (1 - q_2) p\theta$$

$$(PC_{1L}) \quad q_2 \left[p_{LH} b_{LH} + (1 - p_{LH})(1 - p)\theta \right] + (1 - q_2) \left[p_{LL} b_{LL} + (1 - p_{LL}) \left(\frac{\theta}{2} \right) \right]$$

$$\geq q_2 (1 - p)\theta + (1 - q_2) \left(\frac{\theta}{2} \right)$$

Since the rejection of the proposed mechanism leads to war, the right-hand sides are the interim payoffs under war. Similar to the case of disputant 1, the incentive compatibility constraints and the participation constraints for disputant 2 of each type can be formulated as follows:

$$(IC_{2H}) \quad q_1 \left[p_{HH}(1 - b_{HH}) + (1 - p_{HH}) \left(\tfrac{\theta}{2} \right) \right] + (1 - q_1) \left[p_{LH}(1 - b_{LH}) + (1 - p_{LH}) p\theta \right]$$

$$\geq q_1 \left[p_{HL}(1 - b_{HL}) + (1 - p_{HL}) \left(\tfrac{\theta}{2} \right) \right] + (1 - q_1) \left[p_{LL}(1 - b_{LL}) + (1 - p_{LL}) p\theta \right]$$

$$(IC_{2L}) \quad q_1 \left[p_{HL}(1 - b_{HL}) + (1 - p_{HL})(1 - p)\theta \right] + (1 - q_1) \left[p_{LL}(1 - b_{LL}) + (1 - p_{LL}) \left(\tfrac{\theta}{2} \right) \right]$$

$$\geq q_1 \left[p_{HL}(1 - b_{HL}) + (1 - p_{HL})(1 - p)\theta \right] + (1 - q_1) \left[p_{LH}(1 - b_{LH}) + (1 - p_{LH}) \left(\tfrac{\theta}{2} \right) \right]$$

$$(PC_{2H}) \quad q_1 \left[p_{HH}(1 - b_{HH}) + (1 - p_{HH}) \left(\tfrac{\theta}{2} \right) \right] + (1 - q_1) \left[p_{LH}(1 - b_{LH}) + (1 - p_{LH}) p\theta \right]$$

$$\geq q_1 \left(\tfrac{\theta}{2} \right) + (1 - q_1) p\theta$$

$$(PC_{2L}) \quad q_1 \left[p_{HL}(1 - b_{HL}) + (1 - p_{HL})(1 - p)\theta \right] + (1 - q_1) \left[p_{LL}(1 - b_{LL}) + (1 - p_{LL}) \left(\tfrac{\theta}{2} \right) \right]$$

$$\geq q_1 (1 - p)\theta + (1 - q_1) \left(\tfrac{\theta}{2} \right)$$

As the mediator's expected payoff is the average of the payoffs of disputants weighted by the bias λ, it can be formulated as follows:

$$\begin{aligned} W_\lambda\left(p_\tau, b_\tau)_{\tau\in\{H,L\}^2}\right) &= E_\tau[\lambda(p_\tau b_\tau^1 + (1-p_\tau)d_\tau^1) + (1-\lambda)(p_\tau b_\tau^2 + (1-p_\tau)d_\tau^2)] \\ &= (2\lambda-1)E_\tau[p_\tau(b_\tau^1 - d_\tau^1)] + (1-\lambda)(1-\theta)E_\tau p_\tau + \text{ constant terms} \\ &\quad \text{or, equivalently,} \\ &= (1-2\lambda)E_\tau[p_\tau(b_\tau^2 - d_\tau^2)] + \lambda(1-\theta)E_\tau p_\tau + \text{ constant terms,} \end{aligned}$$

where b_τ^i and d_τ^i are the shares of disputant i under peace and under war, respectively. Specifically, $b_\tau^1 = 1 - b_\tau^2 = b_\tau$, $d_{HH}^1 = d_{LL}^1 = \theta/2$, $d_{HL}^1 = p\theta$, $d_{LH}^1 = (1-p)\theta$ such that $d_\tau^1 + d_\tau^2 = \theta$.

Remark 4. *The expected payoff of the mediator W_λ consists of two distinct components: The first component weighted by $(2\lambda-1)$ is the ex-ante expected gain of disputant 1 relative to his payoff under war, and the second component weighted by $(1-\lambda)(1-\theta)$ is the ex-ante expected probability of peace. Suppose that the mediator is extremely biased toward disputant 1, i.e. $\lambda = 1$. Then, the mediator coincides completely with the ex-ante gain of disputant 1. If the chosen mediator is impartial, i.e. $\lambda = \frac{1}{2}$, then the mediator only cares about the expected peace probability without any concern for a specific allocation of the pie under peace. Lastly, suppose that the mediator is completely on the side of disputant 2, i.e. $\lambda = 0$. Then, the resulting expression can be shown to be the ex-ante gain of disputant 2. We relegate this to the Appendix.*

Explicitly, we may express the mediator's payoff from the viewpoint of disputant 1 as follows:

$$W_\lambda\left(p_\tau, b_\tau\right)_{\tau\in\{H,L\}^2}\Big) = (2\lambda-1)\left[q_1q_2p_{HH}\left(b_{HH}-\frac{\theta}{2}\right)+q_1(1-q_2)p_{HL}\left(b_{HL}-p\theta\right)\right]$$

$$+(2\lambda-1)\left[q_2(1-q_1)p_{LH}\left\{b_{LH}-(1-p)\theta\right\}+(1-q_1)(1-q_2)p_{LL}\left(b_{LL}-\frac{\theta}{2}\right)\right]$$

$$+(1-\lambda)(1-\theta)\left[q_1q_2p_{HH}+q_1(1-q_2)p_{HL}+(1-q_1)q_2p_{LH}+(1-q_1)(1-q_2)p_{LL}\right]$$

+constant terms,

Hence, the optimal mediation programme **(P)** for the mediator is to determine p_τ and b_τ for each $\tau \in \{H,L\}^2$ to maximize the expected payoff of the mediator[6]:

$$\text{(P)} \quad \max_{(p_\tau,b_\tau)_{\tau\in\{H,L\}^2}} W_\lambda\left(p_\tau, b_\tau\right)_{\tau\in\{H,L\}^2}\Big)$$

subject to the interim incentive compatibility constraints and the interim participation constraints for both disputants: (IC_{1H}), (IC_{1L}), (PC_{1H}), (PC_{1L}), (IC_{2H}), (IC_{2L}), (PC_{2H}), and (PC_{2L}).

Equilibrium Definition To define an equilibrium, we first define the payoff of each disputant. For a strategy profile $(v_H^1, v_L^1, v_H^2, v_L^2)$, the posterior beliefs q_1 and q_2 are calculated via Bayes' rule whenever it is applicable. Define $\gamma_i = \frac{q_i}{1-q_i}$ for $i = 1, 2$. For any type profile $\tau = (\tau_1, \tau_2) \in \{H, L\}^2$ such that a mediator is rejected, i.e. $v_{\tau_1}^1 \neq v_{\tau_2}^2$, each disputant's payoff is his equilibrium payoff in War-and-Peace game. Specifically, if $\gamma_1 \leq \delta$ or $\gamma_2 \leq \delta$, the payoff for disputant i of L-type is $q_j(1-p)\theta + (1-q_j)(1/2)$, and for H-type it is $q_j\left(\frac{\theta}{2}\right) + (1-q_j)p\theta$

[6]For the derivation of the expected payoff of the mediator, see Appendix.

for i's opponent j, $j \neq i$. Otherwise, if $\gamma_1 > \delta$ and $\gamma_2 > \delta$, disputant i's payoff is $1/2$ for both types.

Let $\mathcal{E}(q_1, q_2)$ be the set of equilibria in War-and-Peace game with the beliefs q_1 and q_2. Whenever a mediator is accepted, i.e. for some type profile $\tau = (\tau_1, \tau_2) \in \{H, L\}^2$ such that $v^1_{\tau_1} = v^2_{\tau_2} = 1$, the payoffs are determined by the optimal mediation mechanism $M(\lambda, q_1, q_2)$.

The solution concept we use is Perfect Bayesian equilibrium. An equilibrium is a strategy profile $(v^1_H, v^1_L, v^2_H, v^2_L)$ satisfying the following:

- For his opponent's nomination strategy (v^j_H, v^j_L), a mechanism $M(\lambda, q_1, q_2)$, and an equilibrium of War-and-Peace game $\mathcal{E}(q_1, q_2)$, disputant i of type $\tau \in \{H, L\}$ maximizes his expected payoff.

- For the mediator's bias λ and the posterior beliefs q_1 and q_2, the optimal mechanism $M(\lambda, q_1, q_2)$ solves **(P)**.

- Given a strategy profile $(v^1_H, v^1_L, v^2_H, v^2_L)$, the posterior belief q_1 and q_2 are determined by Bayes' rule whenever it is applicable.

The whole structure of the model, somewhat complicated though, is depicted in Figure 3.2

3.3 Optimal Mechanism

In analyzing the model, we shall focus on pure strategy equilibria. That is, we shall analyze the following cases: (i) $v^i_H = v^i_L$, which we refer to as *pooling strategy*, and (ii) $v^i_H \neq v^i_L$, so-called *separating strategy*, for $i = 1, 2$. Therefore, we shall describe the features of an optimal mechanism in each case. Proofs are

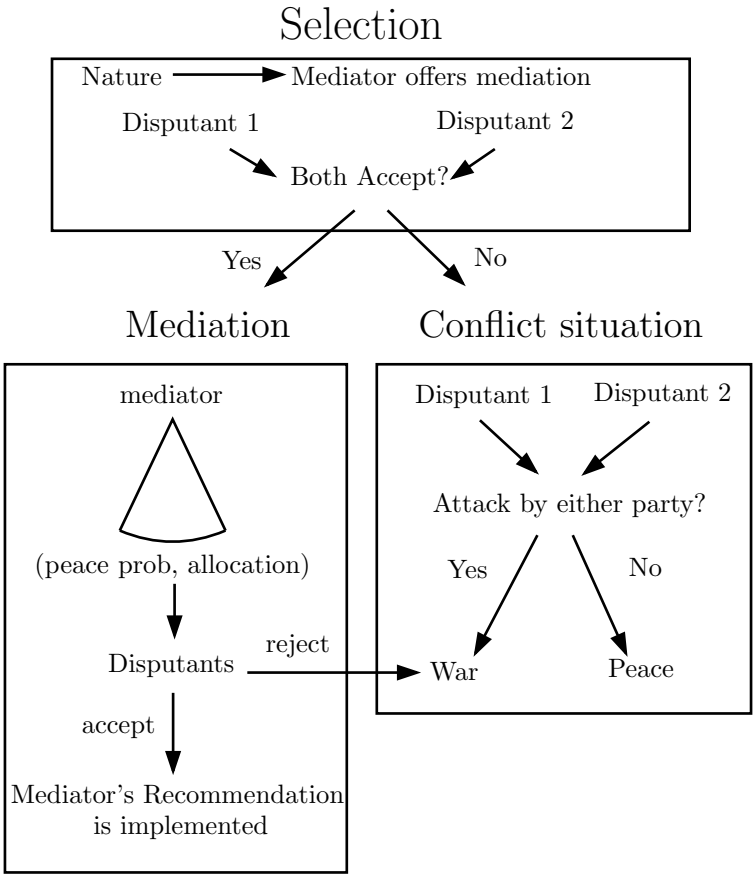


Figure 3.2: War-and-Peace Game with Mediator Selection

relegated to the Appendix. For more parsimonious analysis, we shall work with the following notation: Let $\gamma_i = \frac{q_i}{1-q_i}$ denote the likelihood of disputant i being H-type. We thus denote the mechanism as $M(\lambda, \gamma_1, \gamma_2)$.

First of all, we consider an optimal mechanism under pooling strategy. If the nomination stage is successful, no information is revealed in the nomination stage. The posterior belief shared by the mediator and both disputants is thus identical to the initial belief: $\gamma_1 = \gamma_2 = \gamma < 1$.

In presenting the optimal mechanism, we shall report only the peace probabilities and the rent obtained by each disputant[7]. Specifically, let $du_{i\tau_i}$ denote the (informational) rent of disputant $i = 1, 2$ of type $\tau_i \in \{H, L\}$ after normalizing by $(1-\theta)(1-q)$. That is, the disputant i's interim payoff is $du_{i\tau_i}(1-\theta)(1-q)$ + disputant i's war payoff. Hence, we shall present p_τ and $du_{i\tau_i}$ for $\tau = (\tau_1, \tau_2) \in \{H, L\}^2$.

When the mediator is unbiased ($\lambda = 1/2$), HMS solve for an optimal mechanism by naturally assuming values of some choice variables by imposing symmetry: $p_{HL} = p_{LH}$, $b_{HL} = b_{LH}$, and $b_{HH} = b_{LL} = 1/2$. This leads one to consider constraints without identifying individual disputants. The details about the solution and the approach for the optimal mechanism for a unbiased mediator can be found in HMS(2015).

Lemma 12 (Optimal Mechanism under Pooling Strategy with Unbiased Mediator, HMS(2015))**.** *Suppose that the mediator is unbiased* $\left(\lambda = \frac{1}{2}\right)$*. Then, the optimal mechanism* $M(1/2, \gamma, \gamma)$ *satisfies the following:*

- *The incentive compatibility constraints of L-type and the participation constraints of H-type bind, and the others do not.*
- *For* $\gamma \leq \frac{\delta}{2}$*, L-type dyads* (L, L) *do not fight* $(p_{LL} = 1)$*, asymmetric dyads* (H, L) *and* (L, H) *enjoy peace with probability* $p_{HL} = p_{LH} = \frac{1}{1+\delta-2\gamma} \in (0, 1)$*, H-type dyads* (H, H) *always fight* $(p_{HH} = 0)$*. The resulting interim payoffs that disputants of each type obtain in addition to their payoff under war, are* 0 *for H-type and* $\left(\frac{1}{1+\delta-\gamma}\right)\left(\frac{1+\delta}{2}\right)$ *for L-type.*

[7] We do not report the optimal split of the pie for each type profile under the mechanism. For our analysis, only the rent obtained by each disputant under mechanism is relevant. Moreover, the optimal split of the pie is indeterminate as there are multiple optima.

- *For $\gamma > \frac{\delta}{2}$, L-type dyads and asymmetric dyads do not fight ($p_{LL} = p_{HL} = p_{LH} = 1$) and H-type dyads fight with probability $p_{HH} = \frac{2\gamma-\delta}{\gamma(1+\delta-\gamma)} \in (0,1)$. The resulting rents are 0 for H-type and $\left(\frac{\gamma+1}{1+\delta-\gamma}\right)\left(\frac{1+\delta}{2}\right)$ for L-type.*

When the mediator is biased ($\lambda \neq 1/2$), it is actually not easy to figure out an optimal mechanism. Unlike the case of an unbiased mediator, the optimal mechanism proposed by a biased one treats disputants differently depending on the direction of a mediator's bias. Moreover, when it comes to a disputant favored by the mediator, it is not easy to see whether a participation constraint or an incentive compatibility constraint does or does not bind at the optimum. We relegate the detailed description of how we resolve these difficulties and the relevant proofs to the Appendix. In presenting the optimal mechanism for a biased mediator, we shall only report the case where the mediator is biased in favor of disputant 1 ($\lambda > 1/2$). This is without loss of generality because one can easily obtain the optimal mechanism by exchanging the roles of disputants.

Lemma 13 (Optimal Mechanism under Pooling Strategy with biased Mediator)**.** *Suppose that the mediator is biased in favor of disputant 1 $\left(\lambda > \frac{1}{2}\right)$. Let $\lambda > \hat{\lambda} \equiv \frac{1+\delta}{2(1+\delta-\gamma)}$. Then, the optimal mechanism $M(\lambda, \gamma, \gamma)$ determines the peace probabilities and each disputant's rent (p_τ and $du_{i\tau_i}$ for $\tau = (\tau_1, \tau_2) \in \{H, L\}^2$) as follows:*

(1) When the mediator is extremely biased toward disputant 1 ($\lambda > \hat{\lambda}$), the incentive compatibility constraints of both disputants of L-type, disputant 1 of H-type, and the participation constraints of disputant 2 of H-type bind, i.e. (IC_{1L}), (IC_{2L}), (IC_{1H}), and (PC_{2H}) bind. Peace is attained when the opponent, disputant 2, is of L-type: $p_{LL} = p_{HL} = 1$. In the

remaining cases, war breaks out: $p_{HH} = p_{LH} = 0$. *Disputant* 1*'s rent is* $du_{1H} = 1 - \frac{1+\delta}{2(1+\gamma)}$ *for* H*-type, and* $du_{1L} = 1 - \frac{\gamma(1+\delta)}{2(1+\gamma)}$ *for* L*-type. The opponent, disputant* 2*, obtains his expected payoff under war regardless of its type, i.e.* $du_{2H} = du_{2L} = 0$.

(2) When the mediator is moderately biased $\left(\lambda \in \left(\frac{1}{2}, \hat{\lambda}\right]\right)$*, the participation constraints of* H*-type and the incentive compatibility constraints of* L*-type bind and the others do not. That is,* (IC_{1L})*,* (IC_{2L})*,* (PC_{1H})*, and* (PC_{2H}) *bind.*

(a) For $\gamma > \delta/2$*, every dyad except for the* H*-type dyad* (H, H) *does not fight* $(p_{LL} = p_{HL} = p_{LH} = 1)$*. The* H*-type dyad fight with probability* $p_{HH} = \frac{2\gamma-\delta}{\gamma(1+\delta-\gamma)} \in (0, 1)$*. For both disputants,* H*-types obtain its expected payoff under war:* $du_{1H} = du_{2H} = 0$*.* L*-types enjoy the rent* $du_{1L} = du_{2L} = \left(\frac{\gamma+1}{1+\delta-\gamma}\right)\left(\frac{1+\delta}{2}\right)$.

(b) For $\gamma \leq \delta/2$*, dyads when disputant* 2 *is* L*-type do not fight* $(p_{LL} = p_{HL} = 1)$*. For the* H*-type dyad, war always breaks out,* $p_{HH} = 0$*. The remaining dyad* (L, H) *enjoys peace with probability* $p_{LH} = \frac{1-\delta+2\gamma}{1+\delta-2\gamma} \in (0, 1)$*. Disputant* 1 *of* L*-type, toward which the mediator is biased, enjoys the informational rent of* $du_{1L} = \frac{1+\delta}{2}$*. Disputant* 2 *of* L*-type also enjoys the less amount of informational rent* $du_{2L} = \left(\frac{1-\delta+2\gamma}{1+\delta-2\gamma}\right)\left(\frac{1+\delta}{2}\right) < du_{1L}$*. For both disputants,* H*-types obtain their expected payoffs under war, i.e.* $du_{1H} = du_{2H} = 0$.

Notice that with a biased mediator, the peace probabilities for the asymmetric dyads and the rents for L-type disputants are disproportionately allocated

across disputants. To be specific, the biased mediators allocates more resources to the L-type disputant they favor, while allowing the opponent of L-type enjoy peace with certainty. This does not seem intuitive at first sight: Why does the mediator give a favor to the disputant in her opposition in terms of peace probability rather than to the disputant she favors? To understand this, suppose that the mediator is biased in favor of disputant 1. In order to provide the truth-telling incentive for disputant 2 of L-type, the mediator must give away some informational rent. This can be done by raising peace probability p_{HL} to disputant 2 of L-type or, alternatively, by raising the share $1 - b_{HL}$ allocated to it under peace. They affect the payoff of disputant 1 differently: The former also benefits disputant 1 of H-type, but the latter does only harm to disputant 1 of either type. Consequently, the mediator chooses the cheaper way, thus guaranteeing peace to disputant 2 of L-type.

Corollary 3. *Suppose that a mediator is biased in favor of one disputant, say disputant* 1*. An optimal mechanism exhibits the following features:*

- *Disputant* 2 *of L-type enjoys no less chance of peace than disputant* 1 *of the same type, i.e.* $p_{HL} = p_{LL} = 1 \geq p_{LH}$*.*
- *Disputant* 1 *of L-type obtains the information rent no less than disputant* 2 *of the same type:* $du_{1L} \geq du_{2L}$*.*
- *A strict inequality holds for both in the above, if the mediator is either moderately biased with* $\gamma \leq \delta/2$*, or extremely biased.*

That is, the optimal mechanism allocates more shares of the pie to the favored party while guaranteeing more peace probabilities to the disfavored party.

Now, turning to the expected peace probability achieved under an optimal mechanism, we show that a biased mediator, even with the bias, does not necessarily perform worse than a unbiased one if her bias is not severe.

Theorem 6. *In an optimal mechanism under pooling strategy, a biased mediator achieves the same expected peace probability as an unbiased one if and only if the mediator has a moderate level of bias,* $\lambda \in \left(1-\hat{\lambda}, \frac{1}{2}\right) \cup \left(\frac{1}{2}, \hat{\lambda}\right)$*, where* $\hat{\lambda} = \frac{1+\delta}{2(1+\delta-\gamma)}$.

Proof. It is easy to see that an extremely biased mediator achieves a lower expected peace probability than an unbiased one. For the other direction, if $\gamma > \delta/2$, then it is obvious because the peace probabilities are identical for both a moderately biased mediator and a unbiased one. For $\gamma \leq \delta/2$, the peace probabilities under a moderately biased mediator differ from those under a unbiased one only for asymmetric dyads. However, the expected peace probability assigned for asymmetric dyads are the same: $q(1-q)\left(1+\frac{1-\delta+\gamma}{1+\delta-2\gamma}\right) = 2q(1-q)\left(\frac{1}{1+\delta-\gamma}\right)$. □

This is striking because the biased mediator does not care about achieving peace per se. The crucial factor that lies behind this result is the fact that war is costly: If war breaks out, some proportion $(1-\theta)$ of the pie is lost. By achieving peace, a mediator may offer more shares to her ally by using this resource that would have been lost under war. However, the mediator's incentive to attain peace is limited by the informational rent that must be given to the opponent of L-type for providing him with the truth-telling incentive. As long as she is not extremely biased, the mediator does not find this informational rent costly. Therefore, the moderately biased mediator happens to maximize the

peace probability as the unbiased mediator does. If the mediator is extremely biased, however, the mediator would find the informational rent given to the opponent too costly. Accordingly, she chooses to give no informational rent to her opponent by reducing the peace probability.

Now, we consider an optimal mechanism when disputants fully reveal their private information during the nomination stage. That is, $v_H^1 \neq v_L^1$ and $v_H^2 \neq v_L^2$. As both disputants' types become common knowledge, the optimal mechanism and the resulting payoff of each disputant are simply computed as follows:

Lemma 14 (Optimal Mechanism under Separating Strategy)**.** *The optimal mechanism under separating strategy achieves peace with certainty. In addition, a biased mediator allocates the share to the opponent (the disputant against whom she is against) only to make the opponent accept the mechanism. Specifically, the following hold:*

(a) If the mediator is unbiased ($\lambda = 1/2$)*, the disputant* 1*'s share under peace is* $b_{HH} = b_{LL} = 1/2$*,* $b_{HL} = p\theta + \frac{1-\theta}{2}$*, and* $b_{LH} = (1-p)\theta + \frac{1-\theta}{2}$*.*

(b) If the mediator is biased toward disputant 1 $\left(\lambda > \frac{1}{2}\right)$*, then* $b_{HH} = b_{LL} = 1 - \frac{\theta}{2}$*,* $b_{HL} = 1 - (1-p)\theta$*,* $b_{LH} = 1 - p\theta$*.*

(ii) If the mediator is biased against disputant 1 $\left(\lambda < \frac{1}{2}\right)$*, then the disputant* 1*'s share under peace is as follows:* $b_{HH} = b_{LL} = \frac{\theta}{2}$*,* $b_{HL} = p\theta$*, and* $b_{LH} = (1-p)\theta$*, and* $b_{LL} = \frac{\theta}{2}$*.*

3.4 Equilibria: Incidence of Biased Mediators

As we noted in the previous section, we focus on the pure strategy equilibria. For the pooling-strategy equilibria, we have the following result:

Theorem 7. *Suppose that disputants employ pooling strategies. Let* $\hat{\lambda} = \frac{1+\delta}{2(1+\delta-\gamma)}$. *Then, the following:*

- *Suppose that disputants are more likely to face H-type* $\left(\gamma \geq \frac{\delta(1+\delta)}{2(2+\delta)}\right)$.
  - *(a) For* $\lambda \in (1-\hat{\lambda}, \hat{\lambda})$, *a strategy profile* $(v_H^1, v_L^1, v_H^2, v_L^2)$ *satisfying* $v_H^1 = v_L^1 = v_H^2 = v_L^2 = 1$ *is a pooling-strategy equilibrium.*
  - *(b) All pooling-strategy equilibria within* $\lambda \in (1-\hat{\lambda}, 1/2)$ *are all outcome-equivalent. Moreover, so are all pooling-strategy equilibria within* $\lambda \in (1/2, \hat{\lambda})$.
- *Suppose that disputants are less likely to face H-type* $\left(\gamma < \frac{\delta(1+\delta)}{2(2+\delta)}\right)$. *Then, there is a unique pooling-strategy equilibrium in which only the unbiased mediator* $(\lambda = 1/2)$ *is accepted.*

The above theorem tells that a disputant would agree to accept the mediator biased in favor of his opponent if he is likely to face the H-type opponent. Otherwise, the disputant would not accept a mediator unless the mediator is impartial. If a disputant anticipates that there are more to lose than to gain in the conflict situation, he would rather to be engaged in a mediation, even when he expects the mediator to be biased against him. This simple intuition, however, does not provide an explanation why the disputant would gain under mediation, despite the fact that the mediator stands on the opposite side of him. Figure 3.3 illustrates this idea.

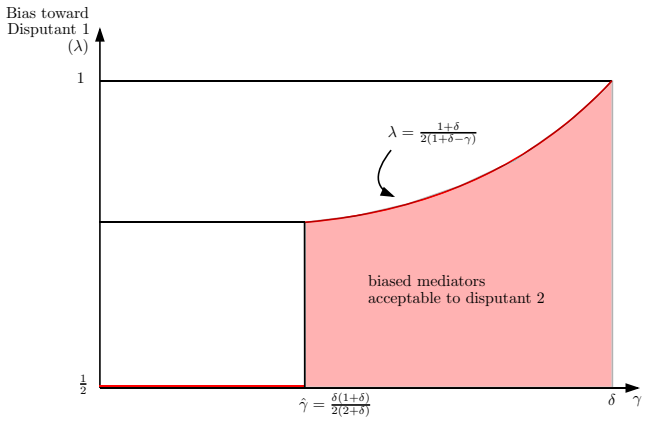


Figure 3.3: Incidence of a Biased Mediator: $\lambda \geq 1/2$

It is the informational rent that provides an answer. For illustration, suppose that the mediator is biased in favor of disputant 1. Recall that under mediation, disputant 2 of H-type's participation constraint holds with equality. In other words, he is indifferent between engaging in mediation and continuing the conflict. For the disputant 2 of L-type, he has an incentive for pretending to be H-type. To prevent this, the mediator needs to give the informational rent, even though the mediator wishes to allocate more shares to her ally. Therefore, even when the mediator is biased against him, disputant 2 would gain under mediation.

Focusing on separating-strategy equilibria ($v_H^1 \neq v_L^1$ and $v_H^2 \neq v_L^2$). We obtain the following result:

Theorem 8. *Suppose that disputants employ separating strategies. There exists a unique equilibrium in which L-type dyad (L, L) accepts only an unbiased mediator ($v_L^1 = v_L^2 = 1$) and H-type dyad (H, H) rejects on such a mediator ($v_H^1 = v_H^2 = 0$).*

The intuition behind this result is rather simple. As there is always an incentive for L-type to mimic H-type, L-type would deviate to accept, whenever there is a mediation for an asymmetric dyad, say (H, L) or (L, H). Similarly, there exists no equilibrium where the H-type dyad successfully agree on a biased mediator. Moreover, it is not an equilibrium that the H-type dyad successfully agree on a unbiased mediator. By deviating to match with L-type under no mediation, H-type obtains $p\theta$ which is larger than his payoff of $1/2$ under mediation. When L-type dyad (L, L) successfully selects a biased mediator, H-type would deviate for his payoff of $1/2$ is smaller than the deviation payoff of $1-\theta/2$. When L-type dyad (L, L) agrees on the unbiased mediator, the resulting split would be $(1/2, 1/2)$. H-type has no incentive to deviate, for this is equal to H-type's payoff under no mediation. Moreover, L-type also would not deviate to match with H-type under no mediation, for the deviation payoff of $(1-p)\theta$ is smaller than the equilibrium payoff of $1/2$.

Notice that at the separating-strategy equilibrium, the ex-ante peace probability is the same as the one in the War-and-Peace game. In addition, it is less than the one at any pooling-strategy equilibria. That is, the separating-strategy equilibrium is Pareto-dominated by any pooling-strategy equilibria. Therefore, we restrict our attention to the pooling-strategy equilibria.

3.5 Effectiveness of Biased Mediators

From the results in the previous section, we are now able to discuss the issue regarding the effectiveness of biased mediators. As it is demonstrated in Theorem 6, in general, a biased mediator performs worse than a unbiased one, for

those mediator with the extreme level of bias achieves the lower expected peace probability. This stands by a general concern about the mediator's bias.

Nevertheless, Theorem 7 reveals that one may disregard such a concern. Although extremely biased mediators are not equally effective as the unbiased one, these mediators would never be chosen in equilibrium. In other words, if one observes that a biased third party acts as a mediator, such a mediator would have a moderate level of bias, and more importantly, she would be equally effective as a unbiased mediator.

Corollary 4. *In any equilibria where a biased mediator is accepted, such a mediator is equally effective as an unbiased one in resolving a conflict.*

By looking at the above corollary, one may wonder about the connection between the effectiveness of a mediator and the demand of disputants for the mediator. As we argue earlier in the introduction, for the disputants, peace is not the end itself. Each disputant cares only about how much he would gain from the likely outcome of mediation, while comparing with the outcome from an ongoing conflict. However, our result shows that only the effective mediators are selected by disputants. Does this mean that disputants demand a mediator who is likely to be effective? We argue that our result does not allow such an interpretation. Extremely biased mediators, although they achieve a higher expected peace probability than the disputants may achieve in a conflict situation (War-and-Peace game), fail to be selected in equilibrium due to the objection by one party. Then, why do we see a connection between the effectiveness and the mediation incidence?

The connection lies in how the mediator's bias affects the allocation of the

informational rent and the peace probability between the two parties. According to our results, we may classify all mediators according to the mediation outcome they propose into the following three categories: unbiased, moderately biased, and extremely biased[8]. The unbiased mediator allocates the peace probability and the informational rent equally across disputants. On the other hand, the biased mediators tend to allocate more informational rent to her ally, while allocating more peace probabilities to the opposite party.

What distinguishes the extremely biased ones from the moderately biased ones is that for the former, no informational rent is allocated to the party whom the mediator is biased against. As we discussed earlier in the section on the optimal mechanism, this is due to the trade-off faced by the biased mediator. To reiterate, maximizing the expected peace probability would be beneficial to the mediator by allowing her to allocate more resources to the disputant whom she is biased in favor of without wasting them in war. However, achieving peace incurs a cost to the mediator. She needs to provide the informational rent, especially even to the disputant she is not favor of. If her bias is not extreme, this cost does not constrain the mediator's willingness to promote peace. Otherwise, the mediator would give no rent to the opposite party by compromising her benefit from achieving peace.

A disputant, when making his decision in the nomination stage, would consider this likely outcome of mediation under a mediator who is extremely biased in favor of his opponent. Especially, when he is L-type, he would see that the likely outcome, even under mediation, is indifferent to the outcome under war.

[8]Although a biased mediator in favor of disputant 1 makes a different recommendation from one in favor of disputant 2, we include them into one category, for the recommendations they propose exhibits symmetry.

As he may enjoy a peaceful outcome with some positive probability in a given conflict situation, he would rather to take a bet by refusing any mediation by an extremely biased third party.

3.6 Conclusion

In this paper, we study how mediator bias affects the initiation as well as the effectiveness of mediation. Specifically, we investigate two different but closely related issues: (i) why, and under which circumstances a disputant is willing to accept a mediator who is biased in favor of his opponent, and (ii) whether such a biased mediator, if accepted, is equally effective as an unbiased one in promoting peace.

To this end, we construct a simple model in which disputants make a joint decision of whether to accept a third-party who is potentially biased in favor of one party as their mediator. Our model adds a novel feature of mediator bias to the previous mechanism design approach to mediation (Fey and Ramsay, 2010; Hörner et al., 2015). This leads to the optimal mechanism that is qualitatively different from the one assuming the mediator's impartiality. Specifically, the biased mediator allocates the peace probabilities and the share of the pie under peace differently across disputants: the favored party enjoys more interim payoffs, while the disfavored party, in return, is guaranteed more peace probabilities.

From this investigation of an optimal mechanism under a biased mediator, we show that if a conflict is less likely to end up with a peaceful outcome, disputants would accept a biased mediator. This result relies on the presence of

private information. To elicit a disputant's private information during the mediation process, a mediator has to provide an informational rent to the disputant even when she is biased against him. Hence, the disputant would be willing to accept mediation if the alternative (conflict) is likely to ends up with war. This is consistent with an empirical finding by Melin et al. (2013).

A novel result we find out, in relation to the effectiveness of a biased mediator, is that if accepted, a biased mediator achieves peace equally as an unbiased one. This implies that accounting for an endogenous selection, a mediator's bias is independent of her effectiveness in resolving a conflict. This relies on the fact that when the mediator is moderately biased, promoting peace as much as possible is beneficial to her. To be specific, as war is socially wasteful, promoting peace would allow a mediator to serve her ally's interest. This, however, increases the need for the mediator to give more informational rent to the disfavored party. As long as the level of bias is moderate, providing an informational rent does not compromise the benefit of promoting peace to the mediator.

Appendix A

Appendix for Chapter 3

A.1 War-and-Peace Game with Arbitrary Beliefs

In this section, we analyze the game described as the environment by allowing the probability of disputant i being H-type to differ. That is, disputant 1 is of H-type with probability q_1, and disputant 2 is of H-type with probability q_2. Let α_τ^i denote the probability that disputant i of type $\tau \in \{H, L\}$ chooses to stay. Then, the expected payoff of each disputant for a given mixed strategy profile $(\alpha_H^1, \alpha_L^1, \alpha_H^2, \alpha_L^2)$ can be computed as follows:

$$\begin{aligned} u_{1H}(\alpha_H^1, \alpha_H^2, \alpha_L^2) &= q_2 \left(\frac{\alpha_H^1 \alpha_H^2}{2} + \frac{(1 - \alpha_H^1 \alpha_H^2)\theta}{2} \right) + (1 - q_2) \left[\frac{\alpha_H^1 \alpha_L^2}{2} + (1 - \alpha_H^1 \alpha_L^2) p\theta \right] \\ u_{1L}(\alpha_L^1, \alpha_H^2, \alpha_L^2) &= q_2 \left(\frac{\alpha_L^1 \alpha_H^2}{2} + (1 - \alpha_L^1 \alpha_H^2)(1 - p)\theta \right) + (1 - q_2) \left[\frac{\alpha_L^1 \alpha_L^2}{2} + \frac{(1 - \alpha_L^1 \alpha_L^2)\theta}{2} \right] \end{aligned}$$

Disputant 2's interim expected payoff for each type can be similarly defined by exchanging the indices in the superscripts and subscripts. Thus, we shall proceed by focusing on disputant 1. To solve for an equilibrium, we first compute

how each type's payoff changes as his choice changes:

$$\begin{aligned}
\frac{\partial u_{1H}}{\partial \alpha_H^1} &= q_2 \alpha_H^2 \left(\frac{1-\theta}{2} \right) - (1-q_2)\alpha_L^2 \left(p\theta - \frac{1}{2} \right) \\
&= (1-q_2) \left(\frac{1-\theta}{2} \right) \left[\gamma_2 \alpha_H^2 - \alpha_L^2 \delta \right] \\
\frac{\partial u_{1L}}{\partial \alpha_L^1} &= q_2 \alpha_H^2 \left(\frac{1}{2} - (1-p)\theta \right) + (1-q_2)\alpha_L^2 \left(\frac{1-\theta}{2} \right) \\
&= (1-q_2) \left(\frac{1-\theta}{2} \right) \left[\gamma_2 \alpha_H^2 (\delta + 2) + \alpha_L^2 \right]
\end{aligned}$$

As $\frac{\partial u_{1L}}{\partial \alpha_L^1} \geq 0$, it is easy to see that 'Stay' ($\alpha_L^i = 1, i = 1, 2$) is a weakly dominant strategy for L-type. For H-type, however, the H-type opponent's strategy matters to determine the best response. First of all, we have the following lemma:

Lemma 15. *Suppose that L-type chooses his weakly dominant strategy "stay". For disputant i of H-type, if $\gamma_j \leq \delta$ for $j \neq i$, "attack" is a weakly dominant strategy, i.e. "attack" yields a strictly higher expected payoff than "stay" unless the other disputant j of H-type chooses peace with probability* 1.

Proof. Consider, without loss of generality, disputant 1 of H-type. His expected utility is

$$\begin{aligned}
\frac{\partial u_{1H}}{\partial \alpha_H^1} &= (1-q_2) \left(\frac{1-\theta}{2} \right) \left[\gamma_2 \alpha_H^2 - \delta \right] \\
&\leq (1-q_2) \left(\frac{1-\theta}{2} \right) \gamma_2 \left(\alpha_H^2 - 1 \right) \leq 0
\end{aligned}$$

and equality holds if and only if $\gamma_2 = \delta$. □

For $\gamma_2 > \delta$, the best response of the H-type disputant depends the H-type opponent's strategy. If the H-type opponent's strategy is to "Attack" with probability more than $\alpha_H^2 < \frac{\delta}{\gamma_2}$, the best response would be to "Attack" as well. Based on this one can easily obtain the following:

Proposition 1. *The set of undominated Bayesian Nash equilibria of the War-and-Peace game consists of the following:*

(1) $\gamma_1 \leq \delta$ *or* $\gamma_2 \leq \delta$*: L-type chooses to stay and H-type chooses to attack. The equilibrium outcome is war with probability* $(1 - q_1)(1 - q_2)$.

(2) $\gamma_1 > \delta$ *and* $\gamma_2 > \delta$*:*

(a) L-type chooses to stay and H-type chooses to attack. The equilibrium outcome is war with probability $(1 - q_1)(1 - q_2)$.

(b) Both types chooses to stay. The equilibrium outcome is peace.

(c) (mixed strategy equilibrium) L-type chooses to stay and H-type disputant i chooses to stay with probability $\frac{\delta}{\gamma_i}$*. The equilibrium outcome is peace with probability*

$$\left[q_1\left(1 - \frac{\delta}{\gamma_1}\right) - 1\right]\left[q_2\left(1 - \frac{\delta}{\gamma_2}\right) - 1\right]$$

For $\gamma_1 > \delta$ and $\gamma_2 > \delta$, the payoff of H-type in each equilibrium is $u_{1H}(a) = u_{1H}(c) = q_2\left(\frac{\theta}{2}\right) + (1 - q_2)p\theta$ and $u_{1H}(b) = \frac{1}{2}$. The equilibrium (b) yields the highest payoff: $u_{1H}(b) > u_{1H}(a) = u_{1H}(c)$. For $\gamma_1 > \delta$ and $\gamma_2 > \delta$, the equilibrium strategy profile in which H-type chooses to "Stay" Pareto-dominates the other equilibrium strategy profiles. We thus assume that the most efficient equilibrium is chosen among the multiple undominated Bayesian Nash equilibria. In all, we have the following:

Corollary 5. *The set of the most efficient undominated Bayesian Nash equilibria of the War-and-Peace game consists of the following:*

(1) $\gamma_1 \leq \delta$ *or* $\gamma_2 \leq \delta$*: L-type chooses to stay and H-type chooses to attack. The equilibrium outcome is war with probability* $(1-q_1)(1-q_2)$*. The expected payoff for disputant* i *of L-type is* $q_j(1-p)\theta + (1-q_j)\left(\frac{1}{2}\right)$*, and for H-type the expected payoff is* $q_j\left(\frac{\theta}{2}\right) + (1-q_j)p\theta$ *for* $j \neq i$*.*

(2) $\gamma_1 > \delta$ *and* $\gamma_2 > \delta$*: Both types chooses to stay. The equilibrium outcome is peace. The expected payoff for both L-type and H-type is* $\frac{1}{2}$*.*

In short, L-type always chooses to stay, while H-type chooses to stay if and only if $\gamma > \delta$*.*

A.2 Formulation of The Mediation Programme

Formulation of a direct mechanism A direct mechanism consists of the following two functions:

- an action rule $\phi : \{H, L\}^2 \to \triangle(\{A, S\}^2)$ and
- a split rule $\beta : \{H, L\}^2 \to \Delta = \{(b_i', b_2') | b_1' + b_2' = 1\}$ that specifies the share b_i' of disputant $i = 1, 2$.

where $\triangle(\{A, S\}^2)$ is the set of probability distributions over the set $\{A, S\}^2$. Note that an action rule can be shortened to be a rule that assigns peace probability for each type profile. Specifically, let $\phi(\tau) = (\phi(\tau)(SS), \phi(\tau)(SA), \phi(\tau)(AS), \phi(\tau)(AA))$ for a type profile $\tau \in T = \{H, L\}$. Note that a war may break out unilaterally and that the payoff relevant information contained in an action rule is whether

a war breaks out or not. Therefore, one may summarize the action rule by defining the peace probability as $p_\tau \equiv \phi(\tau)(SS)$ and $1 - p_\tau \equiv \phi(\tau)(SA) + \phi(\tau)(AS) + \phi(\tau)(AA)$ for $\tau \in \{H, L\}^2$. For a split rule $\beta(\tau) = (\beta_1(\tau), \beta_2(\tau))$, one can consider only the share of disputant 1 because the share of disputant 2 can be computed simply by $b'_2 = 1 - b'_1$. Hence, define b_τ to denote the share of disputant 1 under peace. To sum up, a direct mechanism is a tuple $\Gamma = (p_{HH}, p_{HL}, p_{LH}, p_{LL}, b_{HH}, b_{HL}, b_{LH}, b_{LL})$.

Derivation of the expected payoff of the mediator Let b_τ and d_τ denote the size of the pie allocated to disputant 1 when the type profile is τ under peace and under war, respectively. The size of the pie allocated to disputant 2 under peace is thus $1 - b_\tau$, for the size of the pie is one. Similarly, the share allocated to disputant 2 under war is $\theta - d_\tau$, for the size of the pie shrinks to θ. The mediator's expected payoff, given the type τ of disputant 1, is thus

$$\begin{aligned} w_\lambda(\tau) &= p_\tau[\lambda(b_\tau - d_\tau) + (1-\lambda)\{(1-b_\tau) - (\theta - d_\tau)\}] \\ &= [(2\lambda - 1)p_\tau(b_\tau - d_\tau) + (1-\lambda)(1-\theta)p_\tau] \\ &\quad \text{or, equivalently} \\ &= [(1-2\lambda)p_\tau\{(1-b_\tau) - (\theta - d_\tau)\} + \lambda(1-\theta)p_\tau] \end{aligned}$$

Therefore, the ex-ante expected payoff of the mediator for given q_1 and q_2 is:

$$\begin{aligned}
W_\lambda &= E_\tau[w_\lambda(\tau)] = E_\tau[p_\tau[\lambda(b_\tau - d_\tau) + (1-\lambda)\{(1-b_\tau) - (\theta - d_\tau)\}]] \\
&= (2\lambda - 1)\left[q_1 q_2 p_{HH}\left(b_{HH} - \frac{\theta}{2}\right) + q_1(1-q_2)p_{HL}(b_{HL} - p\theta)\right] \\
&\quad +(2\lambda-1)\left[q_2(1-q_1)p_{LH}(b_{LH} - (1-p)\theta) + (1-q_1)(1-q_2)p_{LL}\left(b_{LL} - \frac{\theta}{2}\right)\right] \\
&\quad +(1-\lambda)(1-\theta)\left[q_1 q_2 p_{HH} + q_1(1-q_2)p_{HL} + (1-q_1)q_2 p_{LH} + (1-q_1)(1-q_2)p_{LL}\right] \\
&\quad \text{or, equivalently,} \\
&= (1-2\lambda)\left[q_1 q_2 p_{HH}\left(1 - b_{HH} - \frac{\theta}{2}\right) + q_1(1-q_2)p_{HL}(1 - b_{HL} - (1-p)\theta)\right] \\
&\quad +(1-2\lambda)\left[q_2(1-q_1)p_{LH}(1 - b_{LH} - p\theta) + (1-q_1)(1-q_2)p_{LL}\left(1 - b_{LL} - \frac{\theta}{2}\right)\right] \\
&\quad +\lambda(1-\theta)\left[q_1 q_2 p_{HH} + q_1(1-q_2)p_{HL} + (1-q_1)q_2 p_{LH} + (1-q_1)(1-q_2)p_{LL}\right]
\end{aligned}$$

The expression in the main body of the paper is the first one.

Lemma 16. *The following statement holds:*

(1) If $\lambda = \frac{1}{2}$, the mediator's expected payoff is the expected probability of peace up to a constant.

(2) If $\lambda = 1$, the mediator's expected payoff is the ex-ante expected gain of disputant 1*.*

(3) If $\lambda = 0$, the mediator's expected payoff is the ex-ante expected gain of disputant 2*.*

Proof. The statements (2) and (3) are obvious from the expressions. To show (1), plug $\lambda = \frac{1}{2}$ into the mediator's expected payoff. Then, we have

$$W_{1/2}(b_H, b_L, p_H, p_L) = \frac{1-\theta}{2}\left[q_1 q_2 p_{HH} + q_1(1-q_2)p_{HL} + (1-q_1)q_2 p_{LH} + (1-q_1)(1-q_2)p_{LL}\right]$$

Since the constant multiplicative term does not affect the maximization problem, we can abstract away to obtain the following: $q_1 q_2 p_{HH} + q_1(1-q_2)p_{HL} + (1-q_1)q_2 p_{LH} + (1-q_1)(1-q_2)p_{LL}$. This is nothing but the expected probability of peace. □

Reformulation of a mediator's problem Let B_τ denote the share of $(1-\theta)$ (that would be lost under war) allocated to disputant 1 of type $\tau \in \{H, L\}$ conditional on the event that peace is achieved, i.e. $B_\tau = \frac{b_\tau - d_\tau}{1-\theta}$ where b_τ and d_τ are the size of the pie allocated to disputant 1 of type τ under peace and under war, respectively. Let $\gamma_i = \frac{q_i}{1-q_i}$ be the likelihood of disputant i $(i = 1, 2)$ being H-type.

The mediation problem by a mediator with the bias λ and the posterior likelihood γ_1 and γ_2 about disputant 1 and 2 being H-type, can be reformulated in the following way:

$$
\begin{aligned}
\max \quad & W_\lambda(B_{HH}, B_{HL}, B_{LH}, B_{LL}, p_{HH}, p_{HL}, p_{LH}, p_{LL}) \\
& = (2\lambda - 1)\left[\gamma_1\gamma_2 p_{HH}B_{HH} + \gamma_1 p_{HL}B_{HL} + \gamma_2 p_{LH}B_{LH} + p_{LL}B_{LL}\right] \\
& +(1-\lambda)\left[\gamma_1\gamma_2 p_{HH} + \gamma_1 p_{HL} + \gamma_2 p_{LH} + p_{LL}\right] \\
(IC_{1H}) \quad & \gamma_2 p_{HH}B_{HH} + p_{HL}B_{HL} \geq \gamma_2 p_{LH}\left(B_{LH} - \tfrac{1+\delta}{2}\right) + p_{LL}\left(B_{LL} - \tfrac{1+\delta}{2}\right) \\
(IC_{1L}) \quad & \gamma_2 p_{LH}B_{LH} + p_{LL}B_{LL} \geq \gamma_2 p_{HH}\left(B_{HH} + \tfrac{1+\delta}{2}\right) + p_{HL}\left(B_{HL} + \tfrac{1+\delta}{2}\right) \\
(PC_{1H}) \quad & \gamma_2 p_{HH}B_{HH} + p_{HL}B_{HL} \geq 0 \\
(PC_{1L}) \quad & \gamma_2 p_{LH}B_{LH} + p_{LL}B_{LL} \geq 0 \\
(IC_{2H}) \quad & \gamma_1 p_{HH}\left(1 - B_{HH}\right) + p_{LH}(1 - B_{LH}) \geq \gamma_1 p_{HL}\left(1 - B_{HL} - \tfrac{1+\delta}{2}\right) + p_{LL}\left(1 - B_{LL} - \tfrac{1+\delta}{2}\right) \\
(IC_{2L}) \quad & \gamma_1 p_{HL}(1 - B_{HL}) + p_{LL}\left(1 - B_{LL}\right) \geq \gamma_1 p_{HH}\left(1 - B_{HH} + \tfrac{1+\delta}{2}\right) + p_{LH}\left(1 - B_{LH} + \tfrac{1+\delta}{2}\right) \\
(PC_{2H}) \quad & \gamma_1 p_{HH}\left(1 - B_{HH}\right) + p_{LH}(1 - B_{LH}) \geq 0 \\
(PC_{2L}) \quad & \gamma_1 p_{HL}(1 - B_{HL}) + p_{LL}\left(1 - B_{LL}\right) \geq 0
\end{aligned}
$$

Note that since $b_\tau \in [0,1]$, $B_{HH}, B_{LL} \in \left[-\frac{\frac{\theta}{2}}{1-\theta}, \frac{1-\frac{\theta}{2}}{1-\theta}\right]$, $B_{HL} \in \left[-\frac{1-\frac{\theta}{2}}{1-\theta}, \frac{\frac{\theta}{2}}{1-\theta}\right]$ and $B_{LH} \in \left[1 - \frac{\frac{\theta}{2}}{1-\theta}, 1 + \frac{1-\frac{\theta}{2}}{1-\theta}\right]$. (If ex-post participation constraints are considered, $p_\tau B_\tau \in [0,1]$ for $\tau \in \{HH, HL, LH, LL\}$.)

One advantage of this reformulation is to make one treat the type-dependent outside options in the participation constraints as if it is type-independent by normalizing them to be zero: Notice the right-hand sides of (PC_{iH}) and (PC_{iL})

for $i = 1, 2$ are all zeros. More importantly, with this reformulation, one may see whether the optimal mediation mechanism is self-enforcing or not by looking at the additional gain or loss to the payoff under war. If the values of B_τ for $\tau \in \{H, L\}^2$ belong to the unit interval in the optimal mechanism, then the mechanism is self-enforcing: The mechanism maximizes the mediator's payoff by allocating the share $(1-\theta)$ while making no disputant worse than its ex-post payoff under war.

A.3 Optimal Mechanism By a Biased Mediator under Pooling Strategy

In this section, we analyze the optimal mechanism under pooling strategy, i.e. $\gamma_1 = \gamma_2 = \gamma < \delta$. The results in this section indeed constitute the proofs for Lemma 13, Theorem 6, and Corollary 3.

Lemma 17. *The participation constraints for each disputant of L-type, (PC_{iL}), is non-binding.*

Proof. One can easily see that for each disputant $i = 1, 2$,

- RHS of (IC_{iL}) > LHS of (IC_{iH}) =LHS of (PC_{iH}) and
- RHS of (IC_{iH}) < LHS of (IC_{iL}) =LHS of (PC_{iL}).

By the above inequalities, (IC_L) and (PC_H) imply (PC_L):

$$\begin{aligned} LHS(PC_{iL}) &= LHS(IC_{iL}) \geq RHS(IC_{iL}), \text{ by } (IC_{iL}) \\ &> LHS(IC_{iH}) = LHS(PC_{iH}), \text{ by } (PC_{iH}) \\ &\geq 0 \end{aligned}$$

□

When the mediator is biased ($\lambda \neq \frac{1}{2}$), we may restrict our attention only to the case in which the mediator is biased toward disputant 1 ($\lambda > \frac{1}{2}$), for the other case corresponds to the case where the mediator is biased in favor of disputant 2.

Lemma 18. *The participation constraint for disputant 2 of H-type and the incentive compatibility constraints of L-type for both disputants are binding at the optimum. That is,* (IC_{1L})*,* (IC_{2L})*, and* (PC_{2H}) *hold with equality.*

Proof. First of all, (PC_{2H}) is binding. If not, raise B_{HH} and B_{LH} simultaneously in order to keep (IC_{1L}). Specifically, if the RHS is away from the LHS by Δ, then raise B_{HH} by $\left(\frac{p_{LH}}{p_{LH}+p_{HH}}\right)\Delta$ while increasing B_{LH} by $\left(\frac{p_{HH}}{p_{LH}+p_{HH}}\right)\Delta$. Both sides of $(IC_{1L}$ thus increase by the same amount. This procedure leads to the increase in the value of the objective function without violating the other two constraints. For (IC_{2L}), if it is not binding, one may raise the value of the objective function by increasing B_{HL} and B_{LL} simultaneously while keeping (IC_{1L}). Similarly to the previous case, if the discrepancy between the RHS and the LHS of (IC_{2L}) is Δ, then raise B_{LL} by $\left(\frac{p_{HL}}{p_{HL}+p_{LL}}\right)\Delta$ and B_{HL} by $\left(\frac{p_{LL}}{p_{HL}+p_{LL}}\right)\Delta$.

Lastly, (IC_{1L}) is binding at the optimum. Suppose that (IC_{1L}) holds with strict inequality. Notice first that due to (PC_{1H}), at least one of B_{HH} or B_{HL} is non-negative. Moreover, (IC_{2L}) implies that either $1 - B_{HL}$ or $1 - B_{LL}$ is non-negative. If $B_{HL} \in [0, 1]$, we may raise p_{HL} and thus the value of W_λ without violating the other two constraints, (IC_{2L}) and (PC_{2H}). Suppose that $B_{HL} > 1$. By (IC_{2L}), either $B_{LL} < 0$ or $B_{LL} \in [0, 1]$. In the former case, raise p_{LL}. In

the latter case, one may raise p_{HL} and p_{LL} simultaneously without violating the other two constraints. □

By the previous lemma, the mediation programme can be expressed as follows:

$$\max \quad \lambda\left(\gamma p_{HL}+p_{LL}\right)+\left[\lambda\gamma-(2\lambda-1)\left(\frac{1+\delta}{2}\right)\right]\left(\gamma p_{HH}+p_{LH}\right)$$

$$(IC_{1H}) \quad \gamma p_{LH}+p_{LL}\geq\gamma p_{HH}+p_{HL}$$

$$(PC_{1H}) \quad p_{LL}+\left(\gamma-\tfrac{1+\delta}{2}\right)p_{LH}\geq\left[\left(\tfrac{1+\delta}{2}\right)^2-\left(\tfrac{1+\delta}{2}-\gamma\right)^2\right]p_{HH}+\left(\tfrac{1+\delta}{2}-\gamma\right)p_{HL}$$

$$(IC_{2H}) \quad \gamma p_{HL}+p_{LL}\geq\gamma p_{HH}+p_{LH}$$

with the following constraints for B_τ with $\tau\in\{H,L\}^2$:

$$(IC_{1L}) \quad \gamma p_{LH}B_{LH}+p_{LL}B_{LL}=\gamma p_{HH}\left(B_{HH}+\frac{1+\delta}{2}\right)+p_{HL}\left(B_{HH}+\frac{1+\delta}{2}\right)$$

$$(IC_{2L}) \quad \gamma p_{HL}(1-B_{HL})+p_{LL}(1-B_{LL})=\left(\gamma p_{HH}+p_{LH}\right)\left(\frac{1+\delta}{2}\right)$$

$$(PC_{2H}) \quad \gamma p_{HH}(1-B_{HH})+p_{LH}(1-B_{LH})=0$$

First of all, notice that $p_{LL}=1$ at the optimum, for it appears only on the left-hand side of each constraint. Then, we analyze the programme in two steps: In the first step, we treat p_{HH} and p_{LH} as fixed and solve for p_{HL} that maximizes $\lambda\left(\gamma_1 p_{HL}+1\right)$. In the second step, after plugging p_{LH}, we choose p_{HH} and p_{LH} that maximizes the objective function W_λ.

Step 1 : For fixed p_{HL}, p_{LL}, the optimal mechanism solves the following programme:

$$
\begin{aligned}
(\mathbf{P1}) \quad \max \quad & \lambda\left(\gamma p_{HL}+1\right) \\
\text{s.t.} \quad & (IC_{1H}) \quad 1-p_{HL} \geq \gamma(p_{HH}-p_{LH}) \\
& (PC_{1H}) \quad 1-\left(\frac{1+\delta}{2}-\gamma\right) p_{HL} \geq \gamma(1+\delta-\gamma) p_{HH}+\left(\frac{1+\delta}{2}-\gamma\right) p_{LH} \\
& (PROB) \quad p_{HL} \in [0,1]
\end{aligned}
$$

If $\gamma \leq \frac{1+\delta}{2}$, then the feasible set for p_{HL} is identified by the following constraints:

$$
\begin{aligned}
& (IC_{1H}) \quad p_{HL} \leq 1-\gamma(p_{HH}-p_{LH}) \\
& (PC_{1H}) \quad p_{HL} \leq \frac{1-\gamma(1+\delta-\gamma) p_{HH}-\left(\frac{1+\delta}{2}-\gamma\right) p_{LH}}{\frac{1+\delta}{2}-\gamma} \\
& (PROB) \quad p_{HL} \in [0,1]
\end{aligned}
$$

For the coefficient $\lambda\gamma$ on p_{HL} is positive, the optimal solution for p_{HL}, if it exists, is

$$
\min \left\{1, 1-\gamma(p_{HH}-p_{LH}), \frac{1-\gamma(1+\delta-\gamma) p_{HH}-\left(\frac{1+\delta}{2}-\gamma\right) p_{LH}}{\frac{1+\delta}{2}-\gamma}\right\} \geq 0
$$

If $\gamma \in \left(\frac{1+\delta}{2}, \delta\right)$, then the feasible set for p_{HL} can be written as the following:

$$
\begin{aligned}
& (IC_{1H}) \quad p_{HL} \leq 1-\gamma(p_{HH}-p_{LH}) \\
& (PC_{1H}) \quad p_{HL} \geq \frac{1-\gamma(1+\delta-\gamma) p_{HH}-\left(\frac{1+\delta}{2}-\gamma\right) p_{LH}}{\frac{1+\delta}{2}-\gamma} \\
& (PROB) \quad p_{HL} \in [0,1]
\end{aligned}
$$

or, alternatively, $p_{HL} \in \left[\max\left\{0, \frac{1-\gamma(1+\delta-\gamma)p_{HH}-\left(\frac{1+\delta}{2}-\gamma\right)p_{LH}}{\frac{1+\delta}{2}-\gamma}\right\}, \min\{1, 1-\gamma(p_{HH}-p_{LH})\}\right]$. The optimal solution for p_{HL}, if it exists, thus occurs at the supremum of the feasible set:

$$p_{HL} = \min\{1, 1-\gamma_2(p_{HH}-p_{LH})\}.$$

Step 2 : Firstly, we remark on the objective function for each values of p_{HL}.

(1) When $p_{HL} = 1$, the objective function $W_\lambda(p_{HH}, p_{LH})$ is written as follows:

$$W_\lambda^{(1)} \equiv \lambda(\gamma+1) + [(1+\delta)/2 - \lambda(1+\delta-\gamma)](\gamma p_{HH} + p_{LH})$$

Let $\hat{\lambda} = \frac{1+\delta}{2(1+\delta-\gamma)}$. If $\lambda < \hat{\lambda}$, then the coefficients on p_{HH} and p_{LH} are positive.

(2) When $p_{HL} = 1-\gamma(p_{HH}-p_{LH})$, the objective function $W_\lambda(p_{HH}, p_{LH})$ is written as follows:

$$W_\lambda^{(2)} \equiv \lambda(\gamma+1) - (2\lambda-1)\left(\frac{1+\delta}{2}\right)\gamma p_{HH} + \left[\left(\frac{1+\delta}{2}\right) - \lambda(1+\delta-\gamma-\gamma^2)\right] p_{LH}$$

The coefficient on p_{HH} is negative.

(3) When $p_{HL} = \frac{1-\gamma(1+\delta-\gamma)p_{HH}-\left(\frac{1+\delta}{2}-\gamma\right)p_{LH}}{\frac{1+\delta}{2}-\gamma}$, the objective function is thus written as follows:

$$W_\lambda^{(3)} \equiv \left(\frac{1+\delta}{2}\right)\left[\frac{\lambda}{\frac{1+\delta}{2}-\gamma} + \frac{\gamma p_{HH}}{\frac{1+\delta}{2}-\gamma}\left\{\left(\frac{1+\delta}{2}-\gamma\right) - \lambda(1+\delta-\gamma)\right\} + p_{LH}(1-2\lambda)\right].$$

For $\lambda \in (1/2, 1)$, both coefficients on p_{HH} and p_{LH} are negative, for $\gamma\left[\left(\frac{1+\delta}{2}-\gamma\right) - \lambda(1+\delta-\gamma)\right] = \frac{1+\delta}{2}(1-2\lambda) - \gamma(1-\lambda) < 0$ and $1-2\lambda < 0$.

Now, we analyze the mediation programme for the following cases: (A) $\gamma \leq \frac{1+\delta}{2}$ and (B) $\gamma \in \left(\frac{1+\delta}{2}, \delta\right)$.

(A) $\gamma \leq \frac{1+\delta}{2}$**:** We first characterize the feasible set (conditional on p_{HL}). As $\gamma < \frac{1+\delta}{2}$, either $p_{HL} = 1$, $p_{HL} = 1-\gamma_2(p_{HH}-p_{LH})$, or $p_{HL} = \frac{1-\gamma(1+\delta-\gamma)p_{HH}-\left(\frac{1+\delta}{2}-\gamma\right)p_{LH}}{\frac{1+\delta}{2}-\gamma}$. Then, for $p_{HL} = 1$, the feasible set for (p_{HH}, p_{LH}) is defined by the following inequalities:

$$p_{LH} \geq p_{HH}$$

$$p_{LH} \leq -\left(\frac{2\gamma(1+\delta-\gamma)}{1+\delta-2\gamma}\right) p_{HH} + \frac{1-\delta+2\gamma}{1+\delta-2\gamma}$$

As for $p_{HL} = 1-\gamma(p_{HH}-p_{LH})$, the feasible set is thus characterized by

$$p_{LH} \leq p_{HH}$$

$$p_{LH}(\gamma+1) \leq -\left(\frac{\gamma(1+\delta)}{1+\delta-2\gamma}\right) p_{HH} + \frac{1-\delta+2\gamma}{1+\delta-2\gamma}$$

Lastly, for $p_{HL} = \frac{1-\gamma(1+\delta-\gamma)p_{HH}-\left(\frac{1+\delta}{2}-\gamma\right)p_{LH}}{\frac{1+\delta}{2}-\gamma}$, the feasible set is defined by following inequalities:

$$p_{HL} = \frac{2-2\gamma(1+\delta-\gamma)p_{HH}}{1+\delta-2\gamma} - p_{LH} \geq 0$$

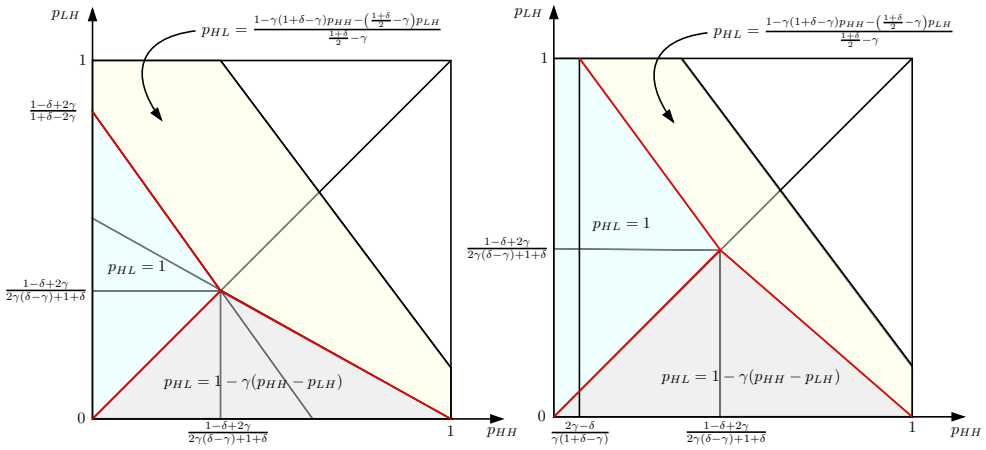
$$p_{HH}, p_{LH} \in [0,1]$$

$$p_{LH} \geq -\left(\frac{2\gamma(1+\delta-\gamma)}{1+\delta-2\gamma}\right) p_{HH} + \frac{1-\delta+2\gamma}{1+\delta-2\gamma}$$

$$p_{LH}(\gamma+1) \geq -\left(\frac{\gamma(1+\delta)}{1+\delta-2\gamma}\right) p_{HH} + \frac{1-\delta+2\gamma}{1+\delta-2\gamma}$$

The feasible set in each case is thus depicted in Figure A.1.

(1) $p_{HL} = 1$: For $\lambda < \hat{\lambda} \equiv \frac{1+\delta}{2(1+\delta-\gamma)}$, both p_{HH} and p_{LH} contribute positively to the objective function, but they cannot both be one due to the constraint



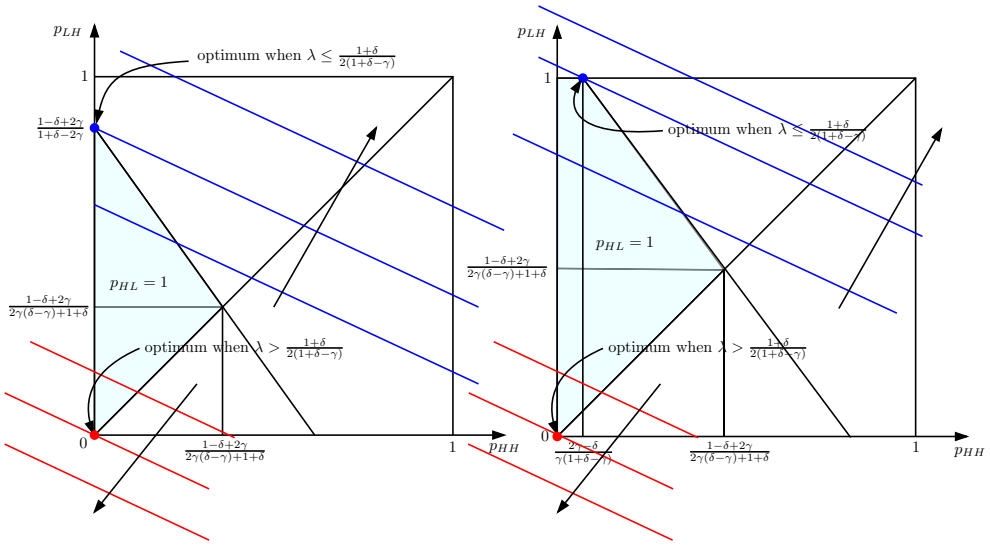
(a) $\gamma \leq \frac{\delta}{2}$ (b) $\gamma > \frac{\delta}{2}$

Figure A.1: Feasible set for (p_{HH}, p_{LH}): $\gamma \leq \frac{1+\delta}{2}$

(PC_{1H}). As the marginal rate of substitution (MRS) γ is less steeper than the slope of (PC_{1H}) which is $\frac{\gamma(1+\delta-\gamma)}{\frac{1+\delta}{2}-\gamma}$, raising p_{LH} always contributes more to the objective function than raising p_{HH} does. Thus, the optimum occurs at the point that maximizes p_{LH} in sacrifice of p_{HH}. When $\gamma > \delta/2$, p_{LH} cannot be larger than one, and p_{HH} cannot decrease further. Therefore, $p_{LH} = 1$ and $p_{HH} = \frac{2\gamma-\delta}{\gamma(1+\delta-\gamma)}$. On the other hand, if $\gamma \leq \delta/2$, (PC_{1H}) binds and $p_{LH} = \frac{1-\delta+2\gamma}{1+\delta-2\gamma}$ while $p_{HH} = 0$.

If $\lambda \geq \hat{\lambda}$, then both p_{HH} and p_{LH} contribute negatively to the objective function. Hence, $p_{HH} = p_{LH} = 0$ at the optimum, for it is feasible.

(2) $p_{HL} = 1 - \gamma(p_{HH} - p_{LH})$: If $\lambda \leq \hat{\lambda}$, then the coefficient on p_{LH} is positive and larger than the absolute value of the coefficient on p_{HH}: $\frac{1+\delta}{2} - \lambda(1 + \delta - \gamma(\gamma+1)) > (2\lambda - 1)\gamma\frac{1+\delta}{2}$. That is, raising p_{LH} contributes more to the objective function than reducing p_{HH}. The optimum thus occurs at $p_{HH} = p_{LH} = \frac{1-\delta+2\gamma}{2\gamma(\delta-\gamma)+1+\delta}$. At the optimum, $p_{HL} = 1$ and both (PC_{1H})



(a) $\gamma \leq \frac{\delta}{2}$ (b) $\gamma > \frac{\delta}{2}$

Figure A.2: The optimal values for (p_{HH}, p_{LH}): $p_{HL} = 1$

and (IC_{1H}) bind. Otherwise, $p_{HH} = p_{LH} = 0$.

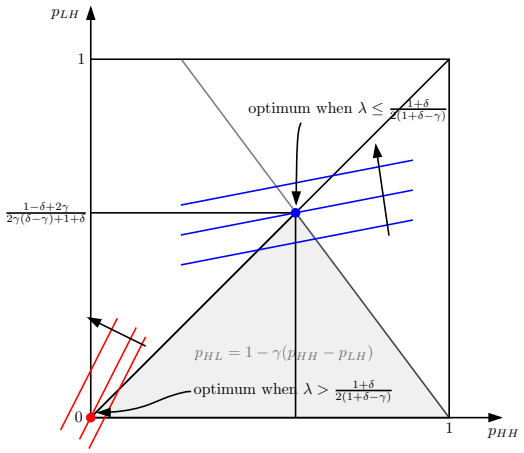
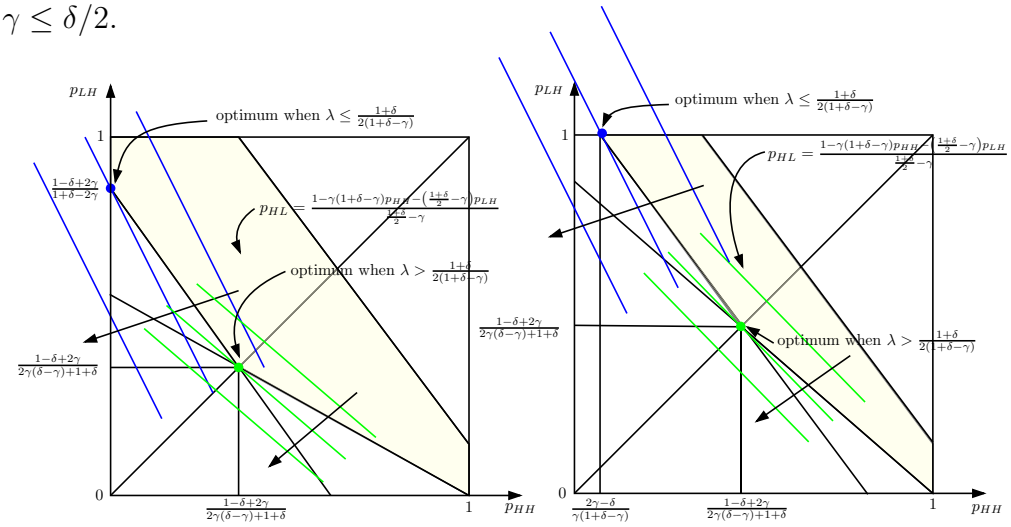


Figure A.3: The optimal values for (p_{HH}, p_{LH}): $p_{HL} = 1 - \gamma(p_{HH} - p_{LH})$

(3) $p_{HL} = \frac{1-\gamma(1+\delta-\gamma)p_{HH}-\left(\frac{1+\delta}{2}-\gamma\right)p_{LH}}{\frac{1+\delta}{2}-\gamma}$: Recall that both coefficients are negative. Thus, decreasing both as much as possible is optimal. Moreover, if $\lambda > \hat{\lambda}$, the marginal rate of substitution (MRS) $\frac{\lambda(1+\delta-\gamma)-\left(\frac{1+\delta}{2}-\gamma\right)}{1-2\lambda}$ is less

steeper than the slope of (PC_{1H}) which is $1+\delta-\gamma$. Therefore, the optimum occurs at $p_{HH}=p_{LH}=\frac{1-\delta+2\gamma}{2\gamma(\delta-\gamma)+1+\delta}$. Otherwise, if $\lambda\leq\hat{\lambda}$, MRS is steeper than the slope of (PC_{1H}), thereby the optimum occurs at $p_{LH}=1$ and $p_{HH}=\frac{2\gamma-\delta}{\gamma(1+\delta-\gamma)}$ for $\gamma>\delta/2$, and $p_{LH}=\frac{1-\delta+2\gamma}{1+\delta-2\gamma}$ and $p_{HH}=0$ for $\gamma\leq\delta/2$.



(a) $\gamma\leq\frac{\delta}{2}$ (b) $\gamma>\frac{\delta}{2}$

Figure A.4: The optimal values for (p_{HH},p_{LH}): $p_{HL}=\frac{1-\gamma(1+\delta-\gamma)p_{HH}-\left(\frac{1+\delta}{2}-\gamma\right)p_{LH}}{\frac{1+\delta}{2}-\gamma}$

In order to determine the optimal mechanism, we compare the values of the objective function. When $\lambda\geq\hat{\lambda}$, all the coefficients on p_{HH} and p_{LH} are negative. Therefore, $p_{HH}=p_{LH}=0$. Indeed, $W_\lambda^{(1)}=W_\lambda^{(2)}\geq W_\lambda^{(3)}$. Suppose otherwise that $\lambda<\hat{\lambda}$. Notice that $W_\lambda^{(1)}=W_\lambda^{(3)}$ and this value is larger than $W_\lambda^{(2)}$.

In all, the optimal mechanism dictates the following:

(i) For $\lambda\geq\hat{\lambda}$, $p_{HH}=p_{LH}=0$ and $p_{LL}=p_{HL}=1$. The interim payoffs for disputant 1 are as follows: $du_{1H}=B_{HL}=1-\frac{1+\delta}{2(1+\gamma)}$, $du_{1L}=B_{LL}=B_{HL}+\frac{1+\delta}{2}=1-\left(\frac{\gamma}{1+\gamma}\right)\left(\frac{1+\delta}{2}\right)$, $du_{2H}=\gamma(1-B_{HH})+(1-B_{LH})=0$, and $du_{2L}=0$.

(ii) For $\lambda < \hat{\lambda}$, $p_{LL} = 1$ and

(a) if $\gamma > \frac{\delta}{2}$, then $p_{LL} = p_{HL} = p_{LH} = 1$ and $p_{HH} = \frac{2\gamma - \delta}{\gamma(1+\delta-\gamma)}$. The interim payoffs for disputant 1 are $du_{1H} = 0$ and $du_{1L} = (\gamma p_{HH} + 1)\left(\frac{1+\delta}{2}\right)$. For disputant 2, $du_{2H} = 0$ and $du_{2L} = (\gamma p_{HH} + 1)\left(\frac{1+\delta}{2}\right)$. The allocation $(B_{HH}, B_{HL}, B_{LH}, B_{LL})$ is thus determined by the following: $du_{1H} = \gamma p_{HH} B_{HH} + B_{HL} = 0$, $du_{1L} = \gamma B_{LH} + B_{LL} = (\gamma p_{HH} + 1)\left(\frac{1+\delta}{2}\right) = \left(\frac{1+\gamma}{1+\delta-\gamma}\right)\left(\frac{1+\delta}{2}\right)$, $du_{2H} = \gamma p_{HH}(1 - B_{HH}) + (1 - B_{LH}) = 0$, and $du_{2L} = \gamma(1 - B_{HL}) + (1 - B_{LL}) = (\gamma p_{HH} + 1)\left(\frac{1+\delta}{2}\right)$.

(b) if $\gamma \leq \frac{\delta}{2}$, then $p_{HL} = 1$, $p_{LH} = \frac{1-\delta+2\gamma}{1+\delta-2\gamma}$, and $p_{HH} = 0$. The interim payoffs for disputant 1 and disputant 2 as well as the allocation are determined as follows: $du_{1H} = B_{HL} = 0$, $du_{1L} = \gamma p_{LH} + B_{LL} = \left(\frac{1+\delta}{2}\right)$, $du_{2H} = p_{LH}(1 - B_{LH}) = 0$, and $du_{2L} = \gamma + (1 - B_{LL}) = p_{LH}\left(\frac{1+\delta}{2}\right)$. That is, $B_{HL} = 0$, $B_{LH} = 1$, $B_{HH} = B_{LL} = \frac{1+\delta}{2} - p_{LH} = \frac{\left(\frac{1+\delta}{2}\right)^2 - \gamma(\gamma+1)}{\frac{1+\delta}{2} - \gamma}$.

(B) $\gamma \in \left(\frac{1+\delta}{2}, \delta\right)$**:** We first characterize the feasible set (conditional on p_{HL}). As $\gamma_1 > \frac{1+\delta}{2}$, either $p_{HL} = 1$ or $p_{HL} = 1 - \gamma(p_{HH} - p_{LH})$ under the condition that $p_{HL} \geq \frac{\gamma(1+\delta-\gamma)p_{HH} + (\frac{1+\delta}{2} - \gamma)p_{LH} - 1}{\gamma - \frac{1+\delta}{2}}$. Then, for $p_{HL} = 1$, the feasible set for (p_{HH}, p_{LH}) is defined by the following inequalities:

$$p_{LH} \geq p_{HH}$$

$$p_{LH} \geq \left(\frac{\gamma(1+\delta-\gamma)}{\gamma - \frac{1+\delta}{2}}\right) p_{HH} + \frac{1-\delta+2\gamma}{1+\delta-2\gamma}$$

On the other hand, for $p_{HL} = 1 - \gamma_2(p_{HH} - p_{LH})$, the feasible set is thus characterized by

$$p_{LH} \leq p_{HH}$$

$$p_{LH}(\gamma + 1) \geq \left[\frac{\gamma\left(\frac{1+\delta}{2}\right)}{\gamma - \frac{1+\delta}{2}}\right] p_{HH} + \frac{1 - \delta + 2\gamma}{1 + \delta - 2\gamma}$$

The feasible set in each case is thus depicted in Figure A.5. The optimum occurs

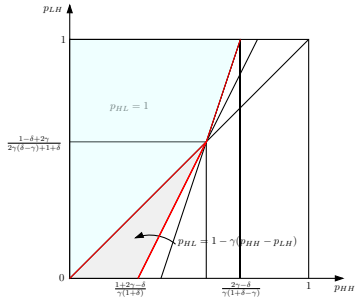
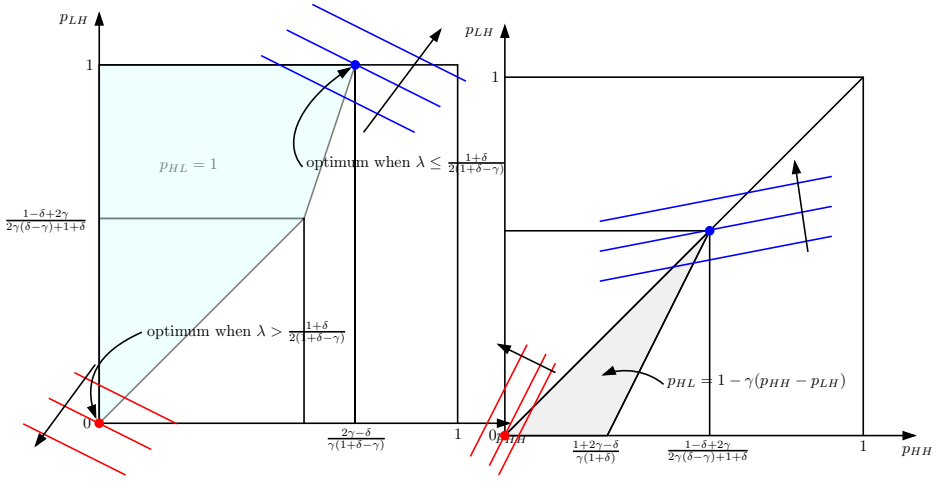


Figure A.5: Feasible set for (p_{HH}, p_{LH}): $\gamma > \frac{1+\delta}{2}$

as in the case of $\gamma > \delta/2$ in (A)



(a) $p_{HL} = 1$ (b) $p_{HL} = 1 - \gamma(p_{HH} - p_{LH})$

Figure A.6: The optimal values for (p_{HH}, p_{LH}): $\gamma > \frac{1+\delta}{2}$

In all, the optimal mechanism dictates the following:

(i) For $\lambda \geq \hat{\lambda}$, $p_{HH} = p_{LH} = 0$ and $p_{LL} = p_{HL} = 1$. The interim payoffs for disputant 1 are as follows: $du_{1H} = B_{HL} = 1 - \frac{1+\delta}{2(1+\gamma)}$, $du_{1L} = B_{LL} = B_{HL} + \frac{1+\delta}{2} = 1 - \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{1+\delta}{2}\right)$, $du_{2H} = \gamma(1 - B_{HH}) + (1 - B_{LH}) = 0$, and $du_{2L} = 0$.

(ii) For $\lambda < \hat{\lambda}$, $p_{LL} = p_{HL} = p_{LH} = 1$ and $p_{HH} = \frac{2\gamma - \delta}{\gamma(1+\delta-\gamma)}$. The interim payoffs for disputant 1 are $du_{1H} = 0$ and $du_{1L} = (\gamma p_{HH} + 1)\left(\frac{1+\delta}{2}\right)$. For disputant 2, $du_{2H} = 0$ and $du_{2L} = (\gamma p_{HH} + 1)\left(\frac{1+\delta}{2}\right)$. The allocation $(B_{HH}, B_{HL}, B_{LH}, B_{LL})$ is thus determined by the following: $du_{1H} = \gamma p_{HH} B_{HH} + B_{HL} = 0$, $du_{1L} = \gamma B_{LH} + B_{LL} = (\gamma p_{HH} + 1)\left(\frac{1+\delta}{2}\right)$, $du_{2H} = \gamma p_{HH}(1 - B_{HH}) + (1 - B_{LH}) = 0$, and $du_{2L} = \gamma(1 - B_{HL}) + (1 - B_{LL}) = (\gamma p_{HH} + 1)\left(\frac{1+\delta}{2}\right)$.

A.4 Proofs for Theorem 7 and Theorem 8.

Proofs for Pooling-Strategy Equilibria: Theorem 7 Without loss of generality, we consider a deviation by disputant 1 when $\lambda \leq 1/2$, for it is symmetric for disputant 2. If deviating, disputant 1's belief is γ and disputant 2's belief is γ', which is arbitrary. Let $du^d_{i\tau_i}$ denote the payoff of disputant i of type τ_i. Suppose that disputant 1 is H-type. Then, disputant $1H$ would attack, and war breaks out unilaterally, regardless of an out-of-equilibrium belief γ' of disputant 2. The payoff under deviation is $du^d_{1H} = 0$. On other hand, suppose that disputant 1 is L-type. Then, the deviation outcome is determined by disputant 2's out-of-equilibrium belief γ'. When $\gamma' \leq \delta$, disputant $2H$ would attack, and war breaks out with probability q. disputant $1L$ obtains $q(1-p)\theta + (1-q)/2$, and thus $du^d_{1L} = 1/2$. For $\gamma' > \delta$, disputant $2H$ would stay, and war

never breaks out. The resulting payoff for $1L$ is $du^d_{1L} = \gamma(\delta + 1) + 1/2$.

(i) Suppose that the chosen mediator is moderately biased against disputant 1, i.e. $\lambda \in \left(1 - \frac{1+\delta}{2(1+\delta-\gamma)}, \frac{1}{2}\right)$. As $du_{1H} = 0$, disputant $1H$ does not gain by deviation. Moreover, for $\gamma > \delta/2$, the mediation payoff is the same as in the previous case, and then disputant $1L$ would not deviate if $\gamma' \leq \delta$. For the remaining case of $\gamma \leq \delta/2$, the mediation payoff is $du_{1L} = \left(\frac{1-\delta+2\gamma}{1+\delta-2\gamma}\right)\left(\frac{1+\delta}{2}\right)$. For $\gamma' \leq \delta$, we have

$$du_{1L} - du^d_{1L} \geq 0 \iff \left(\frac{1-\delta+2\gamma}{1+\delta-2\gamma}\right)\left(\frac{1+\delta}{2}\right) - \frac{1}{2} \geq 0 \iff \gamma \geq \frac{\delta(1+\delta)}{2(2+\delta)}$$

If $\gamma' > \delta$, $du_{1L} - du^d_{1L} = -\frac{\delta^2+(\delta-2\gamma)(1+2\gamma(1+\delta))}{2(1+\delta-2\gamma)} < 0$ Therefore, for $\gamma \geq \frac{\delta(1+\delta)}{2(2+\delta)} \geq 0$ and $\gamma' \leq \delta$, the deviation is not profitable. Otherwise, disputant $1L$ would deviate.

(ii) Suppose that the chosen mediator is unbiased: $\lambda = 1/2$. Recall that the mediation payoffs are the same as those under the moderately biased mediator if $\gamma > \delta/2$. We thus consider the case where $\gamma \leq \delta/2$. The mediation payoff is $du_{lL} = \left(\frac{1}{1+\delta-\gamma}\right)\left(\frac{1+\delta}{2}\right)$ and $du_{1H} = 0$. Obviously, disputant $1H$ does not have any incentive to deviate. For disputant $1L$, if $\gamma' \leq \delta$, we have $du_{1L} - du^d_{1L} = \frac{\gamma}{2(1+\delta-\gamma)} > 0$, and thus deviation is not profitable. When $\gamma' > \delta$, $du_{1L} - du^d_{1L} = -\gamma\frac{(1+2\delta)+(\delta-\gamma)(1+\delta)}{2(1+\delta-\gamma)} < 0$, i.e. disputant $1L$ would deviate.

(iii) Suppose that the chosen mediator is extremely biased against disputant 1. That is, $\lambda < 1 - \frac{1+\delta}{2(1+\delta-\gamma)}$. Under mediation, $du_{1H} = du_{1L} = 0$. Obviously, disputant $1H$ has no incentive for deviation. However, disputant $1L$ would always deviate because $du_{1L} - du^d_{1L} = -du^d_{1L} < 0$.

In all, if $\gamma' \leq \delta$ and a disputant would not deviate from the mediator against him if and only if $\gamma > \frac{\delta(1+\delta)}{2(2+\delta)}$.

Proofs for Separating Strategy Equilibrium: Theorem 8 The proof proceeds in the series of the following lemmas.

Lemma 19. *There exists no separating-strategy equilibrium in which (H, H)-dyad selects a mediator*

Proof. Suppose not, i.e. $v_H^1 = v_H^2 = 1$ for some $\lambda \in [0, 1]$. If $\lambda < 1/2$ (bias against 1) is accepted, then disputant $1H$ obtains $\theta/2$. By deviating to match with L-type under no mediation , however, disputant $1H$ would obtain $p\theta > 1/2$. Symmetrically, for $\lambda > 1/2$, disputant $2H$ has a profitable deviation to no mediation. Lastly, if $\lambda = 1/2$, disputant $1H$ obtains $1/2$. Still, matching with $2L$ under no mediation is a profitable deviation. □

Lemma 20. *There exists no separating-strategy equilibrium in which an asymmetric dyad, (H, L) or (L, H), selects a mediator.*

Proof. Consider $v_H^1 = v_L^2 = 1$ and $\lambda > 1/2$ (bias toward disputant 1). Disputant $2L$ obtains $(1 - p)\theta$. By deviating to match with $1L$ under no mediation, he would obtain at least $1/2 > (1-p)\theta$. Similarly, when $\lambda < 1/2$ and $v_L^1 = v_H^2 = 1$, $1L$ would deviate to match with $2L$. Lastly, if $v_H^1 = v_L^2 = 1$ and $\lambda = 1/2$ (or, symmetrically, $v_L^1 = v_H^2 = 1/2$), then disputant $1H$ obtains $p\theta + \frac{1-\theta}{2}$, which is larger than any payoff of $1L$. Hence, $1L$ would mimic $1H$. □

Lemma 21. *There exists no separating-strategy equilibrium such that (L, L)-dyad selects a biased mediator.*

Proof. Without loss of generality, suppose that there exists, i.e. $v_L^1 = v_L^2 = 1$ and $\lambda > 1/2$. Note that H-type disputants obtains $1/2$ under no mediation.

Disputant $1L$ obtains $1 - \frac{\theta}{2}$ which is larger than the payoff of disputant $1H$. Therefore, disputant $1H$ would deviate to pool himself with $1L$. □

The only remaining possibility is that L-types succeeds in selecting an unbiased mediator ($v_L^1 = v_L^2 = 1$ and $\lambda = 1/2$) and H-types do not accept the mediator ($v_H^1 = v_H^2 = 0$). We show that this is indeed a separating equilibrium. Consider disputant 1 without loss of generality. Disputant 1 of H-type obtains $1/2$. If it deviates to mimic L-type, the resulting payoff is $1/2$ no larger than his equilibrium payoff, thus H-type would not deviate. Now, we argue that L-type does not have any profitable deviation. Suppose that L-type deviates to no mediation. Then, his payoff facing H-type is $(1 - p)\theta$, and this is strictly smaller than his equilibrium payoff of $1/2$.

Bibliography

Al-Najjar, N. I. (2009): "Decision makers as statisticians: Diversity, ambiguity, and learning," *Econometrica*, 77, 1371–1401.

Arrow, K. J. (1966): "Exposition of the theory of choice under uncertainty," *Synthese*, 16, 253–269.

Aumann, R. (1976): "Agreeing to disagree," *The Annals of Statistics*, 4, 1236–1239.

Aumann, R. J. (1999a): "Interactive epistemology 1: Knowledge," *International Journal of Game Theory*, 28, 263–300.

——— (1999b): "Interactive epistemology 2: Probability," *International Journal of Game Theory*, 28, 301–314.

Bacharach, M. (1985): "Some extensions of a claim of Aumann in an axiomatic model of knowledge," *Journal of Economic Theory*, 37, 167–190.

Beardsley, K. (2006): "Politics by Means Other than War: Understanding International Mediation," Ph.D. thesis, San Diego, CA: University of California.

Beber, B. (2012): "International mediation, selection effects, and the question of bias," *Conflict Management and Peace Science*, 29, 397–424.

Bercovitch, J. and S. S. Gartner (2008): *International conflict mediation: new approaches and findings*, London: Routledge.

Billingsley, P. (1995): *Probability and measure*, New Jersey: Wiley.

Blackwell, D. and C. Ryll-Nardzewski (1963): "Non-existence of everywhere proper conditional distributions," *Annals of Mathematical Statistics*, 34, 223–225.

Brandenburger, A. and E. Dekel (1987): "Common knowledge with probability 1," *Journal of Mathematical Economics*, 16, 237–245.

Coase, R. H. (1960): "The problem of social cost," in *Classic Papers in Natural Resource Economics*, Springer, 87–137.

Cornfeld, I. P., S. V. Fomin, and Y. G. Sinai (2012): *Ergodic theory*, vol. 245, New York: Springer.

Dubra, J. and F. Echenique (2004): "Information is not about measurability," *Mathematical Social Sciences*, 47, 177–185.

Favretto, K. (2009): "Should peacemakers take sides? Major power mediation, coercion, and bias," *American Political Science Review*, 103, 248–263.

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

Frazier, D. V. and W. J. Dixon (2006): "Third-party intermediaries and negotiated settlements, 1946–2000," *International Interactions*, 32, 385–408.

Halpern, J. Y. (1999): "Hypothetical knowledge and counterfactual reasoning," *International Journal of Game Theory*, 28, 315–330.

Heifetz, A., M. Meier, and B. C. Schipper (2006): "Interactive unawareness," *Journal of economic theory*, 130, 78–94.

Hervés-Beloso, C. and P. K. Monteiro (2013): "Information and sigma-algebras," *Economic Theory*, 54, 405–418.

Holmström, B. and R. Myerson (1983): "Efficient and durable decision rules with incomplete information," *Econometrica*, 51, 1799–1819.

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

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

Maschler, M., E. Solan, and S. Zamir (2013): *Game theory. Translated from the Hebrew by Ziv Hellman and edited by Mike Borns*, New York: Cambridge University Press.

Melin, M. M., S. S. Gartner, and J. Bercovitch (2013): "Fear of rejection: The puzzle of unaccepted mediation offers in international conflict," *Conflict Management and Peace Science*, 30, 354–368.

MEYER, J.-J. C. (2003): "Modal Epistemic and Doxastic Logic," in *Handbook of Philosophical Logic*, ed. by D. M. Gabbay and F. Guenthner, Dordrecht: Springer Netherlands, vol. 10, 1–38.

ROYDEN, H. L. (1988): *Real analysis*, New York: Prentice Hall, 4 ed.

SAVAGE, L. J. (1972): *The foundations of statistics*, New York: Dover Publication.

SHAFER, G. (1986): "Savage revisited," *Statistical science*, 1, 463–485.

SHORTT, R. (1984): "Products of Blackwell spaces and regular conditional probabilities," *Real Analysis Exchange*, 10, 31–41.

SMITH, W. P. (1985): "Effectiveness of the biased mediator," *Negotiation Journal*, 1, 363–372.

TOUVAL, S. (1975): "Biased intermediaries: Theoretical and historical considerations," *Jerusalem Journal of International Relations*, 1, 51–69.

VILLEGAS, C. (1964): "On qualitative probability σ-algebra," *Annals of Mathematical Statistics*, 35, 1787–1796.

VOHRA, R. (1999): "Incomplete information, incentive compatibility, and the core," *Journal of Economic Theory*, 86, 123–147.

WILSON, R. (1978): "Information, efficiency, and the core of an economy," *Econometrica*, 46, 807–816.

ZHANG, Z. (2009): "Comparison of Information Structures with Infinite States of Nature," Ph.D. thesis, Johns Hopkins University.

Curriculum Vitae

Jong Jae Lee received a B.A in 2009 from Sogang University, Republic of Korea. He then enrolled the Ph.D. program in the Department of Economics at the Johns Hopkins University, United States in 2011.

His primary fields of academic interest fall broadly into the area of microeconomic theory, particularly game theory and economics of information with their applications to political economy and economics of organization.

Moreover, throughout his graduate studies he had the opportunity to accumulate extensive teaching experience ranging from Microeconomic Theory, Topics in Political Economy, International Trade, Econometrics, Monetary Analysis, and Corporate Finance to graduate courses such as Microeconomic Theory, Game Theory and Dynamic Optimization.