Spring 2018

Essays on mechanism design under non-Bayesian frameworks

Huiyi Guo
University of Iowa

Recommended Citation

Guo, Huiyi. "Essays on mechanism design under non-Bayesian frameworks." PhD (Doctor of Philosophy) thesis, University of Iowa, 2018.
https://doi.org/10.17077/etd.n9ojxv4p

Follow this and additional works at: https://ir.uiowa.edu/etd
Part of the Economics Commons

ESSAYS ON MECHANISM DESIGN UNDER NON-BAYESIAN FRAMEWORKS

by

Huiyi Guo

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Economics
in the Graduate College of
The University of Iowa

May 2018

Thesis Supervisor: Professor Nicholas Yannelis

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Huiyi Guo

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Economics at the May 2018 graduation.

Thesis Committee: ____________________________
Nicholas Yannelis, Thesis Supervisor

____
Rabah Amir

____
Luciano de Castro

____
Alexandre Poirier

____
Anne Villamil

ACKNOWLEDGEMENTS

I would first and foremost like to express my deepest gratitude to my advisor Nicholas Yannelis, for his guidance, insights, encouragement, and continuous support throughout my entire graduate study. His instruction and help have made my achievement possible and have contributed to my growth in the profession. Besides, I would also like to thank other professors at the University of Iowa, especially my committee members Rabah Amir, Luciano de Castro, Alexandre Poirier, and Anne Villamil for many helpful suggestions and invaluable comments.

I am grateful to the Department of Economics and the Graduate College for generous financial support. My special thanks go to Renea Jay, whose has provided invaluable administrative support throughout the past six years.

Thank you to all my fellow graduate students in the Department of Economics of the University of Iowa, whose insights and support have benefited me both in research and in life. In particular, I am grateful to Wei He for stimulating discussions and invaluable feedback on the first chapter of my thesis.

I am particularly indebted to my family members and friends Guanqi Guo, Zhumei Tan, Yunfan Yu, and Tianlu Zhang, for their patience, support, unconditional love, and belief in me. Without them, the journey would neither have been started nor completed.

ABSTRACT

One important issue in mechanism design theory is to model agents' behaviors under uncertainty. The classical approach assumes that agents hold commonly known probability assessments towards uncertainty, which has been challenged by economists in many fields. My thesis adopts alternative methods to model agents' behaviors. The new findings contribute to understanding how the mechanism designer can benefit from agents' uncertainty aversion and how she should respond to the lack of information on agents' probability assessments.

Chapter 1 of this thesis allows the mechanism designer to introduce ambiguity to the mechanism. Instead of informing agents of the precise payment rule that she commits to, the mechanism designer can tell agents multiple payment rules that she may have committed to. The multiple payment rules are called ambiguous transfers. As agents do not know which rule is chosen by the designer, they are assumed to make decisions based on the worst-case scenario. Under this assumption, this chapter characterizes when the mechanism designer can obtain the first-best outcomes by introducing ambiguous transfers. Compared to the standard approach where the payment rule is unambiguous, first-best mechanism design becomes possible under a broader information structure. Hence, there are cases when the mechanism designer can benefit from introducing ambiguity.

Chapter 2 assumes that the mechanism designer does not know agents' probability assessments about others' private information. The mechanisms designed to implement the social choice function thus should not depend on the probability assessments, which are

called robust mechanisms. Different from the existing robust mechanism design literature where agents are always assumed to act non-cooperatively, this chapter allows them to communicate and form coalitions. This chapter provides necessary and almost sufficient conditions for robustly implementing a social choice function as an equilibrium that is immune to all coalitional deviations. As there are social choice functions that are only implementable with coalitional structures, this chapter provides insights on when agents should be allowed to communicate. As an extension, when the mechanism designer has no information on which coalitions can be formed, this chapter also provides conditions for robust implementation under all coalition patterns.

Chapter 3 assumes that agents are not probabilistic about others' private information. Instead, when they hold ambiguous assessments about others' information, they make decisions based on the worst-case belief. This chapter provides necessary and almost sufficient conditions on when a social choice goal is implementable under such a behavioral assumption. As there are social choice goals that are only implementable under ambiguous assessments, this chapter provides insights on what information structure is desirable to the mechanism designer.

PUBLIC ABSTRACT

One important issue in mechanism design theory is to model agents' behaviors under uncertainty. The classical approach assumes that agents hold commonly known probability assessments towards uncertainty, which has been challenged by economists in many fields. My thesis adopts alternative methods to model agents' behaviors. The new findings contribute to understanding how the mechanism designer can benefit from agents' uncertainty aversion and how she should respond to the lack of information on agents' probability assessments.

Chapter 1 of this thesis allows the mechanism designer to introduce ambiguity to the mechanism. Instead of informing agents of the precise payment rule that she commits to, the mechanism designer can tell agents multiple payment rules that she may have committed to. The multiple payment rules are called ambiguous transfers. As agents do not know which rule is chosen by the designer, they are assumed to make decisions based on the worst-case scenario. Under this assumption, this chapter characterizes when the mechanism designer can obtain the first-best outcomes by introducing ambiguous transfers. Compared to the standard approach where the payment rule is unambiguous, first-best mechanism design becomes possible under a broader information structure. Hence, there are cases when the mechanism designer can benefit from introducing ambiguity.

Chapter 2 assumes that the mechanism designer does not know agents' probability assessments about others' private information. The mechanisms designed to implement the social choice function thus should not depend on the probability assessments, which are called robust mechanisms. Different from the existing robust mechanism design literature where agents are always assumed to act non-cooperatively, this chapter allows them to communicate and form coalitions. This chapter provides necessary and almost sufficient conditions for robustly implementing a social choice function as an equilibrium that is immune to all coalitional deviations. As there are social choice functions that are only implementable with coalitional structures, this chapter provides insights on when agents should be allowed to communicate. As an extension, when the mechanism designer has no information on which coalitions can be formed, this chapter also provides conditions for robust implementation under all coalition patterns.

Chapter 3 assumes that agents are not probabilistic about others' private information. Instead, when they hold ambiguous assessments about others' information, they make decisions based on the worst-case belief. This chapter provides necessary and almost sufficient conditions on when a social choice goal is implementable under such a behavioral assumption. As there are social choice goals that are only implementable under ambiguous assessments, this chapter provides insights on what information structure is desirable to the mechanism designer.

TABLE OF CONTENTS

APPENDIX

LIST OF TABLES

Table

CHAPTER 1
MECHANISM DESIGN WITH AMBIGUOUS TRANSFERS

1.1 Introduction

Many transaction mechanisms have uncertain rules. For instance, Priceline Express Deals offer travelers a known price for a hotel stay, but the exact name and location of the hotel remain unknown until the completion of payment. Alternatively, some stores run scratch-and-save promotions. Consumers receive scratch cards during check-out, which reveal discounts, and thus the costs of their purchases remain unknown at the time they decide to buy. As a third example, eBay allows sellers of auction-style listings to set hidden reserve prices.

In all the above mechanisms, the mechanism designer introduces uncertainty about the allocation and/or transfer rule without telling agents the underlying probability distribution. The subjective expected utility model can be adopted to describe agents' decision making without an objective probability. However, since Ellsberg (1961), many studies have challenged this model, arguing that decision makers tend to be ambiguity-averse.[1] Therefore, it is important to understand if and how a mechanism designer can benefit from agents' ambiguity aversion. More specifically, we would like to know whether engineering ambiguity on rules of mechanisms can help the designer achieve the first-best outcome.

[1] There is a huge literature studying ambiguity aversion from the perspective of different fields, including (but not limited to) decision theory (e.g., Gilboa and Schmeidler (1989), Klibanoff et al. (2005)), macroeconomics (e.g., Hansen and Sargent (2001, 2008)), finance (e.g. Chen and Epstein (2002), Garlappi et al. (2006)), and experimental and behavioral economics (e.g., Fox and Tversky (1995), Borghans et al. (2009)).

This paper introduces ambiguous transfers to study two problems: full surplus extraction and interim individually rational and ex-post budget-balanced implementation of any ex-post efficient allocation rule. The problem of full surplus extraction aims to design a mechanism in which agents transfer the entire surplus to the designer. A typical example is to establish an auction such that the auctioneer obtains the first-best revenue. The efficient implementation problem constructs an incentive compatible, individually rational, and budget-balanced mechanism such that the socially optimal outcome emerges as an equilibrium. The designs of bilateral trading protocols and public project financing schemes serve as two examples. In our model, the mechanism designer informs agents of the exact allocation rule. She also commits to one transfer rule, but the communication is ambiguous so that agents only know a set of potential ones. Without knowing the adopted transfer rule, agents are assumed to be ambiguity-averse. More specifically, agents are maxmin expected utility maximizers who make decisions based on the worst-case scenario.

In this paper, the Beliefs Determine Preferences (BDP) property is the key condition for the existence of first-best mechanisms with ambiguous transfers. The property, introduced by Neeman (2004), requires that an agent should hold distinct beliefs about others' private information under different types. Essentially, the property calls for correlated information among agents. In a type space with finite dimension and at least two agents, the BDP property holds for all agents generically.

We show that full surplus extraction can be guaranteed via ambiguous transfers if and only if the BDP property is satisfied by all agents. In addition, any efficient allocation rule is implementable via an interim individually rational and ex-post budget-balanced mechanism with ambiguous transfers if and only if the BDP property holds for all agents. The two characterizations are the primary results of this paper. As an extension, we also characterize the condition for efficient implementation under a private value environment. Then, we establish sufficient conditions for efficient implementation when agents' beliefs are not generated from a common prior. Lastly, we discuss the robustness of our sufficiency results under alternative models of ambiguity aversion.

Our key condition, the BDP property, is weaker than Crémer and McLean (1988)'s Convex Independence condition, which is necessary and sufficient for full surplus extraction via a Bayesian mechanism. Convex Independence, together with the Identifiability condition established by Kosenok and Severinov (2008), is necessary and sufficient for implementing any efficient allocation rule via an interim individually rational and ex-post budget-balanced Bayesian mechanism. Hence, under both problems, this paper requires a strictly weaker condition to obtain the first-best outcome compared to the Bayesian approach. As a result, engineering ambiguity deliberately allows the designer to achieve first-best outcomes that are impossible under the Bayesian mechanisms, and thus the use of ambiguous transfers can enhance social efficiency. For example, ambiguous payments can allow the auctioneer to collect the first-best revenue that is impossible under the standard approach. The social planner can also employ ambiguous trading protocols or tax schemes to realize efficient trades or public projects that are impossible otherwise.

We summarize several advantages of the BDP property below. Firstly, compared to Convex Independence, the BDP property imposes weaker restrictions on the cardinality of the type space. For example, in a two-agent problem, where one agent has two types and the other has three, Convex Independence fails for one agent for sure, but the BDP property holds for both generically. Secondly, the Identifiability condition is relaxed along with its associated restriction on the cardinality of the type space.[2] For example, in a three-agent problem where each agent has two types, the Identifiability property fails with positive probability, but the BDP property holds generically. Thirdly, the Bayesian mechanism design literature documents several negative results on get-balanced implementation with two agents, but the BDP property and ambiguous transfers provide a generic solution to such problems, which are fundamental and important in view of the many bilateral trades and bargains occurring every day.[3] Fourthly, the BDP property is easy to check. To verify this property for an agent, we only need to make sure that she never has identical beliefs under different types.

In this paper, the mechanism designer announces an efficient allocation rule and introduces ambiguity in transfer rules only. To see why we impose this restriction, notice that in our second problem (the implementation problem), the allocation rule is exogenous, and thus, it is natural for the mechanism designer to commit to this allocation rule. In our first problem, since the mechanism designer aims to extract the full surplus instead, she endogenously chooses an ex-ante efficient allocation rule. As the efficient rule is often unique

[2] For the first two points, see Section 1.4 for more details.

[3] For example, Myerson and Satterthwaite (1983) demonstrate the impossibility of efficient bilateral trading with independent information. Matsushima (2007) provides a sufficient condition under which individually rational and budget-balanced implementation with two agents cannot be achieved. Kosenok and Severinov (2008)'s necessary and sufficient conditions never hold simultaneously in two-agent environments, which could also be interpreted as an impossibility result even if correlated information is allowed.

in a finite-type framework, the mechanism designer does not have multiple allocation rules to choose from. In a related paper, Di Tillio et al. (2017) study how second-best revenue in an independent private value auction can be improved if the seller introduces ambiguity in both allocation and transfer rules. In fact, only introducing ambiguous transfers under their environment cannot improve the seller's revenue compared to an unambiguous mechanism, and thus ambiguous allocations play an important role. We discuss more on the relationship with this paper in Section 1.1.1. As a by-product, the restriction on the unambiguous allocation rule also helps clarify the scope and limitations of ambiguous transfers.

The paper proceeds as follows. We review the literature in Section 1.1.1 and introduce the environment in Section 2.2. We formalize the mechanism with ambiguous transfers in Section 1.3. The BDP property is introduced and shown to be necessary and sufficient for our primary results in Section 1.4. Section 1.5 extends our primary results along several directions. The Appendix collects all proofs and some examples.

1.1.1 Literature Review

1.1.1.1 Efficient Mechanisms with Independent Information

How to implement efficient allocations is a classical topic in mechanism design theory that has been widely studied in situations such as public good provision and bilateral trading. Individual rationality is a natural requirement as agents can opt out of the mechanism. As a resource constraint, budget balance requires that agents should finance within themselves for the efficient outcome rather than rely on an outside budget-breaker. When either individual rationality or budget balance is required, the literature provides positive results for efficient mechanism design in private value environments. For instance, the VCG mechanism (Vickrey (1961), Clarke (1971), and Groves (1973)) is ex-post individually rational. The AGV mechanism (d'Aspremont and Gérard-Varet (1979)) is ex-post budget-balanced.

However, the literature documents a tension between efficiency, individual rationality, and budget balance, when agents have independent information. For example, in a private value bilateral trading framework, Myerson and Satterthwaite (1983) prove that it is impossible to achieve efficiency with an individually rational and budget-balanced mechanism in general. With multi-dimensional and interdependent values, Dasgupta and Maskin (2000) and Jehiel and Moldovanu (2001) prove that efficient allocations are generically non-implementable.

One goal of the current paper is to design an efficient, individually rational, and budget-balanced mechanism. But instead of assuming independent information, we show that correlation is necessary and sufficient to achieve the goal.

1.1.1.2 Mechanism Design with Correlated Information

With correlated information, first-best mechanism design becomes possible. Crémer and McLean (1985, 1988) establish two conditions to fully extract agents' surplus in private value auctions, among which the Convex Independence condition is necessary and sufficient for full surplus extraction to be a Bayesian Nash equilibrium. In a fixed finite-dimensional type space, if there are at least two agents and no one has more types than all others' type profiles, the condition holds for all agents under almost every prior. Without restricting the dimension, different notions of genericity are adopted in the literature and various conclusions on genericity of Convex Independence (or the weaker BDP property) are made (e.g., Neeman (2004), Heifetz and Neeman (2006), Barelli (2009), Chen and Xiong (2011, 2013), Gizatulina and Hellwig (2014, 2017)). With continuous types, McAfee and Reny (1992) show that approximate full surplus extraction can be achieved. In addition, the recent papers of Liu (2017) and Noda (2015) prove an intertemporal variant of Convex Independence is sufficient for first-best mechanism design in dynamic environments. By introducing ambiguous transfers, Section 1.4.1 of the current paper shows that a weaker condition, the BDP property, becomes necessary and sufficient for full surplus extraction.

In an implementation problem, the allocation rule is exogenously given. Thus, the mechanism designer constructs incentive compatible transfers to achieve the desired outcome. Under the context of exchange economies, McLean and Postlewaite (2002, 2003a,b) propose the notion of informational size and prove the existence of incentive compatible and approximately efficient outcomes when agents have small informational size.[4] Under a mechanism design framework, McLean and Postlewaite (2004, 2015) implement efficient allocation rules via individually rational mechanisms under the BDP property. In their mechanisms, small outside money is needed even when agents are informationally small. Different from these papers, our mechanism for implementation in Section 1.4 is exactly efficient, individually rational, and budget-balanced without imposing any informational smallness assumption.

[4] For related results, see also Sun and Yannelis (2007, 2008).

A few papers study budget-balanced mechanisms with or without independent information, including Matsushima (1991), Aoyagi (1998), Chung (1999), d'Aspremont et al. (2004), Miller et al. (2007), etc.[5] Among these works, d'Aspremont et al. (2004) propose necessary and sufficient conditions for budget-balanced mechanisms. None of these papers requires individual rationality. Also, they assume that there are at least three agents. In fact, d'Aspremont et al. (2004) indicate an impossibility result in implementing efficient allocations via budget-balanced mechanisms with two agents under correlated information. However, we do require individual rationality, and our mechanism with ambiguous transfers works for environments with at least two agents.

Matsushima (2007), Kosenok and Severinov (2008), and Gizatulina and Hellwig (2010) among others design individually rational and budget-balanced mechanisms. Among them, Kosenok and Severinov (2008) propose the Identifiability condition, which along with the Convex Independence condition, is necessary and sufficient for implementing any ex-ante socially rational allocation rule via an interim individually rational and ex-post budget-balanced Bayesian mechanism. The Identifiability condition is generic with at least three agents and under some restrictions on the dimension of the type space, but Convex Independence and Identifiability never hold simultaneously in a two-agent setting. Thus Kosenok and Severinov (2008) imply an impossibility result in efficient, individually rational, and budget-balanced two-agent mechanism design. In our paper, the BDP property is weaker than Convex Independence, and we do not need Identifiability. Moreover, the BDP

[5] Matsushima (1991), Chung (1999), d'Aspremont et al. (2004) only consider private value utility functions. In this case, incentive compatibility can be achieved via a VCG mechanism, rather than via information correlation. Thus, they allow for independent information.

property holds generically in a finite-dimensional type space with at least two agents, and thus we make the impossible possible for two-agent implementation problems.

1.1.1.3 Mechanism Design under Ambiguity

In the growing literature on mechanism design with ambiguity-averse agents, most of the works assume exogenously that agents hold ambiguous beliefs of others' types. For example, Bose et al. (2006) prove that when agents are more ambiguity-averse than the auctioneer, a full insurance transfer rule is optimal in a private value auction. Bose and Daripa (2009) achieve almost full surplus extraction in a dynamic auction by exploiting the dynamic inconsistency of prior-by-prior updating. Bodoh-Creed (2012) characterizes the revenue-maximizing mechanism with a payoff equivalence theorem. de Castro and Yannelis (2018) prove that all Pareto efficient allocations are incentive compatible when agents' ambiguous beliefs are unrestricted. Accordingly, de Castro et al. (2017a,b) implement all Pareto efficient allocations. Under the private value assumption, Wolitzky (2016) establishes a necessary condition for the existence of an efficient, individually rational, and weak budget-balanced mechanism. In an environment with multi-dimensional and interdependent values, Song (2016) quantifies the amount of ambiguity that is necessary and sometimes sufficient for efficient mechanism design. We do not assume exogenous ambiguity in agents' beliefs, which is the biggest difference between the above papers and our work.

Bose and Renou (2014) and Di Tillio et al. (2017) contrast the above works in that ambiguity is endogenously engineered by the mechanism designer. Before the allocation stage, Bose and Renou (2014) let the mechanism designer communicate with agents via an ambiguous device, which generates multiple beliefs. Their paper characterizes social choice functions that are implementable under this method. Our paper is different from Bose and Renou (2014), as we do not need multiple beliefs on other agents' private information.

Di Tillio et al. (2017) consider the problem of revenue maximization in a private value and independent belief auction. The seller commits to a simple mechanism, i.e., an allocation and transfer rule, but informs agents of a set of simple mechanisms. As all the simple mechanisms generate the same expected revenue (imposed by the Consistency condition), agents do not know the exact rule and thus make decisions based on the worst-case scenario. Compared to the standard Bayesian mechanism, their ambiguous approach yields a higher expected revenue.

In the current paper, ambiguity is engineered in a similar way to Di Tillio et al. (2017). However, instead of studying how ambiguous mechanisms improve second-best revenues under independent beliefs, our paper studies when the first-best outcome in surplus extraction or implementation can be achieved without restricting attention to independent beliefs.

As mentioned before, we fix an efficient allocation rule and only allow for ambiguity in transfer rules, but in Di Tillio et al. (2017)'s mechanism both allocation and transfer rules are ambiguous. Our restriction on unambiguous allocation rule is compatible with Di Tillio et al. (2017)'s Consistency condition. In our full surplus extraction problem, each transfer rule gives the revenue-maximizing designer the first-best revenue. In our implementation problem, since the allocation rule is ex-post efficient and each transfer rule is ex-post budget-balanced, each transfer rule leads to the first-best efficiency. Therefore, Consistency is satisfied. The restriction on unambiguous allocation rule is closely related to two facts: (1) we aim to achieve the first-best outcome, and (2) our argument is confined to a finite type space. Allowing for ambiguity in allocation rules may fail full surplus extraction and implementation. To see this, consider a finite-type environment where the total surplus is maximized by a unique allocation rule. In this case, any other allocation rule is inefficient and has a lower surplus level. As the efficient allocation rule must be used in the mechanisms for full surplus extraction and implementation, and as agents know the designer's objective is to maximize revenue or efficiency, any other rule with a lower surplus level is non-credible to the agents. Hence, multiple allocation rules should not be used in an ambiguous mechanism for first-best outcomes.

The essential factor that enables us to achieve the first-best outcome in a finite-type environment is the correlation in agents' beliefs. In fact, we show correlated information is necessary and sufficient for full surplus extraction and implementation of any efficient allocation, under both interdependent value and private value cases. Correlation also results in different constructions of mechanisms between Di Tillio et al. (2017) and the current paper: in our main section (Section 1.4), we only need two transfer rules, while the number of simple mechanisms in their paper depends on the cardinality of the type space.

In Di Tillio et al. (2017)'s optimal mechanism under independent beliefs and finitely many types, ambiguity in allocation rules plays a role. Therefore, they cannot obtain the first-best outcome. In fact, in a screening or an independent private value auction framework, allowing for ambiguous transfers but not ambiguous allocations does not improve the seller's revenue compared to a standard unambiguous mechanism. However, according to Di Tillio et al. (2017)'s Appendix B, their approach works for full surplus extraction with continuous types. This is because there are infinitely many ex-ante efficient allocation rules, or infinitely many allocation rules that are ex-post efficient almost everywhere. Among them, every two rules are the same except in a null set of the type space. In a continuous type space, if an efficiency-maximizing social planner wants to implement an ex-post efficient allocation rule almost everywhere, she can follow the approach of Di Tillio et al. (2017)'s Appendix B as well. Hence, the current paper only focuses on environments with finitely many types.

1.2 Asymmetric Information Environment

The asymmetric information environment is given by $\mathcal{E} = \{I, A, (\Theta_i, u_i)_{i=1}^N, p\}$.

- Let $I = \{1, ..., N\}$ be a finite set of agents. Assume $N \geq 2$.
- Denote the set of **feasible outcomes** by A.
- Let $\theta_i \in \Theta_i$ be agent i's **type**. For simplicity, denote $\times_{i\in I}\Theta_i$ by Θ, $\times_{j\in I,\, j\neq i}\Theta_j$ by Θ_{-i}, and $\times_{k\in I,\, k\neq i,j}\Theta_k$ by Θ_{-i-j}. Let $|\Theta_i|$ represent the cardinality of Θ_i, where we assume $2 \leq |\Theta_i| < \infty$.[6]
- Each agent i has a quasi-linear **utility function** $u_i(a, \theta) + b$, where $a \in A$ is a feasible outcome, $b \in \mathbb{R}$ is a monetary transfer, and $\theta \in \Theta$ is the realized type profile.

[6] The assumption that $|\Theta_i| \geq 2$ for all i is imposed for simplicity of notation. When at least two agents satisfy this cardinality condition, all theorems and propositions of this paper hold. See Appendix A.2 for more details.

- Let p be a probability distribution on Θ, denoting agents' **common prior**. Let $p(\theta_i)$ and $p(\theta_i, \theta_j)$ represent the marginal distribution of p on θ_i and (θ_i, θ_j) respectively. When agent i has type θ_i, her **belief** is derived from Bayesian updating p, i.e., others have type profile $\theta_{-i} \in \Theta_{-i}$ with probability $p_i(\theta_{-i}|\theta_i)$. For agent $j \neq i$ and type θ_j, we let $p_i(\theta_j|\theta_i)$ denote the marginal belief of $p_i(\cdot|\theta_i) \equiv (p_i(\theta_{-i}|\theta_i))_{\theta_{-i}\in\Theta_{-i}}$ on type θ_j.

The structure of the environment $\mathcal{E}$ is assumed to be common knowledge between the mechanism designer and the agents, but every agent's realized type is her private information.

We impose the following assumption throughout the paper unless otherwise specified.

Assumption 1.2.1: *For all $i, j \in I$ with $i \neq j$, and $(\theta_i, \theta_j) \in \Theta_i \times \Theta_j$, assume $p(\theta_i, \theta_j) > 0$.*

An **allocation rule** $q : \Theta \to A$ is a plan to assign a feasible outcome contingent on agents' realized type profile. An allocation rule q is said to be ex-post **efficient** if $\sum_{i\in I} u_i\big(q(\theta), \theta\big) \geq \sum_{i\in I} u_i\big(q'(\theta), \theta\big)$ for all $q' : \Theta \to A$ and $\theta \in \Theta$.

1.3 Mechanism with Ambiguous Transfers

This section formalizes the mechanism adopted in the paper.

Definition 1.3.1: *A **mechanism with ambiguous transfers** is a triplet $\mathcal{M} = (M, \tilde{q}, \tilde{\Phi})$, where $M = \times_{i\in I} M_i$ is the message space, $\tilde{q} : M \to A$ is a message-contingent allocation rule, and $\tilde{\Phi}$ is a set of message-contingent transfer rules with a generic element $\tilde{\phi} : M \to \mathbb{R}^N$. We call the set $\tilde{\Phi}$ **ambiguous transfers**.*

The mechanism works in the following way. The designer first commits to the message-contingent allocation rule $\tilde{q} : M \to A$ and an arbitrary message-contingent transfer rule $\tilde{\phi} \in \tilde{\Phi}$ secretly. Before reporting messages, agents are informed of the message-contingent allocation rule $\tilde{q}$ and ambiguous transfers $\tilde{\Phi}$, but not $\tilde{\phi}$. After agents report their messages, the mechanism designer reveals $\tilde{\phi}$. Then allocations and transfers are made according to the reported messages as well as $\tilde{q}$ and $\tilde{\phi}$.

In this mechanism, agents face both risk and uncertainty. They merely know the distribution of others' private information, which we interpret as the risk. Their limited knowledge of the exact message-contingent transfer rule chosen by the designer leads to a layer of uncertainty. For each message-contingent transfer rule, agents compute their expected payoffs based on beliefs generated by the common prior. As agents only know the set $\tilde{\Phi}$, we follow the spirit of Gilboa and Schmeidler (1989)'s maxmin expected utility (MEU) and assume that agents make decisions based on the worst-case expected payoff.

A **strategy** of agent i is a mapping $\sigma_i : \Theta_i \to M_i$ where M_i is agent i's message space and $M = \times_{i \in I} M_i$. Like most mechanism design works with ambiguity aversion (e.g., Wolitzky (2016), Di Tillio et al. (2017)), we restrict attention to pure strategies.[7] An **equilibrium** of the mechanism $\mathcal{M} = (M, \tilde{q}, \tilde{\Phi})$ is a strategy profile $\sigma = (\sigma_i)_{i \in I}$ such that

$$\inf_{\tilde{\phi} \in \tilde{\Phi}} \sum_{\theta_{-i} \in \Theta_{-i}} [u_i\big(\tilde{q}(\sigma(\theta_i, \theta_{-i})), (\theta_i, \theta_{-i})\big) + \tilde{\phi}_i(\sigma(\theta_i, \theta_{-i}))] p_i(\theta_{-i}|\theta_i)$$

$$\geq \inf_{\tilde{\phi} \in \tilde{\Phi}} \sum_{\theta_{-i} \in \Theta_{-i}} [u_i\big(\tilde{q}(\sigma'_i(\theta_i), \sigma_{-i}(\theta_{-i})), (\theta_i, \theta_{-i})\big) + \tilde{\phi}_i(\sigma'_i(\theta_i), \sigma_{-i}(\theta_{-i}))] p_i(\theta_{-i}|\theta_i)$$

[7] When there is no ambiguity, the restriction is without loss of generality. When there is ambiguity, depending on how the payoff of playing a mixed strategy is formalized, the restriction could be with or without loss of generality. See Wolitzky (2016) for more details.

for all $i \in I$, $\theta_i \in \Theta_i$, and $\sigma_i' : \Theta_i \to M_i$.

This paper studies two related but different objectives. One is full surplus extraction by a revenue-maximizing mechanism designer, and the other is implementation of an efficient allocation rule via an interim individually rational and ex-post budget-balanced mechanism.

A mechanism with ambiguous transfers $\mathcal{M} = (M, \tilde{q}, \tilde{\Phi})$ is said to **extract the full surplus** if there exists an equilibrium σ such that

$$-\sum_{\theta\in\Theta}\sum_{i\in I}\tilde{\phi}_i(\sigma(\theta))p(\theta) = \max_{\hat{q}:\Theta\to A}\sum_{\theta\in\Theta}\sum_{i\in I}u_i\big(\hat{q}(\theta),\theta\big)p(\theta), \forall\tilde{\phi}\in\tilde{\Phi}. \tag{1.1}$$

The requirement that every $\tilde{\phi} \in \tilde{\Phi}$ achieves the same ex-ante revenue follows from Di Tillio et al. (2017)'s Consistency condition. To see this, suppose some $\tilde{\phi}$ achieves a lower ex-ante revenue compared to another element in $\tilde{\Phi}$. As the mechanism designer's objective is to obtain the highest revenue, $\tilde{\phi}$ is non-credible to buyers. Thus, in this case $\tilde{\phi}$ should not be included in $\tilde{\Phi}$.

A mechanism with ambiguous transfers $\mathcal{M} = (M, \tilde{q}, \tilde{\Phi})$ is said to (partially) **implement** the efficient allocation rule $q : \Theta \to A$, if there exists an equilibrium σ such that $\tilde{q}(\sigma(\theta)) = q(\theta)$ for all $\theta \in \Theta$.

If for each agent $i \in I$, we have $M_i = \Theta_i$, i.e., $M = \Theta$, then $\mathcal{M}$ is said to be a **direct mechanism**. We omit the message space Θ in direct mechanisms. A direct mechanism (q, Φ) satisfies interim **incentive compatibility** if $\inf_{\phi\in\Phi}\sum_{\theta_{-i}\in\Theta_{-i}}[u_i\big(q(\theta_i,\theta_{-i}),(\theta_i,\theta_{-i})\big)+\phi_i(\theta_i,\theta_{-i})]p_i(\theta_{-i}|\theta_i) \geq \inf_{\phi\in\Phi}\sum_{\theta_{-i}\in\Theta_{-i}}[u_i\big(q(\theta_i',\theta_{-i}),(\theta_i,\theta_{-i})\big)+\phi_i(\theta_i',\theta_{-i})]p_i(\theta_{-i}|\theta_i)$ for all $i \in I$, $\theta_i, \theta_i' \in \Theta_i$. Lemma 1.3.1 (on revelation principle) implies that it is without loss of generality to focus on incentive compatible direct mechanisms.

Lemma 1.3.1: *Full surplus extraction can be achieved via a mechanism with ambiguous transfers if and only if there is an incentive compatible direct mechanism with ambiguous transfers* (q, Φ) *that extracts the full surplus. An allocation rule* $q' : \Theta \to A$ *is implementable via a mechanism with ambiguous transfers if and only if there exists an incentive compatible direct mechanism with ambiguous transfers* (q', Φ).

Throughout this paper, the outside option x_0 is normalized to give all agents zero payoffs at all type profiles. A direct mechanism with ambiguous transfers (q, Φ) satisfies interim **individual rationality** if

$$\inf_{\phi \in \Phi} \sum_{\theta_{-i} \in \Theta_{-i}} [u_i\big(q(\theta_i, \theta_{-i}), (\theta_i, \theta_{-i})\big) + \phi_i(\theta_i, \theta_{-i})] p_i(\theta_{-i}|\theta_i) \geq 0$$

for all $i \in I$ and $\theta_i \in \Theta_i$. For both full surplus extraction and implementation, we require that the mechanism should be interim individually rational so that agents participate voluntarily.

A direct mechanism with ambiguous transfers (q, Φ) satisfies ex-post **budget balance** if $\sum_{i \in I} \phi_i(\theta) = 0$ for all $\phi \in \Phi$ and $\theta \in \Theta$. To implement an efficient allocation rule q, we also require the mechanism should be ex-post budget-balanced so that outside money is not needed to finance the efficient outcome. Budget balance is not required for the problem of full surplus extraction because the mechanism designer collects the full surplus.

1.4 Necessary and Sufficient Condition

Our key condition, the Beliefs Determine Preferences property, is introduced by Neeman (2004). It requires that an agent with different types should have distinct beliefs.

Definition 1.4.1: *The **Beliefs Determine Preferences** (BDP) property holds for agent i if there does not exist $\bar{\theta}_i, \hat{\theta}_i \in \Theta_i$ with $\bar{\theta}_i \neq \hat{\theta}_i$ such that $p_i(\cdot|\bar{\theta}_i) = p_i(\cdot|\hat{\theta}_i)$.*

The BDP property requires that agents' beliefs should be correlated. Many forms of correlation, including positively or negatively correlated information, can be accommodated.

The following subsections present the necessary and sufficient condition for full surplus extraction and implementation. The BDP property plays a key role in both results.

1.4.1 Full Surplus Extraction

Theorem 1.4.1: *Given a common prior p, full surplus extraction under any profile of utility functions can be achieved via an interim individually rational mechanism with ambiguous transfers if and only if the BDP property holds for all agents.*

In the Appendix, the proof starts with converting the original problem into finding incentive compatible ambiguous transfers such that every interim individual rationality constraint binds.

The necessity part is proved through constructing utility functions such that full surplus extraction cannot be achieved when the BDP property fails for some agent.

We prove the sufficiency part by constructing a mechanism consisting of two transfer rules. Although there are mechanisms with more transfers that extract the full surplus, to be consistent with the spirit of minimal mechanisms of Di Tillio et al. (2017), we only present the one with two rules.

The construction is decomposed into several lemmas, which are useful for both full surplus extraction and implementation. Lemma A.1.1 shows that for each $i \in I$ and $\bar{\theta}_i, \hat{\theta}_i \in \Theta_i$ satisfying $\bar{\theta}_i \neq \hat{\theta}_i$, there exists a budget-balanced transfer rule $\psi^{\bar{\theta}_i\hat{\theta}_i}$ such that (1) it gives every agent zero expected value when all agents truthfully report, and (2) a type-$\bar{\theta}_i$ agent i achieves a negative expected value when she misreports $\hat{\theta}_i$ and all others truthfully report. This step is proven via Fredholm's theorem of the alternative. Note that $\psi^{\bar{\theta}_i\hat{\theta}_i}$ only needs to satisfy one incentive compatibility constraint. Its existence is guaranteed by the BDP property. Lemmas A.1.2 and A.1.3 construct a linear combination of transfer rules $(\psi^{\bar{\theta}_i\hat{\theta}_i})_{i\in I, \bar{\theta}_i, \hat{\theta}_i \in \Theta_i, \bar{\theta}_i \neq \hat{\theta}_i}$, denoted by ψ, such that ψ is ex-post budget-balanced, gives all agents zero expected values when they truthfully report, and gives an agent a non-zero expected value when she is the only misreporter. Pick an ex-post efficient allocation rule q and let $\eta_i(\theta) = -u_i(q(\theta), \theta)$ for all $i \in I$ and $\theta \in \Theta$.

In the end, let the set of ambiguous transfers for agent i be $\Phi_i = \{\eta_i + c\psi_i, \eta_i - c\psi_i\}$. Notice firstly that η_i transfers agent i's entire surplus to the mechanism designer and secondly ψ_i has zero expected value when every agent truthfully reports. Thus, every interim individual rationality constraint binds. In addition, as ψ_i has non-zero expected value whenever i misreports unilaterally, the lower value out of $\eta_i + c\psi_i$ and $\eta_i - c\psi_i$ is negative under a sufficiently large c. Thus, incentive compatibility can be achieved.

We remark that in the construction of ambiguous transfers, budget balance of ψ among all agents is not necessary for full surplus extraction. Alternatively, we could follow Crémer and McLean (1988) and construct N unrelated transfer rules $(\tilde{\psi}_i)_{i\in I}$ that do not necessarily balance the budget, where each $\tilde{\psi}_i$: (1) gives i zero expected value when all agents truthfully report, and (2) gives i non-zero expected value when she misreports unilaterally. However, requiring budget balance of ψ does not impose a stronger restriction on the common prior p. It allows us to use the same lemmas to study both full surplus extraction and implementation. In addition, we achieve ex-post full surplus extraction. Namely, if the mechanism designer wishes to equate the ex-post revenue and ex-post total surplus, our method still works.

In a finite type space with $N \geq 2$ and $|\Theta_i| \geq 2$ for all i, our necessary and sufficient condition holds for almost every common prior $p \in \Delta(\Theta)$.[8]

Under Bayesian mechanisms, full surplus extraction can be guaranteed if and only if the Convex Independence condition, defined below, holds for all agents.

Definition 1.4.2: *The **Convex Independence** condition holds for agent $i \in I$ if for any type $\bar{\theta}_i \in \Theta_i$ and coefficients $(c_{\hat{\theta}_i})_{\hat{\theta}_i \in \Theta_i} \geq \boldsymbol{0}$, $p_i(\cdot|\bar{\theta}_i) \neq \sum_{\hat{\theta}_i \in \Theta_i \setminus \{\bar{\theta}_i\}} c_{\hat{\theta}_i} p_i(\cdot|\hat{\theta}_i)$.*

The necessary and sufficient condition for Bayesian full surplus extraction holds for almost all priors when $N \geq 2$ and $2 \leq |\Theta_i| \leq |\Theta_{-i}|$ for all $i \in I$. However, when $|\Theta_i| > |\Theta_{-i}|$, the Convex Independence condition fails with positive probability. In Example 1.4.1 where $|\Theta_2| = 3 > |\Theta_1| = 2$, the Convex Independence condition fails for agent 2 under every prior. As another instance, if $N = 3$ and $(|\Theta_1|, |\Theta_2|, |\Theta_3|) = (5, 2, 2)$, it is easy to find a non-negligible set of priors under which agent 1's Convex Independence fails.

The BDP property is weaker than Convex Independence in two aspects. Firstly, the BDP property holds generically even if $|\Theta_i| > |\Theta_{-i}|$. Secondly, the BDP property can address some linear cases of correlation that are ruled out by Convex Independence. When the

[8] If agents without private information are included in I (see Appendix A.2), the BDP property holds generically for all agents if there exists $i, j \in I$ with $i \neq j$ such that $|\Theta_i|, |\Theta_j| \geq 2$.

BDP property holds for all agents but the Convex Independence fails for someone, mechanisms with ambiguous transfers can perform strictly better than Bayesian mechanisms in full surplus extraction.

Intuitively, with multiple transfer rules, an agent's worst-case expected payoffs of different misreports are attained by distinct transfers. Compared to Bayesian mechanisms, we do not need one transfer rule to satisfy all incentive compatibility constraints. Hence, the full surplus can be extracted under a weaker condition than Convex Independence.

Example 1.4.1: *This example demonstrates how ambiguous transfers work. Consider a two-agent environment where one agent has three types, and the other has two. In this case, agent* 2*'s Convex Independence condition never holds. Hence, full surplus extraction cannot be guaranteed via a Bayesian mechanism. However, when ambiguous transfers are allowed, full surplus extraction can be guaranteed under almost all common priors.*

For illustration, consider a common prior $p \in \Delta(\Theta)$ *defined below.*

Table 1.4.1: Prior of Example 1.4.1

p	θ_2^1	θ_2^2	θ_2^3
θ_1^1	0.1	0.2	0.2
θ_1^2	0.2	0.1	0.2

The belief of θ_2^3 *is a convex combination of* θ_2^1 *and* θ_2^2*. Therefore, the Convex Independence condition fails for agent* 2*. We briefly sketch Crémer and McLean (1988)'s argument to see why full surplus extraction is impossible via a Bayesian mechanism in an* *auction with private values satisfying* $\theta_2^1 > \theta_2^2 > \theta_2^3 > \theta_1 > 0$ *for all* $\theta_1 \in \Theta_1$*. Suppose by way of contradiction that a transfer rule to agents,* $\phi = (\phi_1, \phi_2) : \Theta \to \mathbb{R}^2$*, extracts the full surplus. To maximize social surplus, the good should always be allocated to agent* 2*. As agent* 2 *obtains zero surplus at every type, incentive compatibility implies:*

$$
\begin{aligned}
&IC(\theta_2^1\theta_2^3) &\qquad& 0 \geq \theta_2^1 + \tfrac{1}{3}\phi_1(\theta_1^1, \theta_2^3) + \tfrac{2}{3}\phi_1(\theta_1^2, \theta_2^3),\\
&IC(\theta_2^2\theta_2^3) &\qquad& 0 \geq \theta_2^2 + \tfrac{2}{3}\phi_1(\theta_1^1, \theta_2^3) + \tfrac{1}{3}\phi_1(\theta_1^2, \theta_2^3).
\end{aligned}
$$

Averaging them yields $0 \geq \frac{1}{2}\theta_2^1 + \frac{1}{2}\theta_2^2 + \frac{1}{2}\phi_1(\theta_1^1, \theta_2^3) + \frac{1}{2}\phi_1(\theta_1^2, \theta_2^3) > \theta_2^3 + \frac{1}{2}\phi_1(\theta_1^1, \theta_2^3) + \frac{1}{2}\phi_1(\theta_1^2, \theta_2^3)$*. This is a contradiction, as type-*θ_2^3 *agent* 2 *should have non-negative payoff. Hence, the standard Bayesian mechanism design approach cannot extract the full surplus.*

Next, we see how ambiguous transfers can help. Let the set of ambiguous transfers be $\Phi = (\phi^1, \phi^2)$*. Transfers* $\phi^1 = (\phi_1^1, \phi_2^1)$ *and* $\phi^2 = (\phi_1^2, \phi_2^2)$ *are defined as follows.*

$$
\phi_i^1(\theta_1, \theta_2) = \begin{cases} c\psi(\theta_1, \theta_2), & \text{if } i = 1, \\ -\theta_2 - c\psi(\theta_1, \theta_2), & \text{if } i = 2, \end{cases} \qquad \phi_i^2(\theta_1, \theta_2) = \begin{cases} -c\psi(\theta_1, \theta_2), & \text{if } i = 1, \\ -\theta_2 + c\psi(\theta_1, \theta_2), & \text{if } i = 2, \end{cases}
$$

where $c \geq 1.5(\theta_2^1 - \theta_2^3)$*, and* $\psi : \Theta \to \mathbb{R}$ *is given below.*

Table 1.4.2: Side Bet of Example 1.4.1

ψ	θ_2^1	θ_2^2	θ_2^3
θ_1^1	-2	-1	2
θ_1^2	1	2	-2

Notice when both agents truthfully report, for each agent i *and type* $\bar{\theta}_i$*,* $\psi(\bar{\theta}_i, \cdot)$

has zero expected value under belief $p_i(\cdot|\bar{\theta}_i)$. However, when she unilaterally misreports $\hat{\theta}_i \neq \bar{\theta}_i$, $\psi(\hat{\theta}_i, \cdot)$ has a non-zero expected value.

Suppose agents truthfully report, both ϕ^1 and ϕ^2 give the mechanism designer the expected social surplus, $0.3\theta_2^1 + 0.3\theta_2^2 + 0.4\theta_2^3$, and both agents obtain zero expected payoffs.

Then we check incentive compatibility. When type-$\bar{\theta}_1$ agent 1 misreports $\hat{\theta}_1 \neq \bar{\theta}_1$, her worst-case expected payoff is $\min\{\pm \sum_{\theta_2 \in \Theta_2} c\psi(\hat{\theta}_1, \theta_2) p_1(\theta_2|\bar{\theta}_1)\} < 0$ and thus misreporting is not profitable. When type-$\bar{\theta}_2$ agent 2 misreports $\hat{\theta}_2 \neq \bar{\theta}_2$, her worst-case expected payoff is $\min\{\bar{\theta}_2 - \hat{\theta}_2 \pm c \sum_{\theta_1 \in \Theta_1} \psi(\theta_1, \hat{\theta}_2) p_2(\theta_1|\bar{\theta}_2)\} < \bar{\theta}_2 - \hat{\theta}_2$. Therefore, any "upward" misreport is not profitable. As $c \geq 1.5(\theta_2^1 - \theta_2^3)$ and $\theta_2^1 > \theta_2^2 > \theta_2^3$, it is easy to verify the three "downward" incentive compatibility constraints:

$$\begin{array}{ll} IC(\theta_2^1\theta_2^2) & 0 \geq \theta_2^1 - \theta_2^2 - c|\frac{1}{3} \times (-1) + \frac{2}{3} \times 2| = \theta_2^1 - \theta_2^2 - c, \\ IC(\theta_2^1\theta_2^3) & 0 \geq \theta_2^1 - \theta_2^3 - c|\frac{1}{3} \times 2 + \frac{2}{3} \times (-2)| = \theta_2^1 - \theta_2^3 - \frac{2}{3}c, \\ IC(\theta_2^2\theta_2^3) & 0 \geq \theta_2^2 - \theta_2^3 - c|\frac{2}{3} \times 2 + \frac{1}{3} \times (-2)| = \theta_2^2 - \theta_2^3 - \frac{2}{3}c. \end{array}$$

Therefore, the full surplus can be extracted via ambiguous transfers.

The BDP property plays an indispensable role in this example. To see this, consider another prior $\tilde{p}$ satisfying $\tilde{p}_2(\cdot|\theta_2^1) = \tilde{p}_2(\cdot|\theta_2^2)$ and suppose by way of contradiction that full surplus extraction is guaranteed by a set of ambiguous transfers $\tilde{\Phi}$. By truthfully revealing (misreporting), every agent should obtain zero (non-positive) expected payoff. In particular, by misreporting θ_2^2, type-θ_2^1 agent 2 has expected payoff of

$$\inf_{\tilde{\phi} \in \tilde{\Phi}} \{\theta_2^1 + \sum_{\theta_1 \in \Theta_1} \tilde{\phi}_1(\theta_1, \theta_2^2) \tilde{p}_2(\theta_1|\theta_2^1)\} = \theta_2^1 + \inf_{\tilde{\phi} \in \tilde{\Phi}} \sum_{\theta_1 \in \Theta_1} \tilde{\phi}_1(\theta_1, \theta_2^2) \tilde{p}_2(\theta_1|\theta_2^1) \leq 0.$$

As $\tilde{p}_2(\cdot|\theta_2^1) = \tilde{p}_2(\cdot|\theta_2^2)$, the above expression, along with $\theta_2^1 > \theta_2^2$, implies that

$$\theta_2^2 + \inf_{\tilde{\phi}\in\tilde{\Phi}} \sum_{\theta_1\in\Theta_1} \tilde{\phi}_1(\theta_1, \theta_2^2)\tilde{p}_2(\theta_1|\theta_2^2) = \inf_{\tilde{\phi}\in\tilde{\Phi}} \{\theta_2^2 + \sum_{\theta_1\in\Theta_1} \tilde{\phi}_1(\theta_1, \theta_2^2)\tilde{p}_2(\theta_1|\theta_2^2)\} < 0,$$

a contradiction with individual rationality of type-θ_2^2 agent-2. Hence, full surplus extraction cannot be guaranteed.

1.4.2 Implementation

Theorem 1.4.2: *Given a common prior p, any ex-post efficient allocation rule under any profile of utility functions is implementable via an interim individually rational and ex-post budget-balanced mechanism with ambiguous transfers if and only if the BDP property holds for all agents.*

When the BDP property fails, we construct utility functions under which an efficient allocation rule is not implementable, which can prove the necessity part of this theorem.

For the sufficiency part, recall that Lemma A.1.1 has constructed budget-balanced transfer rule ψ that gives agents zero expected values when they truthfully report, and gives agent i non-zero expected value when she misreports unilaterally. Pick any ex-post budget-balanced and interim individually rational transfer rule η. Let the set of ambiguous transfers be $\Phi = \{\eta + c\psi, \eta - c\psi\}$. Incentive compatibility can be achieved by choosing a sufficiently large c.

We remark that efficiency of an allocation rule q does not play a role in the proof. In fact, by combining our proof with that of Kosenok and Severinov (2008), Theorem 1.4.2 can be extended to implement any ex-ante **socially rational** allocation rule q, i.e., q satisfying $\sum_{\theta\in\Theta}\sum_{i\in I} u_i\big(q(\theta), \theta\big)p(\theta) \geq 0$. Hence, a large set of inefficient allocations are implementable under the BDP property.

Kosenok and Severinov (2008) prove that the conditions of Convex Independence and Identifiability are necessary and sufficient for implementing any efficient or ex-ante socially rational allocation rules via an interim individually rational and ex-post budget-balanced Bayesian mechanism.

Definition 1.4.3: *The common prior $p(\cdot)$ satisfies the* ***Identifiability*** *condition if for any $\tilde{p}(\cdot) \neq p(\cdot)$, there exists an agent $i \in I$ and her type $\bar{\theta}_i \in \Theta_i$, with $\tilde{p}(\bar{\theta}_i) > 0$, such that for any $(c_{\hat{\theta}_i})_{\hat{\theta}_i \in \Theta_i} \geq \mathbf{0}$, $\tilde{p}_i(\cdot|\bar{\theta}_i) \neq \sum_{\hat{\theta}_i \in \Theta_i} c_{\hat{\theta}_i} p_i(\cdot|\hat{\theta}_i)$.*

The Identifiability condition is generic in a finite type space with $N = 3$ and $|\Theta_i| \geq 3$ for some $i \in I$ or $N > 3$, but it fails with positive probability otherwise. In particular, Kosenok and Severinov (2008) have remarked that only independent beliefs satisfy this condition when $N = 2$, and thus Convex Independence and Identifiability can never hold simultaneously in two-agent settings. In a budget-balanced Bayesian mechanism without the Identifiability condition, some agent i may have the incentive to misreport in a way that makes the truthful report of some $j \neq i$ appear untruthful. This is because by budget balance, i can benefit from the low expected transfer to j, which is the punishment due to j's (seemingly) untruthful report. However, when the set of ambiguous transfers Φ is used, i does not have such an incentive, because it remains ambiguous whether misreport of j would result in a high or low expected transfer to j. Hence, with ambiguous transfers, we can relax the Identifiability condition.

As the BDP property is weaker than the Convex Independence condition, and we do not need the Identifiability condition, our ambiguous transfers require a weaker condition than Bayesian mechanisms. The difference between our condition and that of Kosenok and Severinov (2008) characterizes when ambiguous transfers perform strictly better than Bayesian mechanisms in implementation of all efficient or ex-ante socially rational allocation rules. In particular, as Convex Independence and Identifiability never hold simultaneously in two-agent settings, ambiguous transfers provide a solution to the impossibility of two-agent individually rational, budget-balanced, and efficient mechanism design generically.

Example 1.4.2: *This example demonstrates how ambiguous transfers work for implementation. As in Example 1.4.1, we still consider a two-by-three environment.*

Recall the common prior p *in Example 1.4.1, for which the Identifiability condition fails as well. We follow Kosenok and Severinov (2008)'s approach to construct utility functions under which an efficient allocation rule is not Bayesian implementable. Let the feasible set of alternatives* A *be* $\{x_0, x_1, x_2\}$*. The outcome* x_0 *gives both agents zero payoffs at all type profiles. The payoffs given by* x_1 *and* x_2 *are presented below, where the first component denotes agent* 1*'s payoff and the second denotes* 2*'s. We assume* $0 < a < B$*.*

Table 1.4.3: Feasible Outcomes in Example 1.4.2

x_1	θ_2^1	θ_2^2	θ_2^3
θ_1^1	$a, 0$	a, a	a, a
θ_1^2	$a, 0$	a, a	a, a

x_2	θ_2^1	θ_2^2	θ_2^3
θ_1^1	a, a	$a - 2B, a + B$	$a, 0$
θ_1^2	a, a	$a - 2B, a + B$	$a, 0$

The efficient allocation rule is $q(\theta_1, \theta_2^1) = x_2$ *and* $q(\theta_1, \theta_2^2) = q(\theta_1, \theta_2^3) = x_1$ *for all* $\theta_1 \in \Theta_1$. *To see* q *is not implementable via an interim individually rational and ex-post budget-balanced Bayesian mechanism, we suppose by way of contradiction that there is a transfer rule* $\phi : \Theta \to \mathbb{R}^N$ *implementing* q. *Multiplying* $IC(\theta_1^1\theta_1^2)$, $IC(\theta_1^2\theta_1^1)$, $IC(\theta_2^1\theta_2^2)$, *and* $IC(\theta_2^2\theta_2^1)$ *by* $0.5, 0.5, 0.3$, *and* 0.3 *respectively, summing them, and taking into account ex-post budget balance yield* $0 \geq -a + B$, *a contradiction.*

To see how ambiguous transfers work, let ϕ^1 *and* ϕ^2 *be the two potential transfer rules, where for all* $(\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$,

$$\phi_i^1(\theta_1, \theta_2) = \begin{cases} c\psi(\theta_1, \theta_2), & \textit{if } i = 1, \\ -c\psi(\theta_1, \theta_2), & \textit{if } i = 2, \end{cases} \qquad \phi_i^2(\theta_1, \theta_2) = \begin{cases} -c\psi(\theta_1, \theta_2), & \textit{if } i = 1, \\ c\psi(\theta_1, \theta_2), & \textit{if } i = 2, \end{cases}$$

$c \geq B$, *and* ψ *is defined in Example 1.4.1. Note that both* ϕ^1 *and* ϕ^2 *are ex-post budget-balanced.*

Type-$\bar{\theta}_i$ *agent* i*'s individual rationality holds, because (1)* $u_i(q(\bar{\theta}_i, \theta_{-i}), (\bar{\theta}_i, \theta_{-i})) = a$ *for all* $i \in I$ *and* $(\bar{\theta}_i, \theta_{-i}) \in \Theta$ *and (2)* $\psi_i(\bar{\theta}_i, \cdot)$*'s expected value is* 0 *under belief* $p_i(\cdot|\bar{\theta}_i)$.

Now we demonstrate one incentive compatibility constraint $IC(\theta_2^2\theta_2^1)$. *Such a misreport gives agent* 2 *a worst-case expected payoff of* $a + B - c|\frac{2}{3} \times (-2) + \frac{1}{3} \times (1)| = a + B - c \leq a$. *The other incentive compatibility constraints can be verified similarly. Therefore, the ambiguous transfers implement* q.

Again, the BDP property plays an essential role. To see this, consider a prior $\tilde{p}$ *satisfying* $\tilde{p}_2(\cdot|\theta_2^1) = \tilde{p}_2(\cdot|\theta_2^2)$. *Suppose by way of contradiction that the interim individually rational and ex-post budget-balanced ambiguous transfers* $\tilde{\Phi}$ *implement* q. *Then the* *following inequalities hold:*

$$IC(\theta_2^1\theta_2^2) \quad \inf_{\tilde{\phi}\in\tilde{\Phi}}\{a + \sum_{\theta_1\in\Theta_1} \tilde{\phi}_2(\theta_1,\theta_2^1)\tilde{p}_2(\theta_1|\theta_2^1)\} \geq \inf_{\tilde{\phi}\in\tilde{\Phi}} \sum_{\theta_1\in\Theta_1} \tilde{\phi}_2(\theta_1,\theta_2^2)\tilde{p}_2(\theta_1|\theta_2^1),$$

$$IC(\theta_2^2\theta_2^1) \quad \inf_{\tilde{\phi}\in\tilde{\Phi}}\{a + \sum_{\theta_1\in\Theta_1} \tilde{\phi}_2(\theta_1,\theta_2^2)\tilde{p}_2(\theta_1|\theta_2^2)\} \geq \inf_{\tilde{\phi}\in\tilde{\Phi}}\{a + B + \sum_{\theta_1\in\Theta_1} \tilde{\phi}_2(\theta_1,\theta_2^1)\tilde{p}_2(\theta_1|\theta_2^2)\}.$$

As $\tilde{p}_2(\cdot|\theta_2^1) = \tilde{p}_2(\cdot|\theta_2^2)$, summing the two expressions yields $2a \geq a + B$, a contradiction. Hence, implementation via an interim individually rational and ex-post budget-balanced mechanism with ambiguous transfers cannot be guaranteed.

1.5 Extension

1.5.1 Implementation under Private Value Environments

When proving the necessity part of Theorem 1.4.2, we construct a profile of interdependent value utility functions. Some may wonder if the BDP property is necessary for implementation under private value environments. We will show at least $N-1$ agents satisfying the BDP property is necessary and sufficient for ex-post efficient, interim individually rational, and ex-post budget-balanced implementation under all private value utility functions. We will also demonstrate that the condition is strictly weaker than the one needed for Bayesian implementation under private value environments.

A utility function u_i is said to have **private value** if $u_i\big(a,(\theta_i,\theta_{-i})\big) = u_i\big(a,(\theta_i,\theta'_{-i})\big)$ for all $\theta_i \in \Theta_i$, $\theta_{-i},\theta'_{-i} \in \Theta_{-i}$, and $a \in A$. We denote $u_i\big(a,(\theta_i,\theta_{-i})\big)$ by $u_i(a,\theta_i)$ in this case.

Theorem 1.5.1: *Given a common prior p, any ex-post efficient allocation rule under any profile of private value utility functions is implementable via an interim individually ratio-*

nal and ex-post budget-balanced mechanism with ambiguous transfers if and only if the BDP property holds for at least $N-1$ agents.

We prove the necessity part by construction again, but the utility functions have private values. For the sufficiency part, we first construct transfers such that $N-1$ agents are incentive compatible. Then by allocating all the surplus to the remaining agent and aligning her incentives with the mechanism designer, the agent will also report truthfully in the private value environment, i.e., when all agents have private values.

Recall that efficiency of the allocation rule q does not play any role in Theorem 1.4.2, and that one can implement inefficient but ex-ante socially rational allocation rules if all agents satisfy the BDP property. However, when only $N-1$ agents satisfy the BDP property, efficiency of q plays a role in this proof, where we let the agent whose BDP property fails be a budget breaker. Example A.1.1 in the Appendix illustrates that an inefficient allocation rule may not be implementable if just $N-1$ agents satisfy the BDP property.

Compared to Theorem 1.4.2, Theorem 1.5.1 implies that we only need a weaker condition for implementation if focusing on private value environments. But according to Theorem 1.5.1, even if ambiguous transfers are allowed and we confine our analysis to private value environments, we can always find non-implementable allocations when information is independent across agents.

To compare ambiguous transfers with Bayesian mechanisms, we present the following necessary condition for Bayesian implementation under private value environments.

Proposition 1.5.1: *Given a common prior p, if any ex-post efficient allocation rule q under* *any profile of private value utility functions is implementable via an interim individually rational and ex-post budget-balanced Bayesian mechanism, then the Convex Independence condition holds for at least $N-1$ agents.*

The necessary condition of Proposition 1.5.1 is stronger than the necessary and sufficient one of Theorem 1.5.1. Hence, when the condition of Theorem 1.5.1 holds but the one of Proposition 1.5.1 fails, ambiguous transfers perform strictly better than Bayesian mechanisms in implementation under private value environments.[9]

By strengthening the necessary condition of Proposition 1.5.1 with the Identifiability condition, we can adapt the argument of Kosenok and Severinov (2008) to give a sufficiency result on Bayesian implementation of efficient allocations under private values. Hence, when Identifiability and the condition of Proposition 1.5.1 hold, ambiguous transfers do not perform better than Bayesian mechanisms.

1.5.2 No Common Prior

This subsection adopts Aumann (1976)'s "agree to disagree" framework to study ambiguous transfers. Namely, we relax the assumption that beliefs are generated by a common prior but still assume common knowledge of their structure. See Morris (1995) for a review of the justifications of modeling with and without a common prior.

The common prior assumption is used in proofs of our previous theorems. In fact,

[9] The necessary condition of Proposition 1.5.1 is not sufficient for Bayesian implementation. In fact, when $N = 2$, for each common prior, one can construct an efficient allocation rule under a private value environment that is not implementable via individually rational and budget-balanced Bayesian mechanisms. Hence, ambiguous transfers improve upon Bayesian mechanisms under two-agent private value environments generically.

without a common prior, it is not hard to construct examples where the BDP property is no longer sufficient for implementation under interdependent values.[10] Hence in this section, we provide sufficient conditions under which efficient allocations are implementable via an interim individually rational and ex-post budget-balanced mechanism with ambiguous transfers. We also demonstrate with examples that ambiguous transfers can implement Bayesian non-implementable allocations.

In Bayesian mechanism design literature, Bergemann et al. (2012), Smith (2010), and Börgers et al. (2015) have documented results related to ex-post efficiency maximization under the "agree to disagree" framework. Without requiring individual rationality and budget balance, Bergemann et al. (2012) show that the BDP property is sufficient for Bayesian implementation of efficient allocations, but the current paper requires interim individual rationality and ex-post budget balance. Smith (2010) compares the welfare of two different mechanisms on public good provision, and Börgers et al. (2015) provide a sufficient condition on when agents' interim payoffs can be arbitrarily increased, given there is an incentive compatible mechanism. Different from Smith (2010) and Börgers et al. (2015), the current section provides a general condition on when the first-best efficiency is implementable.

In this subsection, $p_i(\cdot|\theta_i)$ still represents the belief of type-θ_i agent i, although the beliefs are not generated by a common prior, i.e., there does not exist $p \in \Delta(\Theta)$ with

[10] Without a common prior, full surplus extraction can still be guaranteed via ambiguous transfers when the BDP property holds for all agents. However, full surplus extraction does not mean revenue maximization. By utilizing the lack of common prior between the mechanism designer and agents, the mechanism designer can arbitrarily increase ex-ante revenue. Therefore, we do not study this problem in this section.

$p(\theta_i) > 0$ for all $\theta_i \in \Theta_i$ such that every $p_i(\cdot|\theta_i)$ is obtained by Bayesian updating p. We start with replacing Assumption 1.2.1 with the following one throughout this subsection because without a common prior, the notation $p(\theta_i, \theta_j)$ is not well defined.

Assumption 1.5.1: *For each $i, j \in I$, $i \neq j$, and $(\theta_i, \theta_j) \in \Theta_i \times \Theta_j$, assume $p_i(\theta_j|\theta_i) > 0$.*

Given this assumption, when $N \geq 3$, the notation $p_i(\theta_{-i-j}|\theta_i, \theta_j) \equiv \frac{p_i(\theta_j, \theta_{-i-j}|\theta_i)}{p_i(\theta_j|\theta_i)}$ is well-defined.

Below we introduce a condition called the No Common Prior* property, which strengthens the assumption that agents' beliefs are not generated by a common prior. For all $i \neq j$, θ_i, and θ_j, let $p_j(\theta_i, \cdot|\theta_j)$ be the vector $\left(p_j(\theta_i, \theta_{-i-j}|\theta_j)\right)_{\theta_{-i-j} \in \Theta_{-i-j}}$ when $N \geq 3$, and be the number $p_j(\theta_i|\theta_j)$ when $N = 2$.

Definition 1.5.1: *Agent i satisfies the **No Common Prior*** (NCP*) property if there do not exist types $\bar{\theta}_i \neq \hat{\theta}_i$, a prior $\mu \in \Delta(\Theta)$, and constants $\bar{C} > 0$ and $\hat{C} > 1$ such that*

1. $\mu(\theta_j) > 0$ and $\mu(\theta_{-j}|\theta_j) = p_j(\theta_{-j}|\theta_j)$ for all $(j, \theta_j) \neq (i, \hat{\theta}_i)$;

2. $\hat{C} p_i(\theta_j, \cdot|\hat{\theta}_i) = p_i(\theta_j, \cdot|\bar{\theta}_i) + \bar{C}\frac{p_i(\theta_j|\bar{\theta}_i)}{p_j(\bar{\theta}_i|\theta_j)} p_j(\hat{\theta}_i, \cdot|\theta_j)$ for all $j \neq i$ and θ_j.

When there is a common prior over Θ, one can show the NCP* property is equivalent to the BDP property. Without a common prior over Θ, the statement of the NCP* property cannot be simplified. It requires that there should not exist two types $\bar{\theta}_i \neq \hat{\theta}_i$ simultaneously satisfying the following two properties, (1) all beliefs except the one of $\hat{\theta}_i$ come from a common prior, and (2) the beliefs of $\hat{\theta}_i$ and $\bar{\theta}_i$ are correlated with other agents' beliefs in a certain way.

The NCP* property is very weak except in a two-by-two type space. We introduce below a simple sufficient condition called the NCP** property. If this property holds, then the NCP* property is satisfied by all $i \in I$.

Definition 1.5.2: *Given beliefs $\left(p_i(\cdot|\theta_i)\right)_{i \in I, \theta_i \in \Theta_i}$, the NCP** property holds if $N \geq 3$ and there are agents $i \neq j$ and types $\bar{\theta}_i \neq \hat{\theta}_i$, $\bar{\theta}_j \neq \hat{\theta}_j$, such that the probability distributions over Θ_{-i-j} satisfy $p_i(\cdot|\bar{\theta}_i, \bar{\theta}_j) \neq p_j(\cdot|\bar{\theta}_i, \bar{\theta}_j)$ and $p_i(\cdot|\hat{\theta}_i, \hat{\theta}_j) \neq p_j(\cdot|\hat{\theta}_i, \hat{\theta}_j)$.*

The NCP** property says there are at least three agents and the heterogeneity between agents' beliefs is not too weak. There should be two agents whose beliefs towards the rest of the agents are different at two type profiles. Note this property is stated across agents instead of for a particular agent. Since beliefs are not generated by a common prior, the weak heterogeneity requirement is easy to satisfy.

The following theorem provides a sufficient condition for implementation via ambiguous transfers when there is no common prior.

Theorem 1.5.2: *Given beliefs $\left(p_i(\cdot|\theta_i)\right)_{i \in I, \theta_i \in \Theta_i}$ that are not generated by a common prior, if the BDP and NCP* properties hold for all agents, then any ex-post efficient allocation rule under any profile of utility functions is implementable via an interim individually rational and ex-post budget-balanced mechanism with ambiguous transfers.*

Similar to Theorem 1.4.2, efficiency of an allocation rule q does not play a role in this proof. The set of implementable allocation rules is larger. Indeed, given q, if there exists an ex-post budget-balanced transfer rule η such that $\sum_{\theta_{-i} \in \Theta_{-i}} [u_i(q(\theta_i, \theta_{-i}), (\theta_i, \theta_{-i})) +$

$\eta_i(\theta_i, \theta_{-i})]p_i(\theta_{-i}|\theta_i) \geq 0$ for all $i \in I$ and $\theta_i \in \Theta_i$, we can implement q.

We remark that the sufficient condition in Theorem 1.5.2 is not necessary for implementation. However, we can identify a weaker condition that is necessary. One can introduce a new property that is similar to Definition 1.5.1, except for replacing "$\bar{C} > 0$" with "$\bar{C} > 1$". If this weaker property or the BDP property fails for some agent, then we can construct a non-implementable example following the necessity part of Theorem 1.4.2.

The example below shows a case where ambiguous transfers perform better than Bayesian mechanisms.

Example 1.5.1: *Under the following beliefs without a common prior, the efficient allocation rule q is not Bayesian implementable, but it is implementable via ambiguous transfers.*

Table 1.5.1: Beliefs in Example 1.5.1

$p_1(\tilde{\theta}_2\|\tilde{\theta}_1)$	θ_2^1	θ_2^2	θ_2^3
θ_1^1	$\frac{7}{28}$	$\frac{12}{28}$	$\frac{9}{28}$
θ_1^2	$\frac{13}{28}$	$\frac{12}{28}$	$\frac{3}{28}$

$p_2(\tilde{\theta}_1\|\tilde{\theta}_2)$	θ_2^1	θ_2^2	θ_2^3
θ_1^1	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$
θ_1^2	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{3}$

The feasible set of alternatives, payoffs, and the efficient allocation rule are identical to those in Example 1.4.2, except that $0 < 8.5a < B$ is imposed. Suppose by way of contradiction that there exists a Bayesian payment rule from agent 1 to 2, denoted by ϕ, that implements q. By multiplying $IR(\theta_1^1)$, $IR(\theta_1^2)$, $IR(\theta_2^1)$, $IR(\theta_2^2)$, $IR(\theta_2^3)$, $IC(\theta_1^1\theta_1^2)$, $IC(\theta_1^2\theta_1^1)$, $IC(\theta_2^1\theta_2^2)$, $IC(\theta_2^1\theta_2^3)$, $IC(\theta_2^2\theta_2^1)$, and $IC(\theta_2^3\theta_2^2)$ by 7, 7, 3, 8, 3, 3.5, 3.5, 3, 3, 4, and 3, and summing up, we obtain $0 \geq 4B - 34a$, a contradiction. Hence, q is not Bayesian *implementable.*

It is easy to verify that both agents satisfy the BDP and NCP properties. Then by Theorem 1.5.2,* q *is implementable via ambiguous transfers.*

As an illustration, we demonstrate why the NCP property holds for both agents. The second condition in its definition can be changed into* $\hat{C}\frac{p_i(\theta_j|\hat{\theta}_i)}{p_i(\theta_j|\bar{\theta}_i)} = 1 + \bar{C}\frac{p_j(\hat{\theta}_i|\theta_j)}{p_j(\bar{\theta}_i|\theta_j)}$ *for any* $j \neq i$ *and* θ_j*, when beliefs have full support. Consider* $(i, \bar{\theta}_i, \hat{\theta}_i) = (1, \theta_1^1, \theta_1^2)$*, the NCP* property does not hold because there does not exist* $\bar{C} > 0$ *and* $\hat{C} > 1$ *such that* $\hat{C}(\frac{13}{7}, 1, \frac{1}{3}) = (1, 1, 1) + \bar{C}(2, 1, 0.5)$*. A symmetric argument applies to* $(i, \bar{\theta}_i, \hat{\theta}_i) = (1, \theta_1^2, \theta_1^1)$*. Agent* 2 *satisfies the NCP* property because for each pair* $(\bar{\theta}_2, \hat{\theta}_2)$*, the first equation in the NCP* property fails.*

Compared to Theorem 1.5.2, we have the following weaker sufficient condition for implementation of efficient allocations under private values without a common prior. Like Theorem 1.5.1, the efficiency of the allocation rule plays an important role in the proof.

Theorem 1.5.3: *Given beliefs* $\left(p_i(\cdot|\theta_i)\right)_{i \in I, \theta_i \in \Theta_i}$ *that are not generated by a common prior, if there do not exist* $i \neq j$ *such that the BDP property fails for* i *and the NCP* property fails for* j*, then any ex-post efficient allocation rule under any profile of private value utility functions is implementable via an interim individually rational and ex-post budget-balanced mechanism with ambiguous transfers.*

The sufficient conditions of Theorem 1.5.3 are weak. Recall that when the NCP** property holds, the NCP* property is satisfied by all agents, and thus the sufficient conditions hold. When the BDP property holds for all agents, the sufficient conditions hold as well.

We remark that the sufficient condition in Theorem 1.5.3 is not necessary for implementation. However, a weaker condition is necessary. One can introduce a new property that is similar to Definition 1.5.1, except for replacing "$\bar{C} > 0$" with "$\bar{C} > 1$". If this weaker property fails for some agent and the BDP property fails for another agent, then we can construct a non-implementable example.

The example below shows that there are cases when ambiguous transfers perform better than Bayesian mechanisms.

Example 1.5.2: *In this example of bilateral trading, the efficient allocation rule q is not Bayesian implementable, but it is implementable via ambiguous transfers.*

Agent 1 is the buyer and 2 is the seller. Outcomes in $A = \{x_0, x_1\}$ are feasible, where x_0 represents no trade. The payoffs of x_1, trading, for both agents are given below.

Table 1.5.2: Payoffs of Trading of Example 1.5.2

x_1	θ_2^1	θ_2^2
θ_1^1	4, -3.5	4, -0.5
θ_1^2	1, -3.5	1, -0.5

The efficient allocation rule satisfies $q(\theta_1^2, \theta_2^1) = x_0$ and $q(\theta) = x_1$ for all other θ. The beliefs satisfy $p_1(\theta_2^1|\theta_1^1) = 0.3$, $p_1(\theta_2^1|\theta_1^2) = 0.2$, $p_2(\theta_1^1|\theta_2^1) = 0.3$, and $p_2(\theta_1^1|\theta_2^2) = 0.25$, which are not generated by a common prior.

To see q is not Bayesian implementable, suppose by way of contradiction that there *exists an interim individually rational and ex-post budget-balanced Bayesian mechanism that implements* q*. Let* ϕ *denote the payment from agent* 1 *to* 2*. Multiply* $IC(\theta_1^1\theta_1^2)$*,* $IR(\theta_1^2)$*,* $IC(\theta_1^2\theta_1^1)$*,* $IR(\theta_2^1)$*, and* $IC(\theta_2^2\theta_2^1)$ *by* $4, 10, 1, 10,$ *and* 8 *respectively, and then add them up. We obtain* $0 \geq 0.9$*, which is a contradiction. Therefore,* q *is not Bayesian implementable.*

However, as the BDP property holds for both agents, we know from Theorem 1.5.3 that q *is implementable via ambiguous transfers.*

1.5.3 Other Ambiguity Aversion Preferences

To check the robustness of our result, we look at alternative preferences of ambiguity aversion in this subsection. One is the α-maxmin expected utility (α-MEU) as in Ghirardato and Marinacci (2002), and the other is the smooth ambiguity aversion preferences of Klibanoff et al. (2005). Even though these preferences differ from Gilboa and Schmeidler (1989), the mechanism designer can still benefit from agents' ambiguity aversion.

Ghirardato and Marinacci (2002) introduce the α-MEU, which is a generalization of the MEU. Under an environment described in Section 2.2, a type-θ_i agent i with **α-maxmin expected utility** has the following interim utility level from participating and reporting truthfully when Φ is the set of ambiguous transfers:

$$\alpha \inf_{\phi\in\Phi}\{\sum_{\theta_{-i}\in\Theta_{-i}} u_i\big(q(\theta_i,\theta_{-i}),(\theta_i,\theta_{-i})\big)p_i(\theta_{-i}|\theta_i)+\sum_{\theta_{-i}\in\Theta_{-i}}\phi_i(\theta_i,\theta_{-i})p_i(\theta_{-i}|\theta_i)\}$$
$$+(1-\alpha)\sup_{\phi\in\Phi}\{\sum_{\theta_{-i}\in\Theta_{-i}} u_i\big(q(\theta_i,\theta_{-i}),(\theta_i,\theta_{-i})\big)p_i(\theta_{-i}|\theta_i)+\sum_{\theta_{-i}\in\Theta_{-i}}\phi_i(\theta_i,\theta_{-i})p_i(\theta_{-i}|\theta_i)\},$$

where $\alpha \in [0,1]$. An agent is said to be ambiguity-averse if $\alpha > 0.5$. All previous sections adopt the MEU preferences, which correspond to the case $\alpha = 1$.

Under the α-MEU preferences with $\alpha > 0.5$, Theorem 1.4.2, as well as the sufficiency part of Theorems 1.4.1 and 1.5.1, still holds. We can construct ambiguous transfers under α-MEU in the same way as those under MEU except for choosing a potentially different multiplier c.

An agent i with **smooth ambiguity aversion** has a utility function of

$$\int_{\pi\in\Delta(\Phi)} v\bigg(\int_{\phi\in\Phi}\Big(\sum_{\theta_{-i}\in\Theta_{-i}}[u_i\big(q(\theta_i,\theta_{-i}),(\theta_i,\theta_{-i})\big)+\phi_i(\theta_i,\theta_{-i})]p_i(\theta_{-i}|\theta_i)\Big)d\pi\bigg)d\mu,$$

where

- for each distribution $\pi \in \Delta(\Phi)$, $\pi(\phi)$ measures the subjective density that ϕ is the true transfer rule chosen by the mechanism designer;
- for each distribution $\mu \in \Delta(\Delta(\Phi))$, $\mu(\pi)$ measures the subjective density that $\pi \in \Delta(\Phi)$ is the right density function the mechanism designer uses to choose the transfer rule;
- $v : R \rightarrow R$ is a strictly increasing function that characterizes ambiguity attitude, where a strictly concave v implies ambiguity aversion.

To see ambiguous transfers help under smooth ambiguity aversion preferences, we demonstrate with the Example 1.4.2. Let v be a strictly increasing and strictly concave function. Consider the same transfers as ϕ^1 and ϕ^2 except for a potentially different multiplier c. Then it is easy to verify individual rationality and budget balance. A generic element of $\Delta(\Phi)$ is a Bernoulli distribution between ϕ^1 and ϕ^2. Let μ be the uniform distribution over $\Delta(\Phi)$ for example. As an illustration, we check $IC(\theta_2^2\theta_2^1)$. Truth-telling always gives agent 2 an expected utility of

$$\int_0^1 v(\mu a + (1-\mu)a)d\mu = v(a).$$

By misreporting from θ_2^2 to θ_2^1, agent 2 gets an interim utility of

$$\int_0^1 v\big(\mu(a+B+c) + (1-\mu)(a+B-c)\big)d\mu.$$

For v sufficiently concave or c sufficiently large, the above expression has a value no more than $v(a)$, implying that truth-telling is incentive compatible. One can verify other incentive compatibility constraints as well.

1.6 Conclusion

This paper introduces ambiguous transfers to study full surplus extraction and implementation of an efficient allocation rule via an individually rational and budget-balanced mechanism. We show that the BDP property is necessary and sufficient for both problems, which is weaker than the necessary and sufficient condition for full surplus extraction and implementation via Bayesian mechanisms. Hence, ambiguous transfers can go beyond Bayesian mechanisms. In particular, under two-agent settings, the BDP property offers a solution to overcoming the negative results on bilateral trading problems generically.

CHAPTER 2
ROBUST COALITIONAL IMPLEMENTATION

2.1 Introduction

[1]In implementation theory, if a mechanism can be designed such that all its equilibria coincide with an exogenous social choice function, then the function is said to be fully implementable. Under incomplete information, agents' private information is traditionally modeled by a commonly known type space. Full implementation problems studied under this environment are called the interim or Bayesian implementation problems. However, some details of the type space, especially agents' beliefs, may not be available to the mechanism designer in practice. Therefore, motivated by the Wilson doctrine (Wilson (1985)), Bergemann and Morris (2009, 2011) among others, relax the common knowledge assumption, and adopt a belief-free approach to study when and how a social choice function is fully implementable under all type spaces, which is the *robust implementation* problem.

Most of the solution concepts studied under the interim implementation or robust implementation literature have been non-cooperative. For example, Postlewaite and Schmeidler (1986), Palfrey and Srivastava (1987, 1989), and Jackson (1991) adopt the solution concept of Bayesian Nash equilibrium. Bergemann and Morris (2009, 2011), Penta (2015), Müller (2016), and Ollár and Penta (2017) among others study rationalizable implementation or Bayesian Nash implementation under all type spaces. However, in many

[1] This chapter is a joint work with Nicholas C.Yannelis.

aspects of economics, like voting, matching, or network problems, the stability concern has motivated the study of solution concepts that are immune from collusion. When a mechanism designer aims to elicit private information from agents, the need to design mechanisms that are free from collusion, without unwanted equilibria, and robust to agents' belief structures may coexist. For example, the Vickrey auction is a belief-free mechanism that satisfies (individually) ex-post incentive compatibility, but it is not coalitional incentive compatible. Therefore, the mechanism is not stable due to its vulnerability to coalitional manipulations. This issue, along with the multiplicity of equilibria, can explain why the Vickrey auction may not elicit bidders' truthful evaluation in practice.[2] Motivated by this gap in the literature, the current paper introduces coalition patterns into the problem of robust implementation.

Depending on the mechanism designer's knowledge about the coalition pattern of the environment, we study two problems, *robust coalitional implementation* and *robust double implementation*.

When the mechanism designer knows which coalitions can be formed, namely she knows the coalition pattern, we study the problem of robust coalitional implementation. Our solution concept is the interim coalitional equilibrium, which is a refinement of Bayesian Nash equilibrium and is immune to permissible coalitional deviations. When only singleton coalitions are permissible, our implementation concept reduces to the one of robust implementation as an interim Nash equilibrium, which is a standard concept in the interim and robust implementation literature. We call this concept *robust Nash implemen-*

[2] See, e.g., Ausubel et al. (2006) and Rothkopf (2007), for discussions.

tation, in order to highlight the fact that only singleton coalitions are permissible. When all coalitions are permissible, our solution concept becomes the interim strong equilibrium, and our implementation concept is called *robust strong implementation*. With complete information, the strong implementation problem has taken into account all coalitional deviations. See, for example, Maskin (1978), Moulin and Peleg (1982), Dutta and Sen (1991), Pasin (2009), and Korpela (2013). Under an interim implementation setting, Hahn and Yannelis (2001) has studied a related solution concept in exchange economies. In reality, other coalition patterns can also emerge. For example, if only coalitions with cardinality no more than two can be formed, our solution concept is consistent with the spirit of the pair-wise stable Nash equilibrium in the network literature. With complete information, Suh (1996) has studied the full implementation problem under general coalition patterns.

When the mechanism designer does not know which coalitions can be formed, we study the problem of robust double implementation. Namely, only agents themselves know the coalition pattern, but the designer does not have this information. In this case, if the designer wishes to implement a social choice function regardless of the coalition pattern and under all type spaces, the function needs to be *robustly double implementable* as an interim Nash equilibrium and as an interim strong equilibrium. This question adds one more layer of uncertainty to the designer's problem, besides her uncertainty about agents' type space. In a complete information environment, Maskin (1979) and Suh (1997) have considered related problems.

We establish necessary and almost sufficient conditions for robust implementation in each of the above problems. Specifically, robust coalitional incentive compatibility and robust coalitional monotonicity are necessary and almost sufficient for robustly coalitional implementation of a social choice function. Robust double implementation requires a stronger set of incentive compatibility and monotonicity conditions.

Our modeling of robust coalitional implementation provides new insights for social choice functions that are not robustly Nash implementable. The interim coalitional equilibrium is a refinement of the interim Nash equilibrium, but this does not mean our robust coalitional implementation is more demanding than robust Nash implementation. This is because full implementation not only requires the existence of good equilibria that lead to social choice outcomes, but also requires the non-existence of bad equilibria that result in outcomes different from the social choice function. Robust coalitional incentive compatibility can guarantee the existence of good equilibria. It is stronger than ex-post incentive compatibility, and thus for partial implementation, robust coalitional implementation implies robust Nash implementation. However, the condition to dissolve bad equilibria, robust coalitional monotonicity, is not stronger than the robust monotonicity condition. This gives us leeway to implement some social choice functions that are not robustly Nash implementable. Also, there are social choice functions that are robustly Nash implementable but not robustly coalitional implementable. As a result, the mechanism designer may wish to facilitate or ban agents' communication in order to implement goals that can be achieved exclusively under the cooperative or non-cooperative framework. Robust double implementation guarantees robust coalitional implementation under all coalition patterns, which is demanding. This fact implies the importance for the mechanism designer to learn the coalition pattern of the environment.

From a technical point of view, we construct a new mechanism in order to prove the sufficiency of our conditions for robust coalitional or double implementation. Our basic idea follows the canonical method in the implementation literature. To relax the full support assumption that usually plays a role in the Bayesian implementation literature, we also incorporate the lottery construction of Bergemann and Morris (2011). Since their mechanism cannot prevent profitable coalitional deviations, the focus on coalitional deviations requires non-trivial modifications.

Several applications of our results are provided. We study three variants of the public good example of Bergemann and Morris (2009). In the first variant, the efficient social choice function is robustly strong implementable if and only if agents have a common value. According to this example, robust Nash implementation doesn't imply robust strong implementation and vice versa. In the second variant, the mechanism designer knows the special coalition pattern, which is a result of geographic isolation. We provide an example of a robustly coalitional implementation social choice function. In the third variant, the mechanism designer does not know agents' coalition pattern and we provide an example of robustly double implementation social choice function.

The paper proceeds as follows. Section 2.2 presents the primitives of the environment. The concept of full implementation is given in Section 2.3. We provide necessary and almost sufficient conditions on robust coalitional implementation in Section 2.4 and 2.5. In Section 2.6, we study robust double implementation. Section 2.7 provides applications. In Section 2.8, we discuss a few possible extensions and open questions of this paper.

2.2 Asymmetric Information Environment

We first consider an asymmetric information environment without any specification on beliefs, namely a **payoff environment**, given by $\mathcal{E} = \{I, A, (\Theta_i, u_i)_{i=1}^n\}$, where

- $I = \{1, ..., n\}$ is the set of agents;
- A is **the set of feasible outcomes**, i.e., the set of all lotteries over a deterministic feasible outcome set X;
- $\Theta = \Theta_1 \times ... \times \Theta_n$ is the **payoff type set**, and $\theta_i \in \Theta_i$ is agent i's **payoff type**;
- $u_i : X \times \Theta \to \mathbb{R}$, agent i's **utility function**, represents agent i's utility of consuming a pure outcome $a \in X$, when the realized payoff type profile is $\theta \in \Theta$; then extend the domain of u_i to $A \times \Theta$ so that for any $a \in A = \Delta(X)$ with density function $\mu(\cdot)$, $u_i(a, \theta) = \int_{x \in X} u_i(x, \theta)\mu(x)dx$; assume that the utility function is bounded on A.[3]

A **type space** is a collection $\mathcal{T} = (T_i, \hat{\theta}_i, \pi_i)_{i=1}^I$, where

- $t_i \in T_i$ is a **type** of agent i, which represents agent i's private information; the set of all type profiles is denoted by $T = \prod_{i \in I} T_i$ and a generic element is denoted by $t = (t_1, t_2, ..., t_n)$;
- agent i with type t_i has a payoff type $\hat{\theta}_i(t_i)$, which is defined by an onto mapping $\hat{\theta}_i : T_i \to \Theta_i$; let $\hat{\theta} : T \to \Theta$ be the mapping defined by $\hat{\theta}(t) = \big(\hat{\theta}_1(t_1), ..., \hat{\theta}_n(t_n)\big)$ for all $t \in T$;
- agent i with a type t_i has a **belief type** $\pi_i(t_i)$, which is a probability distribution over $T_{-i} = \prod_{j \neq i} T_j$, assigning probability $\pi_i(t_i)[t_{-i}]$ to the event that others have the type

[3] Both the integral form of the utility function and the boundedness assumption are used and explained in the proof of Theorem 2.5.1.

profile $t_{-i} = (t_j)_{j \neq i}$; $\Delta(T_{-i})$ is the set of all probability distributions on T_{-i}.

The literature on interim or robust implementation under general mechanisms usually focuses on finite (e.g., Postlewaite and Schmeidler (1986), Palfrey and Srivastava (1989), Jackson (1991)) or countable (e.g., Bergemann and Morris (2011)) type spaces. Following them, we assume that Θ_i and T_i are countable sets.

Notice that in a type space $\mathcal{T}$, if for all $i \in I$ and $t_i \in T_i$, $\pi_i(t_i)[\cdot]$ has full support over T_{-i}, then the type space is said to be a **full support type space**. If for all $i \in I$, there is a one-to-one mapping between T_i and Θ_i, then the type space is called a **payoff type spaces**.

A **social choice function** $f : \Theta \to A$ is an exogenous allocation rule contingent on agents' payoff types.

2.3 Full Implementation

A **mechanism** is a pair $(M, g) = (\prod_{i \in I} M_i, g)$, where M_i is the set of all messages that agent i can submit, i.e., the **message space** of agent i.

An **outcome function** is a mapping $g : M \to A$, which assigns to each message profile a feasible outcome. Agent i's **strategy** $\sigma_i : T_i \to M_i$ is a private information contingent plan of submitting messages.

A **strategy profile** is given by $\sigma = (\sigma_1, \sigma_2, ..., \sigma_n)$. For simplicity, denote by σ_S the strategy for all agents in $S \subseteq I$ and by σ_{-S} the strategy for all agents not in S.

Full implementation requires that the set of equilibrium outcomes of a mechanism should coincide with the social choice function.

Definition 2.3.1: *Under a type space $\mathcal{T}$, a mechanism (M, g)* ***fully implements*** *a social choice function f if the following two conditions are satisfied:*

1. *there exists an equilibrium $\sigma : T \to M$ of the mechanism (M, g) such that $g\big(\sigma(t)\big) = f(\hat{\theta}(t))$ for all $t \in T$;*
2. *if σ is an equilibrium of the mechanism (M, g), then $g\big(\sigma(t)\big) = f(\hat{\theta}(t))$ for all $t \in T$.*

If the first requirement is satisfied, then the social choice function is said to be **partially implemented** by (M, g).

When the type space is common knowledge among the mechanism designer and the agents, we call the full implementation problem an **interim implementation** problem. In reality, the mechanism designer may not know agents' belief structure. To implement the social choice function regardless of agents' belief structure, the designer can only rely on mechanisms that are belief-free. If there exists a mechanism (M, g) that fully implements f in all type spaces associated with the payoff environment Θ, then the social choice function is said to be **robustly implementable**.

To take into account the stability concern, we allow coalitions to be formed. A **coalition** is a non-empty subset of I. A **coalition pattern**, denoted by $\mathcal{S}$, is a collection of coalitions, representing the set of coalitions that can be formed. In reality, not all coalitions are of interest, e.g., in a marriage question, only coalitions with cardinality of two or less are considered. Also, some coalitions are not permissible due to culture differences, language barriers, or geographic isolation. As a result, we do not necessarily require $\mathcal{S}$ to include all non-empty subsets of I. We assume that all singleton coalitions of I are included in $\mathcal{S}$, i.e., agents can always choose not to communicate with others. Before Section 2.6, we assume the mechanism designer knows the coalition pattern $\mathcal{S}$. In Section 2.6, we relax this assumption.

Prior to defining the notion of an interim coalitional equilibrium, we introduce some notations to describe how agents update their beliefs after knowing other coalition members' private information. Let the symbol $\setminus$ denote the difference between two sets. For each distribution $\pi_i(t_i^*)[\cdot]$ and $S \ni i$, the notation $\pi_i(t_i^*)[t^*_{S\setminus\{i\}}]$ represents the marginal probability that the coalition $S\setminus\{i\}$ has type profile $t^*_{S\setminus\{i\}}$. For $S \subseteq I$, $t_S^* \in T_S$, and $i \in S$, if the marginal probability $\pi_i(t_i^*)[t^*_{S\setminus\{i\}}] > 0$, Bayes' rule can be applied. In this case, we let $\pi_i(t_i^*)[t_{-i}|t^*_{S\setminus\{i\}}]$ be the conditional probability that t_{-i} is the true type profile of agents in $I \setminus \{i\}$, given that $S\setminus\{i\}$ has a type profile $t^*_{S\setminus\{i\}}$. If the marginal probability $\pi_i(t_i^*)[t^*_{S\setminus\{i\}}] = 0$, Bayes' rule cannot be applied. In this case we assume that agent i updates her belief into $\pi_i(t_i^*)[\cdot|t^*_{S\setminus\{i\}}]$, which is an arbitrary but commonly known distribution satisfying $\pi_i(t_i^*)[t^*_{S\setminus\{i\}}] = 0$.[4]

This paper adopts the notion of interim coalitional equilibrium. It is immune to all permissible coalitional deviations.

Definition 2.3.2: *Given a type space $\mathcal{T}$, the strategy profile σ^* is an* ***interim coalitional equilibrium*** *of the mechanism (M, g) if there does not exist $S \in \mathcal{S}$, $t_S^* \in T_S$, and $\sigma'_S : T_S \to M_S$, such that for all $i \in S$,*

$$\sum_{t_{-i}\in T_{-i}} u_i\Big(g\big(\sigma'_S(t_S^*), \sigma^*_{-S}(t_{-S})\big), \hat{\theta}(t_S^*, t_{-S})\Big)\pi_i(t_i^*)[t_{-i}|t^*_{S\setminus\{i\}}]$$

[4] Some other papers impose specific updating rules when Bayes' rule fails, e.g., Penta (2015) and Müller (2016).

$$> \sum_{t_{-i}\in T_{-i}} u_i\Big(g\big(\sigma^*(t^*_S, t_{-S})\big), \hat{\theta}(t^*_S, t_{-S})\Big)\pi_i(t^*_i)[t_{-i}|t^*_{S\setminus\{i\}}].$$

When only singleton coalitions are permissible, Definition 2.3.2 reduces to the **interim Nash equilibrium**.[5] When all coalitions are permissible, i.e., $\mathcal{S} = 2^I\setminus\emptyset$, the interim coalitional equilibrium generalizes the strong equilibrium to incomplete information. We call this notion an **interim strong equilibrium**. It is clear that an interim strong equilibrium is stronger than an interim coalitional equilibrium and the latter is stronger than an interim Nash equilibrium.

Given a coalition pattern, a social choice function f is said to be **robustly coalitional implementable** if there is a mechanism (M, g) that implements f as an interim coalitional equilibrium in all type spaces. Specifically, a social choice function f is said to be **robustly strong implementable** if there is a mechanism (M, g) that implements f as an interim strong equilibrium in all type spaces. When there exists a mechanism (M, g) implementing f as an interim Nash equilibrium under all type spaces, then f is **robustly Nash implementable**, which is a standard robust implementation concept in the literature.

In the above definition of an interim coalitional equilibrium, we assume that the coalition members truthfully pool their information within the coalition. This is not essential to establish a set of necessary and almost sufficient conditions for robust coalitional implementation. In Section 2.8, we provide a discussion on an alternative definition without the information pooling assumption.

[5] It is usually called a Bayesian Nash equilibrium when beliefs are generated by a common prior. But as agents may not have a common prior, we follow the robust implementation literature and call it an interim Nash equilibrium.

2.4 Necessary Conditions

In this section, we introduce conditions that are necessary for robust coalitional implementation. We show that if a social choice function is robustly coalitional implementable, it satisfies robust coalitional incentive compatibility and robust coalitional monotonicity. The (almost) sufficiency of the conditions will be proved in Section 2.5.

2.4.1 Incentive Compatibility

In a type space $\mathcal{T}$, agent i's **deception** of her type is a mapping $\alpha_i : T_i \to T_i$, i.e., under α_i, the type-t_i agent reports $\alpha_i(t_i)$. We denote by α the deception profile $(\alpha_1, ..., \alpha_n)$.

It could be the case that no individual has the incentive to manipulate her private information, but a coalition has such an incentive. To prevent a coalitional deviation from truth-telling, we need the following condition of interim coalitional incentive compatibility.

Definition 2.4.1: *In a type space $\mathcal{T}$, a social choice function f is **interim coalitional incentive compatible** if there is no $S \in \mathcal{S}$, $t_S^* \in T_S$, and $\alpha_S\colon T_S \to T_S$ such that for all $i \in S$,*

$$\sum_{t_{-i}\in T_{-i}} u_i\Big(f\big(\hat{\theta}(\alpha_S(t_S^*), t_{-S})\big), \hat{\theta}(t_S^*, t_{-S})\Big)\pi_i(t_i^*)[t_{-i}|t_{S\setminus\{i\}}^*]$$
$$> \sum_{t_{-i}\in T_{-i}} u_i\Big(f\big(\hat{\theta}(t_S^*, t_{-S})\big), \hat{\theta}(t_S^*, t_{-S})\Big)\pi_i(t_i^*)[t_{-i}|t_{S\setminus\{i\}}^*].$$

Following standard arguments in implementation theory, it is straightforward to see that the above condition is necessary for a social choice function to be implementable as an interim coalitional equilibrium. We thereby omit the proof.

When the type space is not common knowledge, we need the following robust coalitional incentive compatibility condition.

Definition 2.4.2: *A social choice function f is* ***robust coalitional incentive compatible*** *if for all $S \in \mathcal{S}$ and $\theta'_S \neq \theta^*_S$, there exists $i \in S$ such that*

$$u_i\big(f(\theta^*_S, \theta_{-S}), (\theta^*_S, \theta_{-S})\big) \geq u_i\big(f(\theta'_S, \theta_{-S}), (\theta^*_S, \theta_{-S})\big) \textit{ for all } \theta_{-S} \in \Theta_{-S}.$$

In the Appendix, Proposition B.1.1 shows that f is robust coalitional incentive compatible if and only if it is interim coalitional incentive compatible in all type spaces. As interim coalitional incentive compatibility is necessary for interim coalitional implementation, we have the following proposition.

Proposition 2.4.1: *If a social choice function f is robustly coalitional implementable, then f is robust coalitional incentive compatible.*

Robust coalitional incentive compatibility is stronger than ex-post incentive compatibility, which can be obtained by restricting permissible coalitions to be singletons. The more permissible coalitions there are, the stronger the robust coalitional incentive compatibility condition is. This condition can be less demanding is special environments. For example, in a two-agent environment, robust coalitional incentive compatibility can be guaranteed by ex-post incentive compatibility and ex-post weak Pareto efficiency.[6]

[6] A social choice function f is said to be **ex-post weak Pareto efficient** if there does not exist $\theta \in \Theta$ and $a \in A$ such that $u_i(a, \theta) > u_i(f(\theta), \theta)$ for all $i \in I$. It is "weak" in the sense that being dominated requires another feasible allocation to strictly improve the payoff of every agent. This condition is not necessary for robust coalitional implementation. To see this, consider a constant social choice function that is not ex-post weak Pareto efficient. It can be implemented by a constant outcome function.

2.4.2 Monotonocity

For full implementation, we need a version of the monotonicity condition. With complete information, Maskin's monotonicity condition is necessary and almost sufficient for Nash implementation, (see, e.g., Maskin (1999), Saijo (1988)). The same condition is also proved to be necessary by Maskin (1978) for strong implementation. Dutta and Sen (1991), Suh (1996), and Pasin (2009) characterize the necessary and sufficient conditions of strong implementation and all of these characterizations involves Maskin's monotonicity condition or its variant. With incomplete information, Postlewaite and Schmeidler (1986), Palfrey and Srivastava (1989), Jackson (1991), and Bergemann and Morris (2011) adopt related monotonicity concepts, which are part of the necessary and sufficient conditions for interim Nash implementation or robust Nash implementation. Hahn and Yannelis (2001) propose a coalitional Bayesian monotonicity condition, which is necessary and sufficient for coalitional Bayesian Nash implementation (a variant of our interim strong implementation concept). We will discuss our connection with these monotonicity conditions after introducing our conditions.

A deception profile of types $\alpha : T \to T$ is **acceptable**, if $f\big(\hat{\theta}(t)\big) = f\big(\hat{\theta}(\alpha(t))\big)$ for all $t \in T$. If the deception profile is not **acceptable**, there exists $t^* \in T$ such that $f\big(\hat{\theta}(t^*)\big) \neq f\big(\hat{\theta}(\alpha(t^*))\big)$. In this case, α is said to be **unacceptable** at t^*.

For the social choice function f, each $S \in \mathcal{S}$ and $t' \in T$, we denote the **reward set** by $H_S^{f,t'}$, which is the collection of reward functions $h : T \to A$ satisfying the following conditions: for all $\bar{S}$ such that $S \subseteq \bar{S} \in \mathcal{S}$ and $t''_{\bar{S}} \in T_{\bar{S}}$, there exists $i \in \bar{S}$ such that

$$\sum_{t_{-i}\in T_{-i}} u_i\Big(f\big(\hat{\theta}(t''_{\bar{S}}, t_{-\bar{S}})\big), \hat{\theta}(t''_{\bar{S}}, t_{-\bar{S}})\Big)\pi_i(t''_i)[t_{-i}|t''_{\bar{S}\setminus\{i\}}]$$
$$\geq \sum_{t_{-i}\in T_{-i}} u_i\big(h(t'_{\bar{S}}, t_{-\bar{S}}), \hat{\theta}(t''_{\bar{S}}, t_{-\bar{S}})\big)\pi_i(t''_i)[t_{-i}|t''_{\bar{S}\setminus\{i\}}].$$

The reward set for coalition S contains functions that are not profitable for any superset of S, denoted by $\bar{S}$, under unanimous truthful reports.

Definition 2.4.3: *Given a type space $\mathcal{T}$, a social choice funcntion f satisfies the* ***interim coalitional monotonicity*** *condition if whenever α is unacceptable at $t^* \in T$, there exists $S \in \mathcal{S}$ and $h \in H_S^{f,\alpha(t^*)}$ such that for all $i \in S$,*

$$\sum_{t_{-i}\in T_{-i}} u_i\Big(h\big(\alpha(t^*_S, t_{-S})\big), \hat{\theta}(t^*_S, t_{-S})\Big)\pi_i(t^*_i)[t_{-i}|t^*_{S\setminus\{i\}}]$$
$$> \sum_{t_{-i}\in T_{-i}} u_i\Big(f\big(\hat{\theta}\big(\alpha(t^*_S, t_{-S})\big)\big), \hat{\theta}(t^*_S, t_{-S})\Big)\pi_i(t^*_i)[t_{-i}|t^*_{S\setminus\{i\}}].$$

The interim coalitional monotonicity condition conveys the following meaning: if all agents follow an unacceptable deception profile α, then there exists a coalition $S \in \mathcal{S}$, that can propose a reward function and benefit from consuming it rather than the social choice function; but under unanimous truthful report, the coalition does not profit from consuming the reward function compared to the social choice function.

Briefly speaking, in various monotonicity conditions under non-cooperative frameworks, when a deception profile is unacceptable, one agent switches her ranking between two feasible outcomes: one reward outcome and one social choice outcome, under two states. In our interim coalitional monotonicity condition, one coalition switches its ranking rather than one agent.[7] In the literature, Hahn and Yannelis (2001)'s coalitional Bayesian monotonicity condition and Pasin (2009)'s coalitional monotonicity condition have a similar feature. Our condition is different from the one in Hahn and Yannelis (2001), since we assume that agents in the same coalition know each others' reported types. Pasin (2009)'s coalitional monotonicity condition is defined under complete information. In that condition, one coalition switches its ranking between one critical element and one social choice outcome. The critical element is not defined in a way that is directly related to our reward function, but according to a lemma in that paper and Proposition 2.4.2 of the current paper, a critical element is equivalent to a reward function when our environment is reduced to complete information.

For the purpose of robust coalitional implementation, we consider the robust coalitional monotonicity condition. In this paper, we use the symbol $\alpha : T \to T$ to denote a deception profile of types and use the symbol $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_n)$ to denote a deception profile of payoff types. A deception of agent i's payoff type is a set-valued mapping $\boldsymbol{\beta}_i : \Theta_i \to 2^{\Theta_i} \backslash \emptyset$. The deception profile is **acceptable** if for any selection $\beta \in \boldsymbol{\beta}$, $f\big(\beta(\theta)\big) = f(\theta)$ for all $\theta \in \Theta$. Otherwise, there exists a selection $\beta \in \boldsymbol{\beta}$ and a payoff type profile $\theta^* \in \Theta$ such that $\theta' = \beta(\theta^*)$ and $f(\theta') \neq f(\theta^*)$. In this case, we say the deception profile is **unacceptable** at the pair (θ^*, θ').

For each $i \in I$ and $\theta_i' \in \Theta_i$, define $\boldsymbol{\beta}_i^{-1}(\theta_i') = \{\theta_i \in \Theta_i | \theta_i' \in \boldsymbol{\beta}_i(\theta_i)\}$, which is the set of all possible true payoff types of agent i, given she reports θ_i'. For any $S \subseteq I$, denote

[7] Note the coalition's ranking is a partial order since one outcome is better if and only if it is better for every member in the coalition.

$\boldsymbol{\beta}_S(\theta_S) = (\boldsymbol{\beta}_i(\theta_i))_{i\in S}$ and $\boldsymbol{\beta}_S(\Theta_S) = \{\boldsymbol{\beta}_S(\theta_S)|\theta_S \in \Theta_S\}$.

For each $S \in \mathcal{S}$ and $\theta' \in \Theta$, the **robust reward set**, $Y_S^{f,\theta'}$, is the collection of all robust reward functions $y : \Theta \to A$ satisfying the following conditions: for all $\bar{S}$ such that $S \subseteq \bar{S} \in \mathcal{S}$ and $\theta''_{\bar{S}} \in \Theta_{\bar{S}}$, there exists $i \in \bar{S}$ such that

$$u_i\big(f(\theta''_{\bar{S}}, \theta_{-\bar{S}}), (\theta''_{\bar{S}}, \theta_{-\bar{S}})\big) \geq u_i\big(y(\theta'_{\bar{S}}, \theta_{-\bar{S}}), (\theta''_{\bar{S}}, \theta_{-\bar{S}})\big) \forall \theta_{-\bar{S}} \in \Theta_{-\bar{S}}.$$

Definition 2.4.4: *A social choice function f satisfies the **robust coalitional monotonicity** condition if whenever the deception profile $\boldsymbol{\beta}$ is unacceptable at the pair (θ^*, θ'), there exists $S \in \mathcal{S}$ such that for any conjectures and distributions $\big(\theta'^{i}_{-S} \in \boldsymbol{\beta}_{-S}(\Theta_{-S}), \psi_i(\cdot) \in \Delta(\boldsymbol{\beta}_{-S}(\theta'^{i}_{-S}))\big)_{i\in S}$, there exists $y \in Y_S^{f,\theta'}$ such that for all $i \in S$,*

$$\sum_{\theta_{-S}\in\boldsymbol{\beta}_{-S}(\theta'^{i}_{-S})} u_i\big(y(\theta'_S, \theta'^{i}_{-S}), (\theta^*_S, \theta_{-S})\big)\psi_i(\theta_{-S})$$

$$> \sum_{\theta_{-S}\in\boldsymbol{\beta}_{-S}(\theta'^{i}_{-S})} u_i\big(f(\theta'_S, \theta'^{i}_{-S}), (\theta^*_S, \theta_{-S})\big)\psi_i(\theta_{-S}).$$

Notice that when $S = I$, in the definition of the robust reward set or the robust coalitional monotonicity condition, we slightly abuse the notation by ignoring all quantifiers θ_{-S} and θ'^{i}_{-S} as well as the weighted sum operation.

The following proposition proves that the robust coalitional monotonicity condition is necessary for robust coalitional implementation, by showing the necessity of interim coalitional monotonicity for interim coalitional implementation.

Proposition 2.4.2: *If a social choice function f is robustly coalitional implementable, then f satisfies the robust coalitional monotonicity condition.*

Proof. Suppose a social choice function f is robustly coalitional implemented by (M, g), but the robust coalitional monotonicity condition fails. From Proposition B.1.2, we know there exists a type space $\mathcal{T}$ in which the interim coalitional monotonicity condition fails.

Suppose in $\mathcal{T}$, there exists $\alpha : T \to T$ and $t^* \in T$ such that $f(\hat{\theta}(\alpha(t^*))) \neq f(\hat{\theta}(t^*))$. As f is implementable, there exists an interim coalitional equilibrium σ^* such that $g\big(\sigma^*(t)\big) = f\big(\hat{\theta}(t)\big)$ for all $t \in T$. Notice that $\sigma^* \circ \alpha$ is not an interim coalitional equilibrium at t^*. Hence, there exists $S \in \mathcal{S}$ and $\sigma'_S : T_S \to M_S$ such that for all $i \in S$,

$$\sum_{t_{-i}\in T_{-i}} u_i\Big(g\Big(\sigma'_S(t^*_S), \sigma^*_{-S}\big(\alpha_{-S}(t_{-S})\big)\Big), \hat{\theta}(t^*_S, t_{-S})\Big)\pi_i(t^*_i)[t_{-i}|t^*_{S\setminus\{i\}}]$$
$$> \sum_{t_{-i}\in T_{-i}} u_i\Big(g\Big(\sigma^*\big(\alpha(t^*_S, t_{-S})\big)\Big), \hat{\theta}(t^*_S, t_{-S})\Big)\pi_i(t^*_i)[t_{-i}|t^*_{S\setminus\{i\}}]. \quad (2.1)$$

Define $h : T \to A$ by $h(t) = g\big(\sigma'_S(t^*_S), \sigma^*_{-S}(t_{-S})\big)$ for all $t \in T$. Then we have for all $i \in S$,

$$\sum_{t_{-i}\in T_{-i}} u_i\Big(h\big(\alpha(t^*_S, t_{-S})\big), \hat{\theta}(t^*_S, t_{-S})\Big)\pi_i(t^*_i)[t_{-i}|t^*_{S\setminus\{i\}}]$$
$$> \sum_{t_{-i}\in T_{-i}} u_i\Big(f\big(\hat{\theta}\big(\alpha(t^*_S, t_{-S})\big), \hat{\theta}(t^*_S, t_{-S})\Big)\pi_i(t^*_i)[t_{-i}|t^*_{S\setminus\{i\}}]. \quad (2.2)$$

Since σ^* is an interim coalitional equilibrium, for all $\bar{S}$ such that $S \subseteq \bar{S} \in \mathcal{S}$ and $t''_{\bar{S}} \in T_{\bar{S}}$, $\big(\sigma'_S(t^*_S), \sigma^*_{\bar{S}\setminus S}(\alpha_{\bar{S}\setminus S}(t^*_{\bar{S}\setminus S}))\big)$ cannot be a profitable message profile to submit compared to $\sigma^*_{\bar{S}}(t''_{\bar{S}})$. Therefore, there exists an agent $i \in \bar{S}$ such that

$$\sum_{t_{-i}\in T_{-i}} u_i\Big(g\big(\sigma^*(t''_{\bar{S}}, t_{-\bar{S}})\big), \hat{\theta}(t''_{\bar{S}}, t_{-\bar{S}})\Big)\pi_i(t''_i)[t_{-i}|t''_{\bar{S}\setminus\{i\}}]$$
$$\geq \sum_{t_{-i}\in T_{-i}} u_i\Big(g\big(\sigma'_S(t^*_S), \sigma^*_{\bar{S}\setminus S}(\alpha_{\bar{S}\setminus S}(t^*_{\bar{S}\setminus S})), \sigma^*_{-\bar{S}}(t_{-\bar{S}})\big), \hat{\theta}(t''_{\bar{S}}, t_{-\bar{S}})\Big)\pi_i(t''_i)[t_{-i}|t''_{\bar{S}\setminus\{i\}}].$$

(2.3)

This means

$$\sum_{t_{-i}\in T_{-i}} u_i\Big(f\big(\hat{\theta}(t''_{\bar{S}}, t_{-\bar{S}})\big), \hat{\theta}(t''_{\bar{S}}, t_{-\bar{S}})\Big)\pi_i(t''_i)[t_{-i}|t''_{\bar{S}\setminus\{i\}}]$$

$$\geq \sum_{t_{-i}\in T_{-i}} u_i\big(h(\alpha_{\bar{S}}(t^*_{\bar{S}}), t_{-\bar{S}}), \hat{\theta}(t''_{\bar{S}}, t_{-\bar{S}})\big)\pi_i(t''_i)[t_{-i}|t''_{\bar{S}\setminus\{i\}}]. \quad (2.4)$$

Therefore, one has $h \in H_S^{f,\alpha(t^*)}$. This fact as well as expression (2.2) implies that interim coalitional monotonicity holds in $\mathcal{T}$, a contradiction. □

With a general coalition pattern and a general type space, the interim monotonicity condition is not necessary for interim coalitional implementation. With a general coalition pattern, the robust monotonicity condition is not necessary for robust coalitional implementation, either. This can be seen from Section 2.7.1, where we present a robustly coalitional implementable example that does not satisfy the robust monotonicity condition.

2.5 Sufficient Conditions

The sufficient conditions to robustly coalitional implement a social choice function f are usually slightly stronger than the necessary conditions. A condition called the "bad outcome property" is added to prove the sufficiency of the conditions introduced in Section 2.4.

Definition 2.5.1: *A social choice function f satisfies the* ***bad outcome property*** *if there exists $\underline{a} \in A$ and $\delta > 0$ such that $u_i(f(\theta'), \theta) - u_i(\underline{a}, \theta) \geq \delta$ for all $i \in I$ and $\theta, \theta' \in \Theta$.*

This bad outcome property requires the existence of an outcome $\underline{a}$ that is strictly worse than any social choice outcome, whenever agents truthfully report or misreport. It is easy to satisfy in a quasilinear environment and in an economy where all private consumptions are non-negative. For a quasilinear environment, agent's utility function has a non-linear part and a monetary transfer part. Taking a sufficiently large monetary transfer from agents can usually serve as a bad outcome. In an economy, if the social choice function allocates to every agent positive level of consumption, then leaving every agent zero consumption can usually serve as a bad outcome. Section 2.8.1 has discussed the situation when the bad outcome property is not satisfied in an economy. In that case, one can replace the property with some other strengthening.

With the bad outcome, no coalition has the incentive to deviate from a social choice function to achieve this outcome. In addition, we can construct an "interior" lottery with the bad outcome, so that unwanted equilibria can be dissolved by an agent's incentive to move further away from the bad outcome.

Introducing the bad outcome property helps us to focus on the new problems raised by incomplete information and can simplify the treatment of problems that have been discussed by the complete information literature. When our framework is reduced to complete information, our sufficient conditions in Sections 2.5 and 2.6 imply Dutta and Sen (1991)'s Condition γ and Suh (1996, 1997)'s Conditions $\eta(\mathcal{J})$ and η (when their social choice correspondence is singleton-valued), which are characterizations for strong implementation, coalitional implementation, and double implementation. Their conditions rely on the existence of a family of sets that cannot be described explicitly and thus are not easy to check. But with the bad outcome property and the lottery construction, the abstract sets in their conditions can be viewed as lotteries of reward functions, the social choice function, and the bad outcome.

Theorem 2.5.1: *If a social choice function f satisfies robust coalitional incentive compatibility, robust coalitional monotonicity, and the bad outcome property, then it is robustly coalitional implementable.*

Proof. We construct a mechanism to robustly coalitional implement f. In the mechanism, each agent i reports a message $m_i = (m_i^1, m_i^2, m_i^3, m_i^4)$, where $m_i^1 \in \Theta_i$, $m_i^2 \in \mathbb{N}_+$, $m_i^3 \in \mathbb{N}_+$, $m_i^4 \in \{y : \Theta \to A\}$. Denote $m = (m_1, m_2, ..., m_n)$. We partition the message space into M^1, M^2, and M^3 as follows:

$M^1 = \{m | m_i = (\cdot, 0, \cdot, \cdot)\, \forall i \in I\}$,

$M^2(S) = \{m | \exists K_1 > 0 \text{ and } y \in Y_S^{f,m^1} \text{ s.t. } m_i = (\cdot, K_1, \cdot, y)\, \forall i \in S,$

$\qquad m_j = (\cdot, 0, \cdot, \cdot)\, \forall j \notin S\}$,

$M^2 = \bigcup_{S \in \mathcal{S}} M^2(S)$,

$M^3 = M \backslash \{M^1 \cup M^2\}$.

Let $\underline{a}$ be a "bad outcome" and $\delta > 0$ be a number described in the in the bad outcome property. Let $\underline{a}_\epsilon(\cdot)$ be $f(\cdot)$ with probability $\epsilon > 0$ and $\underline{a}$ with probability $1 - \epsilon$, where ϵ is sufficiently small such that $u_i(\underline{a}_\epsilon(\theta'), \theta) = \epsilon u_i(f(\theta'), \theta) + (1 - \epsilon) u_i(\underline{a}, \theta) < u_i(\underline{a}, \theta) + \delta \leq u_i(f(\theta), \theta)$ for all $\theta, \theta' \in \Theta$ and $i \in I$. Note that the " $=$ " relies on the additivity of the utility function, which is a result of its integral form. The " $<$ " relies on the boundedness of u_i. By the bad outcome property, we also have $u_i(\underline{a}, \theta) + \epsilon\delta \leq u_i(\underline{a}_\epsilon(\theta'), \theta)$ for all $\theta, \theta' \in \Theta$ and $i \in I$.

If $m \in M^1$, let the outcome allocation be $g(m) = f(m^1)$.

If $m \in M^2$, there exists $S \in \mathcal{S}$ such that $m \in M^2(S)$. Let $g(m)$ be a lottery $\tilde{y}(m^1)$, which has a realization of $y(m^1)$, with probability $K_1/(K_1+1)$, $\underline{a}_\epsilon(m^1)$ with probability $(1/(nK_1+n))\sum_{i\in I}(m_i^4/(m_i^4+1))$, and $\underline{a}$ with probability $(1/(nK_1+n))\sum_{i\in I}(1/(m_i^4+1))$.

If $m \in M^3$, let $g(m)$ be $\underline{a}_\epsilon(m^1)$ with probability $(1/n)\sum_{i\in I}(m_i^4/(m_i^4+1))$ and $\underline{a}$ with probability $(1/n)\sum_{i\in I}(1/(m_i^4+1))$.

The outcomes in M^2 and M^3 are compound lotteries of $\underline{a}$, $\underline{a}_\epsilon(m^1)$, and $y(m^1)$. The additivity of the utility function implies that the higher weight the lottery puts on $\underline{a}_\epsilon(m^1)$ as opposed to $\underline{a}$, the better the outcome is.

Claim 2.5.1: *For any type space $\mathcal{T}$, $\sigma_i^*(t_i) = (\hat{\theta}_i(t_i), 0, \cdot, \cdot)$ for all $i \in I$ and $t_i \in T_i$ constitutes an interim coalitional equilibrium of* (M, g).

Proof: For notational convenience, for any strategy of agent i, $\sigma_i : T_i \to M_i$, we decompose it into $\sigma_i = (\sigma_i^1, \sigma_i^2, \sigma_i^3, \sigma_i^4)$. We wish to show that for any $S \in \mathcal{S}$, $t_S \in T_S$, and strategy profile σ'_S, σ'_S is not a profitable deviation from σ^*_S.

Consider a deviation of a coalition $S \in \mathcal{S}$. (1) Suppose the coalition S with type $t_S \in T_S$ deviates by submitting another profile of the form $\sigma'_i(t_i^*) = (\cdot, 0, \cdot, \cdot)$ for all $i \in S$, then by robust coalitional incentive compatibility, σ'_S is not profitable. (2) Suppose there exists $\underline{S} \in \mathcal{S}$ satisfying $\underline{S} \subseteq S$, $K_1 > 0$, and $y : \Theta \to A$ such that $\sigma'_i(t_i^*) = (\cdot, K_1, y, \cdot)$ for all $i \in \underline{S}$ and $\sigma'_i(t_i^*) = (\cdot, 0, \cdot, \cdot)$ for all $i \in S \backslash \underline{S}$. This deviation either results in a message in $M^2(\underline{S})$ or M^3. In both cases, this is not a profitable deviation for S. (3) Any other deviation makes the message fall in M^3 for sure, which is not profitable for S.

This completes the proof of the claim.

Claim 2.5.2: *In an arbitrary type space $\mathcal{T}$, if σ is an interim coalitional equilibrium of the mechanism (M, g), then $\sigma(t) \in M^1$ for all $t \in T$.*

Proof: Suppose by way of contradiction that there exists $t \in T$ such that $\sigma(t) \notin M^1$. Below we show that there exists an agent $j \in I$ who is strictly better-off with the strategy σ'_j defined as $\sigma'_j(t_j) = \left(\sigma^1_j(t_j), \sigma^2_j(t_j), 1 + \sigma^3_j(t_j), \sigma^4_j(t_j)\right)$ and $\sigma'_j(t'_j) = \sigma_j(t'_j)$ for $t'_j \neq t_j$. This contradicts the fact that σ is an interim coalitional equilibrium.

Suppose that there exists $t \in T$ such that $\sigma(t) \notin M^1$. This implies that there is an agent $j \in I$ with type t_j such that $\sigma^2_j(t_j) > 0$. If j deviates with strategy σ'_j, for all $t'_{-j} \in T_{-j}$, the message either leads to a strictly better lottery in M^2 or a strictly better lottery in M^3. This contradicts the fact that σ is an interim coalitional equilibrium

Claim 2.5.3: *In an arbitrary type space $\mathcal{T}$, if σ is an interim coalitional equilibrium of (M, g), then $g(\sigma(t)) = f(\hat{\theta}(t))$ for all $t \in T$.*

Proof: From the previous claim, we know $g(\sigma(t)) = f(\sigma^1(t))$ for all $t \in T$. Suppose $g(\sigma(t^*)) \neq f(\hat{\theta}(t^*))$ for some $t^* \in T$.

Define a correspondence $\boldsymbol{\beta}$ by $\boldsymbol{\beta}(\theta) = \cup_{\{t \in T | \hat{\theta}(t) = \theta\}} \sigma^1(t)$ for all $\theta \in \Theta$. From the supposition, $\boldsymbol{\beta}$ is not acceptable. Define $\theta^* = \hat{\theta}(t^*)$. Thus, there exists selection $\beta \in \boldsymbol{\beta}$ such that $\theta' = \beta(\theta^*)$ and $f(\theta') \neq f(\theta^*)$.

By the robust coalitional monotonicity condition, there exists $S \in \mathcal{S}$ such that for any conjectures and distributions $\left(\theta'^i_{-S} \in \boldsymbol{\beta}_{-S}(\Theta_{-S}), \psi_i(\cdot) \in \Delta(\boldsymbol{\beta}_{-S}(\theta'^i_{-S}))\right)_{i \in S}$, there exists $y \in Y_S^{f,\theta'}$ such that for all $i \in S$, inequality in Definition 2.4.4 is satisfied for all $i \in S$.

Pick a large integer $K^* > 0$. Define σ_i'' by $\sigma_i''(t_i) = (\sigma_i^1(t_i), K^*, \cdot, y)$ for $i \in S$ and $t_i = t_i^*$, and define $\sigma_i''(t_i) = \sigma_i(t_i)$ elsewhere. Then for the coalition with type profile t_S^*, they know the new message is in $M^2(S)$. When K^* is sufficiently large, this deviation is strictly profitable for S in all type spaces, a contradiction.

In view of the three claims, we have established that (M, g) robustly coalitional implements f. □

2.6 Robust Double Implementation

The previous sections assume that the designer knows which coalitions can be formed in the environment. However, in reality, the mechanism designer may not have any information on the coalition pattern, except that $\mathcal{S}$ contains all singletons of I.

In this section, we extend the approach of Theorem 2.5.1 to study unknown coalition patterns. From a mechanism designer's perspective, this section addresses another layer of uncertainty, i.e., the uncertainty of coalition patterns, in addition to the uncertainty of the type spaces. If the mechanism designer wishes to guarantee a desirable outcome regardless of the coalition patterns, the following implementation concept can be adopted.

If there exists a mechanism (M, g) that robustly coalitional implements a social choice function f for every coalition pattern $\mathcal{S}$, then f is said to be **robustly doubly implementable**. Similarly, if in a type space $\mathcal{T}$, a mechanism (M, g) implements a social choice function f for all coalition pattern $\mathcal{S}$, then f is said to be **interim doubly implementable**.

The name “double implementation” comes from the fact that there are two extreme cases of $\mathcal{S}$. The minimal $\mathcal{S}$ only contains singleton coalitions of I and the maximal $\mathcal{S}$ equals $2^I\backslash\emptyset$. In view of the inclusion relationship between interim Nash equilibrium, interim coalitional equilibrium, and interim strong equilibrium, a social choice function is interim double implementable if and only if there exists (M, g) that implements f as an interim Nash and strong equilibrium. Similarly, robust double implementation is equivalent to requiring the existence of (M, g) that robustly Nash implements f and robustly strong implements f simultaneously.

Strong versions of incentive compatibility and monotonicity are necessary for robust double implementation of f. The strong version of incentive compatibility is the one in Definition 2.4.2 under the coalition pattern $\mathcal{S} = 2^I\backslash\emptyset$, which we call **robust strong incentive compatibility**. The strong version of monotonicity is the one in Definition 2.4.4 under the coalition pattern $\mathcal{S} = 2^I\backslash\emptyset$ and with the restriction that S is a singleton, which we call **robust strong monotonicity**. These conditions are strong, implying that the knowledge of coalition pattern can help the designer to implement a social choice function.

The necessity of robust strong incentive compatibility is easy to see. Now we follow Proposition 2.4.2 to provide a brief argument on why robust strong monotonicity is necessary. First, define the condition of **interim strong monotonicity**, which is the one in Definition 2.4.3 under the coalition pattern $\mathcal{S} = 2^I\backslash\emptyset$ and with the restriction that S is a singleton coalition. By applying Lemma B.1.1, we can prove the equivalence between robust strong monotonicity and interim strong monotonicity under all type spaces. Therefore, it suffices to the prove the necessity of interim strong monotonicity for interim double implementation.

Suppose in $\mathcal{T}$, there exists $\alpha : T \to T$ such that $f(\hat{\theta}(\alpha(t^*))) \neq f(\hat{\theta}(t^*))$ for some $t^* \in T$. As f is interim double implementable in $\mathcal{T}$, there exists a mechanism (M, g) and an interim strong equilibrium σ^* such that $g\big(\sigma^*(t)\big) = f\big(\hat{\theta}(t)\big)$ for all $t \in T$. The supposition as well as the fact that f is interim double implementable in $\mathcal{T}$ imply that $\sigma^* \circ \alpha$ is not an interim Nash equilibrium at t^*. Hence, there exists $i \in I$ and $\sigma_i' : T_i \to M_i$ such that agent i is better-off by deviating. By defining $h : T \to A$ by $h(t) = g\big(\sigma_i'(t_i^*), \sigma_{-i}^*(t_{-i})\big)$ for all $t \in T$, one can follow the argument of Proposition 2.4.2 to establish the necessity of interim strong monotonicity in $\mathcal{T}$.

Subsequently, we strengthen the necessary conditions with the bad outcome property and provide a sufficiency result on robustly doubly implementing a social choice function.

Theorem 2.6.1: *If a social choice function f satisfies robust strong incentive compatibility, robust strong monotonicity, and the bad outcome property, then it is robustly doubly implementable.*

Proof. The mechanism is the same as the one in Theorem 2.5.1 except that the union is taken over all coalitons $S \subseteq I$ when we define M^2.

Following Claim 2.5.1, one can show that for any $\mathcal{T}$, $\sigma_i^*(t_i) = (\hat{\theta}_i(t_i), 0, \cdot, \cdot)$ for all $i \in I$ and $t_i \in T_i$ constitutes an interim strong equilibrium and thus an interim Nash equilibrium. Following Claim 2.5.2, we know in any $\mathcal{T}$, if σ is an interim Nash equilibrium of the mechanism, then $\sigma(t) \in M^1$ for all $t \in T$. Then we modify Claim 2.5.3 by adopting the robust strong monotonicity, so that the deviating coalition in its proof is a singleton. As a result, unwanted outcomes can be dissolved by an individual deviation. □

2.7 Applications

2.7.1 A Robustly Strong Implementable Public Good Example

We modify the transfer rules of the public good example of Bergemann and Morris (2009). The new social choice function is robustly strong implementable if and only if agents have a common value, in which case the function is not robustly Nash implementable.

Consider an environment with n agents, and each Θ_i is a fine grid on $[0, 1]$. The social planner chooses to provide $x_0 \in [0, K]$ units of public good with a cost function $c(x_0) = {x_0}^2/2$. Agent i's utility function is $u_i(x, \theta) = (\theta_i + \gamma \sum_{j \neq i} \theta_j)x_0 + x_i$, where $x_i \in [-M, M]$ is the monetary transfer and $\gamma \in \mathbb{R}$ is a measure of interdependence of valuation. The K and M can bound agents' utility, but we pick sufficiently large K and M so that the social choice function and a bad outcome are feasible. Let the social choice public good provision level be $f_0(\theta) = \big(1 + \gamma(n-1)\big) \sum_{i=1}^{n} \theta_i$. Let the transfer rule be $f_i(\theta) = -\big(1 + \gamma(n-1)\big)\big(\gamma\theta_i \sum_{j \neq i} \theta_j + \theta_i^2/2 + (\sum_{j \neq i} \theta_j)^2/2\big)$ for all $i \in I$. The social choice function is $f = \big(f_0, (f_i)_{i \in I}\big)$.

We claim that f is robustly strong implementable if and only if agents have a common value, i.e. $\gamma = 1$. In this case, each agent's utility function is aligned with the social planner's surplus function. As a result, an unacceptable deception profile lowers social welfare and thus every agent's welfare. To prove the result, we apply Theorem 2.5.1. As the bad outcome property holds, it suffices to focus on robust coalitional incentive compatibility and robust coalitional monotonicity under coalition pattern $\mathcal{S} = 2^I \backslash \emptyset$.

When $\gamma = 1$, suppose agents within a coalition S with θ_S misreport θ'_S and other agents truthfully report. For any agent $i \in S$, her net benefit from this coalitional deviation is n times $(\sum_{j\in S}\theta'_j + \sum_{j\notin S}\theta_j)\sum_{j\in I}\theta_j - [0.5(\sum_{j\in S}\theta'_j + \sum_{j\notin S}\theta_j)^2 + 0.5(\sum_{j\in I}\theta_j)^2] \leq 0$. Therefore, robust coalitional monotonicity condition holds. If $\gamma < 1$, let $S = \{1,2\}$, $\theta_1 = 0$, $\theta_2 = 1$ and $\theta_j = 0$ for all $j \notin S$. Reporting $\theta'_1 = \theta'_2 = 0.5$ makes both agents in S strictly better off. If $\gamma > 1$, let $S = \{1,2\}$, $\theta_1 = \theta_2 = 0.5$ and $\theta_j = 0$ for all $j \notin S$. Reporting $\theta'_1 = 0$ and $\theta'_2 = 1$ makes both agents in S strictly better off. Hence, f is robust coalitional incentive compatible if and only if $\gamma = 1$.

When $\gamma = 1$, assume that there exists a payoff type profile θ^* and an unacceptable deception profile so that the report θ' satisfies $\sum_{i\in I}\theta^*_i \neq \sum_{i\in I}\theta'_i$. Let $S = I$ and y satisfy $y_0(\theta') = n\sum_{i\in I}\theta^*_i$ and $y_i(\theta') = -n(\sum_{i\in I}\theta^*_i)^2/2$ for all $i \in I$. Then one can verify the robust coalitional monotonicity condition.

Compared to the transfer rule in Bergemann and Morris (2009), we have one extra term, $(\sum_{j\neq i}\theta_j)^2/2$. The extra term does rely on agent i's report, and thus f is still ex-post incentive compatible under all γ. The extra term also does not change the "aggregator function". Thus their "contraction property", which is equivalent to robust monotonicity, still holds if and only if the interdependence of preferences is small ($|\gamma| < 1/(n-1)$). Recall that ex-post incentive compatibility and robust monotonicity are necessary and almost sufficient for robust Nash implementation.

2.7.2 A Robustly Coalitional Implementable Public Good Example

This part provides an example of a robustly coalitional implementable social choice function.

Consider a variant of the previous example. Suppose there are two islands in a country, the east one and the west one. Let $I_E = \{1, ..., n_E\}$ denote all citizens on the east island and $I_W = \{1 + n_E, ..., n_W + n_E\}$ denote all citizens on the west island. Citizens on each island do not communicate with those on the other island, but they communicate frequently with those on the same island. Thus, the coalition pattern is given by $\mathcal{S} = \{S \neq \emptyset : S \subseteq I_E \, or \, S \subseteq I_W\}$.

The cost of building a bridge with quality level x_0 between the two islands is $x_0^2/2$. For each agent $i \in I_E$, i's utility from the bridge and a transfer x_i is $(\sum_{j \in I_E} \theta_j)x_0 + x_i$. Similarly, for agent $i \in I_W$, i's utility is $(\sum_{j \in I_W} \theta_j)x_0 + x_i$. In other words, within each island, citizens have common value. But across islands, agents have independent valuation.

The socially optimal quality level is $f_0(\theta) = n_E \sum_{j \in I_E} \theta_j + n_W \sum_{j \in I_W} \theta_j$. Let the transfer of agent $i \in I_E$ be $f_i(\theta) = -n_E(\sum_{j \in I_E} \theta_j)^2/2$ and that of agent $i \in I_W$ be $f_i(\theta) = -n_W(\sum_{j \in I_W} \theta_j)^2/2$. By Theorem 2.5.1, f is robustly coalitional implementable. We only discuss why robust coalitional incentive compatibility and robust coalitional monotonicity hold below.

It is equivalent to view this problem as having two representative agents E and W. Agent E's payoff type v_E is defined by $v_E = \sum_{j \in I_E} \theta_j$. And v_W is defined similarly. When the representative agents with payoff type profile (v_E, v_W) misreport (v'_E, v'_W), agent E's utility is $(n_E v'_E + n_W v'_W)v_E - n_E v'^2_E/2$, and agent W's utility is $(n_E v'_E + n_W v'_W)v_W -$

$n_W v_W'^2/2$.

As only coalitions within I_E or I_W can be formed and every east (west) islander always has the same utility with the representative agent E (W), it is equivalent to say the two agents E and W play non-cooperatively. Truthfully reporting is a strictly dominant strategy for E and W. Thus, ex-post incentive compatibility and robust monotonicity hold for agents E and W. Accordingly, robust coalitional incentive compatibility and robust coalitional monotonicity hold for agents $I_E \cup I_W$ under the coalition pattern $\mathcal{S}$.

2.7.3 A Robustly Double Implementable Public Good Example

We provide an example of a robustly double implementable social choice function. Consider the same example as in Section 2.7.1 except for the following three modifications (1) $\gamma = 0$, (2) $\Theta_i = \{0, 1\}$ for all $i \in I$, and (3) $f_i(\theta) = -\theta_i^2/2$. Robust double implementation is a strong requirement, and thus we restrict the payoff type set to seek for a positive result.

By Theorem 2.6.1, the social choice function f is robustly double implementable. We only verify robust strong incentive compatibility and robust strong monotonicity below.

To establish robust strong incentive compatibility, suppose the true payoff type profile is θ and each agent i in a coalition S misreports θ'_i. For $i \in S$, the net gain from misreporting is

$$\Big(\big(\sum_{j\in S}\theta'_j + \sum_{j\in S^c}\theta_j\big)\cdot\theta_i - \theta_i'^2/2\Big) - \Big(\big(\sum_{j\in I}\theta_j\big)\cdot\theta_i - \theta_i^2/2\Big) = \theta_i \sum_{j\in S, j\neq i}(\theta'_j - \theta_j) - (\theta'_i - \theta_i)^2/2.$$

If there exists $i \in S$ such that $\theta_i = 0$, then the coalition S cannot be strictly better-off by deviating. If $\theta_j = 1$ for all $j \in S$, as $\Theta_j = \{0, 1\}$, it must be the case that $\sum_{j\in S, j\neq i}(\theta'_j -$

$\theta_j) \leq 0$, which also means that S does not benefit from deviating.

To establish robust strong monotonicity, let $\boldsymbol{\beta}$ be an unacceptable deception profile. Then there exist different payoff type profiles θ^* and θ' such that $\theta' \in \boldsymbol{\beta}(\theta^*)$. Suppose agent i satisfies $\theta_i^* \neq \theta_i'$. If $\theta_i^* = 1 > \theta_i' = 0$, let $\epsilon = 1$. If $\theta_i^* = 0 < \theta_i' = 1$, let $\epsilon = -1$. Then define $y = (y_0, (y_j)_{j \in I}) : \Theta \to A$ by $y(\theta) = f\big(\theta_i' + \epsilon, \theta_{-i}\big)$ for all $\theta \in \Theta$. One can see that

$$\sum_{\theta_{-i} \in \boldsymbol{\beta}_{-i}(\theta'^{i}_{-i})} u_i\big(y(\theta_i', \theta'^{i}_{-i}), (\theta_i^*, \theta_{-i})\big)\psi_i(\theta_{-i}) > \sum_{\theta_{-i} \in \boldsymbol{\beta}_{-i}(\theta'^{i}_{-i})} u_i\big(f(\theta_i', \theta'^{i}_{-i}), (\theta_i^*, \theta_{-i})\big)\psi_i(\theta_{-i})$$

for all $\theta'^{i}_{-i} \in \boldsymbol{\beta}_{-i}(\Theta_{-i})$ and $\psi_i(\cdot) \in \Delta(\boldsymbol{\beta}_{-i}(\theta'^{i}_{-i}))$, since

$$[(\theta_i' + \epsilon + \sum_{j \neq i} \theta_j'^{i})\theta_i^* - 0.5(\theta_i' + \epsilon)^2] - [(\theta_i' + \sum_{j \neq i} \theta_j'^{i})\theta_i^* - 0.5(\theta_i')^2] = \epsilon(\theta_i^* - \theta_i') - 0.5\epsilon^2 > 0.$$

For all $\theta''_S \in \Theta_S$ such that $i \in S$, there exists $j \in S$ such that

$$u_j(f(\theta''_S, \theta_{-S}), (\theta''_S, \theta_{-S})) \geq u_j(y(\theta'_S, \theta_{-S}), (\theta''_S, \theta_{-S})) = u_j(f(\theta_i' + \epsilon, \theta'_{S \setminus \{i\}}, \theta_{-S}), (\theta''_S, \theta_{-S}))$$

for all $\theta_{-S} \in \Theta_{-S}$ because of robust strong incentive compatibility. Thus, one has established the robust strong monotonicity condition.

2.8 Discussion

This paper introduces coalitional structures to study belief-free full implementation. Given a coalition pattern, we have established necessary and almost sufficient conditions for robustly fully implementing a social choice function as an interim coalitional equilibrium. Our modeling provides insights into implementing social choice functions that may not be robustly Nash implementable. When the mechanism designer does not know which coalitions can be formed, we also study robust double implementation. We discuss a few possible extensions and open questions of this paper.

2.8.1 Relaxing the Bad Outcome Property in an Economy

In an economy with at least three agents, the bad outcome property can be relaxed if we look at interim coalitional implementation or interim double implementation over all full-support and finite type spaces. Here, the implementation is no longer "belief-free" because we impose the full support assumption. Consider an economy with L private goods and a total resource $(e^1, ..., e^L) \geq \mathbf{0}$. The set of feasible pure outcomes X is defined as $\{(x_1, ..., x_n)|x_i \in \mathbb{R}_+ \forall i \in I, \sum_{i \in I} x_i^l \leq e^l \forall l = 1, ..., L\}$. For pure outcomes, the utility function $u_i(\cdot, \theta)$ is assumed to be strictly increasing in every dimension of private consumption and independent of other agents' consumption. The "zero outcome" is not a "bad outcome" because the strict inequality may not hold in Definition 2.5.1. We only present a mechanism to address coalitional implementation. It shares some features with, but is different from the one in Theorem 2.5.1.

With the same strategy space, we define

$$M^1 = \{m \notin M^2 | \exists i \in I s.t. m_j = (\cdot, 0, 0, \cdot) \forall j \neq i\},$$

$$M^2(S) = \{m | \exists K_1 > 0, \text{ and } y \in Y_S^{f,m^1} \text{ s.t. } m_i = (\cdot, K_1, 0, y) \forall i \in S,$$

$$m_j = (\cdot, 0, 0, \cdot) \forall j \notin S\},$$

$$M^2 = \bigcup_{S \in \mathcal{S}} M^2(S),$$

$$M^3 = M \backslash (M^1 \cup M^2).$$

If $m \in M^1$, let $g(m)$ be $f(m^1)$.

If $m \in M^2(S)$ for some $S \in \mathcal{S}$, let $g(m)$ be $y(m^1)$.

If $m \in M^3$, let $g(m)$ allocate all resources to the agent with the highest m_i^3.

Ties are broken in favor of the agent with the smaller index. The message sets M^1

and M^2 and the outcomes over them are similar to those in Theorem 2.5.1, except that we do not need lotteries, we allow one agent to deviate in a detectable way in M^1, and we require at least $n-1$ agents to have $m_i^3 = 0$ in $M^1 \cup M^2$. The outcome function over M^3 is essentially a winner-take-all integer game. We highlight the features of this mechanism that allow us to relax the bad outcome property.

It is easy to see that for any $\mathcal{T}$, $\sigma_i^*(t_i) = (\hat{\theta}_i(t_i), 0, 0, \cdot)$ for all $i \in I$ and $t_i \in T_i$ constitutes an interim coalitional equilibrium. Notice that if a non-singleton coalition deviates, at most one agent can be strictly better-off. In other words, to prove the existence of the "good" equilibria, we rely on the unevenness of the outcomes in M^3 so that deviating is never a coalition's common interest. Recall that Theorem 2.5.1 relies on the fact that every agent dislikes a bad outcome.

If there exists $t \in T$ under which the submitted message profile $\sigma(t)$ is different from the above-mentioned one, there must be *some* agent j who does not win all the resources under $\sigma(t)$ and can profit from claiming a strictly larger m_j^3 than $\max_{i \in I, t_i \in T_i} \sigma_i^3(t_i)$, which is well-defined in any finite type space. For example, when $\sigma(t) \in M^2$, there exists S such that $\sigma(t) \in M^2(S)$. When $|S| = 1$ ($|S| > 1$), there must be an agent in S^c (S) who does not win all the resources and can deviate. When the belief is full-supported, such a deviation is profitable for the agent. Recall that Theorem 2.5.1 relies on the "openness" of the unwanted outcomes so that *every* agent wishes to deviate with σ_i' and go further away from the bad outcome.

2.8.2 An Alternative Definition of Interim Coalitional Equilibrium

Definition 2.3.2 assumes that agents within a coalition pool their private information, which helps the coalition obtain higher efficiency. One can also study interim and robust coalitional implementation without the information pooling assumption. However, the efficiency level is reduced compared to Definition 2.3.2 in general. We provide one solution concept without the assumption.

Definition 2.8.1: *In a type space $\mathcal{T}$, the strategy profile σ^* is an* ***interim coalitional equilibrium*** *of the mechanism (M, g) if there does not exist $S \in \mathcal{S}$, $t_S^* \in T_S$, and $m_S \in M_S$, such that for all $i \in S$,*

$$\sum_{t_{-i}\in T_{-i}} u_i\Big(g\big(m_S, \sigma^*_{-S}(t_{-S})\big), \hat{\theta}(t_i^*, t_{-i})\Big)\pi_i(t_i^*)[t_{-i}]$$

$$> \sum_{t_{-i}\in T_{-i}} u_i\Big(g\big(\sigma^*(t_i^*, t_{-i})\big), \hat{\theta}(t_i^*, t_{-i})\Big)\pi_i(t_i^*)[t_{-i}].$$

Namely, under any type profile, there does not exist a coalition who can strictly benefit from committing within the coalition to sending a certain message. By setting S to be singletons, we see that the above definition is a refinement of the interim Nash equilibrium.

The necessary and sufficient conditions under this definition are natural extensions of the one in Sections 2.4 and 2.5. It should be noted that the commitment within the coalition to submitting m_S plays a role in the proof. The commitment is realistic, as the messages submitted to the mechanism designer are verifiable.

CHAPTER 3
FULL IMPLEMENTATION UNDER AMBIGUITY

3.1 Introduction

[1]In implementation theory, a mechanism designer aims to elicit information from agents and realize an exogenous social choice set or function. If a mechanism can be designed such that all its equilibria coincide with the social choice set, then the set is said to be fully implementable. When agents have private information, the subjective expected utility framework has been widely adopted in the literature to model agents' preferences. However, since fifty years ago, we have known from Ellsberg (1961) that the subjective expected utility hypothesis is problematic. To this end, non-expected utility decision theory has been developed. In particular, the seminal work of Gilboa and Schmeidler (1989) proposes the maximin expected utility, which is one of the successful alternatives in describing agents' decision making under ambiguity. With maximin expected utility models, new insights emerge in the mechanism design theory. However, the full implementation problem has not been considered yet under the maximin expected utility.

By assuming that agents are maximin expected utility maximizers, we provide a new framework to study full implementation. The maximin preferences postulate that agents have multiple beliefs and make a decision with the worst-case belief. As special cases, this setup includes both the Bayesian framework, where the multi-prior set is a singleton, and the Wald-type maximin preferences, where agents' decision making is based

[1] This chapter is a joint work with Nicholas C.Yannelis.

on the worst event.

In this paper, we allow coalitions to be formed under different patterns. When only singleton coalitions can be formed, our solution concept is the ambiguous Nash equilibrium, which generalizes the Bayesian Nash equilibrium to maximin preferences. When all coalitions are permissible, our solution concept is the ambiguous strong equilibrium, which is an adaptation of the strong (Nash) equilibrium of Aumann (1959) to incomplete information environments. Other coalition patterns can emerge, too. For example, when only coalitions of cardinality less or equal to two are of interest, our solution concept follows the spirit of the pairwise stable Nash equilibrium in the network literature.

We provide a unified approach to study implementation under different coalitional patterns. The conditions of ambiguous coalitional incentive compatibility, ambiguous coalitional monotonicity, local Pareto efficiency, and closure are necessary and almost sufficient for a social choice set to be implementable under a certain coalition pattern.

Alternatively, the mechanism designer may not know the coalition pattern, and thus may require a social choice set to be implementable under all coalition patterns. This is the so-called double implementation problem. We strengthen ambiguous coalitional incentive compatibility and ambiguous coalitional monotonicity into ambiguous strong incentive compatibility and ambiguous strong monotonicity for double implementation.

The conditions of incentive compatibility and monotonicity are usually relatively demanding or difficult to check. However, under the Wald-type maximin preferences and private value utility functions, we provide weak conditions to guarantee ambiguous strong incentive compatibility and ambiguous strong monotonicity. As applications, we doubly implement the set of all ambiguous Pareto efficient social choice functions, the maximin core, and the maximin value under the Wald-type maximin preferences.

The paper proceeds as follows. Section 3.1.1 reviews the related literature. Section 3.2 presents the primitives of the paper. We provide necessary and almost sufficient conditions on ambiguous coalitional implementation in Section 3.3 and 3.4. Section 3.5 provides conditions for double implementation. Section 3.6 focuses on Wald-type maximin preferences and provides easy conditions for ambiguous coalitional implementation and double implementation. Several applications are provided in this section. Section 3.7 concludes the paper.

3.1.1 Literature Review

Unlike full implementation, partial implementation only requires the existence of a truth-telling equilibrium leading to the social choice outcome. Partial implementation is studied under an incomplete information environment, and the main condition is incentive compatibility. The emerging literature on mechanism design with ambiguity averse agents has been focusing on partial implementation. de Castro and Yannelis (2018) prove that the Wald-type maximin preference is the only preference to guarantee that all Pareto efficient allocations are incentive compatible. de Castro et al. (2017a,b) thus partially implement every Pareto efficient allocation as a maximin equilibrium. Bose and Renou (2014), Wolitzky (2016), Song (2016), and the first chapter of the current thesis also study partial implementation of efficient social choice functions with ambiguity averse agents from different perspectives. Other related papers focus on revenue maximization with ambiguity averse agents, e.g., Bodoh-Creed (2012) and Di Tillio et al. (2017). The current paper studies full implementation with ambiguity averse agents, and thus is different from the above-mentioned papers.

The problem of full implementation has been studied extensively in both complete and incomplete information environments. With complete information, Maskin (1999), Saijo (1988), and Repullo (1987) among others show that a monotonicity condition is necessary and almost sufficient for Nash implementation. With incomplete information, the Bayesian implementation literature has established the necessary and almost sufficient conditions to implement a social choice set as a Bayesian Nash equilibrium, e.g., Postlewaite and Schmeidler (1986), Palfrey and Srivastava (1987, 1989), Jackson (1991). In the current paper, we have incomplete information and adopt the maximin preferences instead of the Bayesian framework, which distinguish the current paper from the canonical full implementation papers.

There are also works studying full implementation with coalitional structures. Under the Bayesian framework, Hahn and Yannelis (2001) obtain necessary and almost sufficient conditions for Bayesian strong implementation in exchange economies. Under complete information, Maskin (1978), Moulin and Peleg (1982), and Dutta and Sen (1991) among others provide necessary and sufficient conditions to implement a social choice correspondence as a strong equilibrium. Suh (1996) and Suh (1997) characterize the conditions for a social choice social choice correspondence to be implementable under a specific coalition pattern or doubly implementable. The incomplete information and the maximin preferences differentiate the current work from these papers.

3.2 Environment

Following the seminal work of Gilboa and Schmeidler (1989), we assume that agents have multiple probability assessments towards others' types and that each agent makes decisions by considering the worst-case belief, i.e., agents' preferences are represented by the **maximin expected utility**. Our asymmetric information environment is given by

$$\mathcal{E} = \{I, A, (T_i, \Pi_i, u_i)_{i=1}^n\},$$

where:

- $I = \{1, ..., n\}$ is the set of agents;
- A is **the set of feasible outcomes**, i.e., the set of all lotteries over a pure outcome set X;
- $t_i \in T_i$ is a **type** of agent i, which is agent i's private information; we focus on the case that each T_i is finite; the set of all type profiles is denoted by $T = \prod_{i \in I} T_i$; for a subset $S \subseteq I$, denote $T_S = \prod_{j \in S} T_j$; for an agent i, denote $T_{-i} = \prod_{j \neq i} T_j$;
- agent i with type t_i has an **ambiguous belief** $\Pi_i(t_i)$, where the function $\Pi_i : T_i \to 2^{\Delta(T_{-i})}$ maps each type of agent i into a non-empty, compact, and convex set, in which each element $\pi_i(t_i) \in \Delta(T_{-i})$ is a probability distribution over T_{-i}, assigning probability $\pi_i(t_i)[t_{-i}]$ to the event that others have type profile t_{-i};
- $u_i : X \times T \to \mathbb{R}$, agent i's **utility function**, represents agent i's utility of consuming a pure outcome $a \in X$, when the realized type profile is $t \in T$; then extend the domain of u_i to $A \times T$ so that for any $a \in A = \Delta(X)$ with density function $\mu(\cdot)$, $u_i(a, t) = \int_{x \in X} u_i(x, t)\mu(x)dx$; assume that the utility function is bounded on A;

when $u_i(a,(t_i,t_{-i})) = u_i(a,(t_i,t'_{-i}))$ for all $a \in A$, $t_i \in T_i$, and $t_{-i}, t'_{-i} \in T_{-i}$, the utility function u_i is said to have **private value**, and in this case, we denote the utility function by $u_i(a,t_i)$ for simplicity.

We assume that the above environment is common knowledge among the mechanism designer and all agents. From Epstein and Wang (1996), the common knowledge assumption of a non-Bayesian type space is well-defined.

Following Jackson (1991), we impose the following assumption on the environment: for each $t \in T$, if there exists $i \in I$ such that $\pi_i(t_i)[t_{-i}] = 0$ for all $\pi_i(t_i) \in \Pi_i(t_i)$, then for all $j \neq i$, $\pi_j(t_j)[t_{-j}] = 0$ for all $\pi_j(t_j) \in \Pi_j(t_j)$. In other words, agents agree on the set of type profiles that occur with zero probability under all beliefs. Then define $T^* = \{t \in T | \forall i \in I, \exists \pi_i(t_i) \in \Pi_i(t_i)\, s.t.\, \pi_i(t_i)[t_{-i}] > 0\}$ to be the set of type profiles that occur with positive probability under at least some belief.

Define the **information set** of type-t_i agent i by $\mathcal{F}_i(t_i) = \{(t_i,t_{-i}) \in T^* | \exists \pi_i(t_i) \in \Pi_i(t_i)\, s.t.\, \pi_i(t_i)[t_{-i}] > 0\}$, which is the set of type profiles that occurs with positive probability under some belief of type t_i. Notice that agent i's information sets under different $t_i \in T_i$ form a partition of T^*. Denote the partition by $\mathcal{F}_i = \{\mathcal{F}_i(t_i)\}_{t_i \in T_i}$. A **coalition** is an non-empty subset $S \subseteq I$. For a coalition $S \subseteq I$, let $\mathcal{F}_S$ denote the **common knowledge** of coalition S, which is the finest partition of T^* that is coarser than $\mathcal{F}_i$ for every $i \in S$. Note that for a singleton coalition $S = \{i\}$, $\mathcal{F}_{\{i\}} = \mathcal{F}_i$ and thus we use the two notations interchangeably.

For illustration, consider $I = \{1,2\}$, $T_1 = \{t_1^1, t_1^2\}$ and $T_2 = \{t_2^1, t_2^2\}$. For both agents, the multi-belief set $\Pi_i(t_i^1)$ is the set of all beliefs over T_{-i}, and the multi-belief set $\Pi_i(t_i^2)$ is the singleton $\{\pi_i(t_i^2)\}$, where $\pi_i(t_i^2)[t_{-i}^1] = 1$. Then we have $\mathcal{F}_i(t_i^1) = \{(t_i^1, t_{-i}^1), (t_i^1, t_{-i}^2)\}$, $\mathcal{F}_i(t_i^2) = \{(t_i^2, t_{-i}^1)\}$, $\mathcal{F}_i = \{\mathcal{F}_i(t_i^1), \mathcal{F}_i(t_i^2)\}$, $T = \{(t_1^1, t_2^1), (t_1^1, t_2^2), (t_1^2, t_2^1), (t_1^2, t_2^2)\}$, and $T^* = \mathcal{F}_I = \{(t_1^1, t_2^1), (t_1^1, t_2^2), (t_1^2, t_2^1)\}$.

A **social choice function** is a mapping $f : T \to A$. For agent $i \in I$ with type $t_i \in T_i$, agent i's interim preferences, or maximin expected utility, of consuming f is defined as

$$\min_{\pi_i(t_i) \in \Pi_i(t_i)} \sum_{t_{-i} \in T_{-i}} u_i\big(f(t), t\big) \pi_i(t_i)[t_{-i}].$$

A **social choice set** is a set of social choice functions.

If each ambiguous belief is a singleton, the interim preferences are consistent with the **Bayesian preferences** of Harsanyi (1967), which have been adopted by Jackson (1991) to study full implementation.

Agent i is said to have the **Wald-type maximin preferences**, or extreme ambiguity aversion, if for all $t_i \in T_i$, $\mathcal{F}_i(t_i) = \{t \in T | t_{-i} \in T_{-i}\}$, and $\Pi_i(t_i) = 2^{\Delta(T_{-i})}$. In this case, $T = T^*$ and

$$\min_{\pi_i(t_i) \in \Pi_i(t_i)} \sum_{t_{-i} \in T_{-i}} u_i\big(f(t), t\big) \pi_i(t_i)[t_{-i}] = \min_{t_{-i} \in T_{-i}} u_i\big(f(t), t\big).$$

This preference has been adopted by de Castro et al. (2017b) and de Castro and Yannelis (2018) to resolve the conflict between efficiency and incentive compatibility.

A **mechanism** is a pair $(M, g) = (\prod_{i \in I} M_i, g)$, where M_i is the set of all messages that agent i can submit to the mechanism designer, i.e., M_i is the **message space** of agent i. When $M = T$, the mechanism is a direct mechanism, but we consider a general message space in order to achieve full implementation. An **outcome function** is a mapping $g : M \to A$, which assigns a feasible allocation to each message profile. Agent i's **strategy** $\sigma_i : T_i \to M_i$ is a private information contingent plan of submitting messages. A **strategy profile** is given by $\sigma = (\sigma_1, \sigma_2, ..., \sigma_n)$. For simplicity, denote by σ_S the strategy profile for all agents in $S \subseteq I$.

A mechanism (M, g) **fully implements** a social choice set F, if the following two conditions are satisfied:

1. for any $f \in F$, there exists an equilibrium $\sigma : T \to M$ of the mechanism (M, g) such that $g\big(\sigma(t)\big) = f(t)$ for all $t \in T^*$;
2. if σ is an equilibrium of the mechanism (M, g), then there exists $f \in F$ such that $g\big(\sigma(t)\big) = f(t)$ for all $t \in T^*$.

If the first requirement is satisfied, then the social choice set F is said to be **partially implemented** by (M, g).

In this paper, we provide a unified treatment for implementation under different coalition patterns. A coalition pattern $\mathcal{S}$ is a family of coalitions, representing all permissible coalitions. We assume that for all $i \in I$, $\{i\} \in \mathcal{S}$. Namely, each singleton coalition is a permissible coalition. In Sections 3.3 and 3.4, we assume that the coalition pattern is common knowledge among the mechanism designer and all agents.

Under coalition pattern $\mathcal{S}$, our solution concept is the ambiguous coalitional equilibrium, which is a generalization of the strong (Nash) equilibrium of Aumann (1959) to incomplete information and alternative coalition patterns. If a strategy profile is an ambiguous coalition equilibrium, there does not exist a coalition and a type profile, such that a deviation is profitable for the coalition.

Definition 3.2.1: *A strategy profile σ^* is an **ambiguous coalitional equilibrium** of the mechanism (M, g), if there does not exist $S \in \mathcal{S}$, $t^* \in T^*$, and strategy profile $\sigma'_S : T_S \to M_S$ such that*

$$\min_{\pi_i(t_i^*) \in \Pi_i(t_i^*)} \sum_{t_{-i} \in T_{-i}} u_i\big(g(\sigma'_S(t_i^*, t_{S\setminus\{i\}}), \sigma^*_{-S}(t_{-S})), (t_i^*, t_{-i})\big)\pi_i(t_i^*)[t_{-i}]$$
$$\geq \min_{\pi_i(t_i^*) \in \Pi_i(t_i^*)} \sum_{t_{-i} \in T_{-i}} u_i\big(g(\sigma^*(t_i^*, t_{-i})), (t_i^*, t_{-i})\big)\pi_i(t_i^*)[t_{-i}]$$

for all $i \in S$ and the strict inequality holds for some $i \in S$.

This solution concept is an adaptation of the strong (Nash) equilibrium of Aumann (1959) and the $\mathcal{J}$ equilibrium of Suh (1996) to incomplete information, and an adaptation of Hahn and Yannelis (2001)'s coalitional Bayesian Nash equilibrium to maximin preferences under a specific coalition pattern.

In this paper, a profitable deviation for the coalition S only needs to strictly improve one member's interim payoff instead of every agent's payoffs. This is different from the above-mentioned papers, but it allows us to naturally connect Pareto efficiency and the ambiguous coalitional equilibrium under coalition pattern $\mathcal{S} = 2^I \setminus \emptyset$. Also, for Definition 3.2.1, we assume that there is no information exchange within the coalition. These two features differentiate this solution concept from the one adopted by the second chapter.

When $\mathcal{S} = \{\{1\}, \{2\}, ..., \{n\}\}$, we call the ambiguous coalitional equilibrium an **ambiguous Nash equilibrium**, which is the generalization of the Bayesian Nash equilibrium to maximin expected utility. The solution concept has been adopted by Bose and

Renou (2014) and Wolitzky (2016) among others to study partial implementation. When $\mathcal{S} = 2^I \backslash \emptyset$, we call the ambiguous coalitional equilibrium an **ambiguous strong equilibrium**, which is immune to all coalitional deviations.

3.3 Necessary Conditions

In this section, we introduce conditions that are necessary for ambiguous coalitional implementation. We show that if a social choice set is implementable, it satisfies ambiguous coalitional incentive compatibility, ambiguous coalitional monotonicity, closure, and local Pareto efficiency.

3.3.1 Incentive Compatibility

A **deception** for agent i is a mapping $\alpha_i : T_i \to T_i$, i.e., under α_i, the type-t_i agent reports $\alpha_i(t_i)$ to the mechanism designer. Specifically, the identity mapping $\alpha_i^* : T_i :\to T_i$ is the truthful report. We denote by α the deception profile $(\alpha_1, \alpha_2, ..., \alpha_n)$ and denote by $\alpha_S : T_S \to T_S$ the deception profile $(\alpha_i : T_i \to T_i)_{i \in S}$. The ambiguous coalitional incentive compatibility condition is presented below.

Definition 3.3.1: *A social choice set F satisfies the* ***ambiguous coalitional incentive compatibility*** *condition if there exists no $f \in F$, $S \in \mathcal{S}$, $t^* \in T^*$, and $\alpha_S \colon T_S \to T_S$ such that*

$$\min_{\pi_i(t_i^*) \in \Pi_i(t_i^*)} \sum_{t_{-i} \in T_{-i}} u_i\Big(f\big(\alpha_S(t_i^*, t_{S\backslash\{i\}}), t_{-S}\big), (t_i^*, t_{-i})\Big) \pi_i(t_i^*)[t_{-i}]$$

$$\geq \min_{\pi_i(t_i^*) \in \Pi_i(t_i^*)} \sum_{t_{-i} \in T_{-i}} u_i\big(f(t_i^*, t_{-i}), (t_i^*, t_{-i})\big) \pi_i(t_i^*)[t_{-i}]$$

for all $i \in S$ and the strict inequality holds for some $i \in S$.

We now prove that this condition is necessary for ambiguous coalitional implementation.

Proposition 3.3.1: *If a social choice set F is implementable as an ambiguous coalitional equilibrium, then it satisfies the ambiguous coalitional incentive compatibility condition.*

Proof. Suppose by way of contradiction that there exists $f \in F$, $t^* \in T^*$, and $\alpha_S \colon T_S \to T_S$ such that

$$\min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(f\big(\alpha_S(t_i^*, t_{S\setminus\{i\}}), t_{-S}\big), (t_i^*, t_{-i})\Big)\pi_i(t_i^*)[t_{-i}]$$
$$\geq \min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\big(f(t_i^*, t_{-i}), (t_i^*, t_{-i})\big)\pi_i(t_i^*)[t_{-i}]$$

for all $i \in S$ and the strict inequality holds for some $i \in S$.

As F is implementable, there exists a mechanism (M, g) that implements F as an ambiguous coalitional equilibrium. As $f \in F$, there exists an ambiguous coalitional equilibrium σ^* of (M, g) such that $g(\sigma^*(t)) = f(t)$ for all $t \in T^*$. Denote $\sigma_S^* \circ \alpha_S^* : T_S \to M_S$ the strategy profile defined by $(\sigma_i^* \circ \alpha_i : T_i \to M_i)_{i\in S}$. Therefore,

$$\min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(g\big(\sigma_S^*(\alpha_S(t_i^*, t_{S\setminus\{i\}})), \sigma_{-S}^*(t_{-S})\big), (t_i^*, t_{-i})\Big)\pi_i(t_i^*)[t_{-i}]$$
$$\geq \min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(g\big(\sigma^*(t_i^*, t_{-i})\big), (t_i^*, t_{-i})\Big)\pi_i(t_i^*)[t_{-i}]$$

for all $i \in S$ and the strict inequality holds for some $i \in S$.

As $\sigma_S^* \circ \alpha_S$ is a profitable strategy profile for coalition S, the above inequality contradicts the supposition that σ^* is an ambiguous coalitional equilibrium. Therefore, we have established the necessity of the ambiguous coalitional incentive compatibility condition. □

3.3.2 Monotonicity

For full implementation, we need a version of the monotonicity condition introduced by Maskin (1999) under complete information. Postlewaite and Schmeidler (1986), Palfrey and Srivastava (1989), Jackson (1991), and Hahn and Yannelis (2001) adopt related concepts under asymmetric information. They prove that a variant of Maskin's monotonicity condition is key for implementation.

We define first the notion of unacceptable deceptions. Given $f \in F$, the deception profile $\alpha : T \to T$ is **acceptable**, if there exists $f' \in F$ with $f'(t) = f\big(\alpha(t)\big)$ for all $t \in T^*$. Otherwise, the deception is **unacceptable**.

For a social choice set F, a social choice function $f \in F$, and coalition S, let the set $H^{f,S}$ be the collection of all social choice functions $h : T \to A$ such that there does not exist a deception profile $\beta_S : T_S \to T_S$, a type profile $t^* \in T^*$ such that

$$\min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(h\big(\beta_S(t_i^*, t_{S\setminus\{i\}}), t_{-S}\big), (t_i^*, t_{-i})\Big)\pi_i(t_i^*)[t_{-i}]$$
$$\geq \min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\big(f(t_i^*, t_{-i}), (t_i^*, t_{-i})\big)\pi_i(t_i^*)[t_{-i}]$$

for all $i \in S$ and the strict inequality holds for some $i \in S$.

For the special case that $S = I$, let the set $H^{f,I}$ be the collection of all social choice functions $h : T \to A$ such that there does not exist a deception profile $\beta : T \to T$, a type profile $t^* \in T$, and a social choice function $f' \in F$, such that

$$\min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(h\big(\beta(t_i^*, t_{-i})\big), (t_i^*, t_{-i})\Big)\pi_i(t_i^*)[t_{-i}]$$

$$\geq \min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\big(f'(t_i^*,t_{-i}),(t_i^*,t_{-i})\big)\pi_i(t_i^*)[t_{-i}]$$

for all $i \in I$ and the strict inequality holds for some $i \in I$.

Given a coalition pattern $\mathcal{S}$, a social choice set F, a social choice function $f \in F$, and a coalition $S \in \mathcal{S}$, define the **reward set** H_S^f in the following way:

$$H_S^f = \begin{cases} \bigcap_{\bar{S}\supseteq S,\bar{S}\in\mathcal{S}} H^{f,\bar{S}} & \text{if } I \notin \mathcal{S}, \\ (\bigcap_{\bar{S}\supseteq S,\bar{S}\in\mathcal{S}} H^{f,\bar{S}}) \cap (\bigcap_{f'\in F} H^{f',I}) & \text{otherwise.} \end{cases}$$

A function in the reward set is called a **reward function**. Each set H_S^f is called a **reward set**. A function in the reward set is called a **reward function**.

The ambiguous coalitional monotonicity condition is defined in the following way.

Definition 3.3.2: *A social choice set F satisfies the **ambiguous coalitional monotonicity** condition, if for any social choice function $f \in F$ and unacceptable deception $\alpha : T \to T$, there exists $S \in \mathcal{S}$, $t^* \in T^*$, and $h \in H_S^f$ such that*

$$\min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(h\big(t_i^*, t_{S\setminus\{i\}}, \alpha_S(t_{-S})\big),(t_i^*,t_{-i})\Big)\pi_i(t_i^*)[t_{-i}]$$
$$\geq \min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(f\big(\alpha(t_i^*,t_{-i})\big),(t_i^*,t_{-i})\Big)\pi_i(t_i^*)[t_{-i}]$$

for all $i \in S$ and the strict inequality holds for some $i \in S$.

Proposition 3.3.2: *If a social choice set F is implementable as an ambiguous coalitional equilibrium, then F satisfies the ambiguous coalitional monotonicity condition.*

Proof. Suppose F is implementable as an ambiguous coalitional equilibrium and the deception profile $\alpha : T \to T$ is unacceptable for $f \in F$, we want to establish the ambiguous coalitional monotonicity condition. As F is implementable, there exists a mechanism

(M, g) and its ambiguous coalitional equilibrium σ^* such that $g\big(\sigma^*(t)\big) = f(t)$ for all $t \in T^*$. Since α is unacceptable for f, $\sigma^* \circ \alpha$ is not an ambiguous coalitional equilibrium. Hence, there exists $S \in \mathcal{S}$, $t^* \in T^*$, and strategy profile $\sigma'_S : T_S \to M_S$ such that

$$\min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(g\big(\sigma'_S(t_i^*, t_{S\setminus\{i\}}), \sigma^*_{-S}(\alpha_{-S}(t_{-S}))\big), (t_i^*, t_{-i})\Big)\pi_i(t_i^*)[t_{-i}]$$
$$\geq \min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(g\big(\sigma^*(\alpha(t_i^*, t_{-i}))\big), (t_i^*, t_{-i})\Big)\pi_i(t_i^*)[t_{-i}]$$

for all $i \in S$ and the strict inequality holds for some $i \in S$.

Define $h : T \to A$ by $h(t) = g\big(\sigma'_S(t_S), \sigma^*_{-S}(t_{-S})\big)$ for all $t \in T$. Then we have

$$\min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(h\big(t_i^*, t_{S\setminus\{i\}}, \alpha_S(t_{-S})\big), (t_i^*, t_{-i})\Big)\pi_i(t_i^*)[t_{-i}]$$
$$\geq \min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(f\big(\alpha(t_i^*, t_{-i})\big), (t_i^*, t_{-i})\Big)\pi_i(t_i^*)[t_{-i}]$$

for all $i \in S$ and the strict inequality holds for some $i \in S$.

Now we need to establish that $h \in H^f_S$.

Since σ^* is an ambiguous coalitional equilibrium, for any deception profile β, coalition $\bar{S} \in \mathcal{S}$ with $S \subseteq \bar{S}$, $(\sigma'_S \circ \beta_S, \sigma^*_{\bar{S}\setminus S} \circ \beta_{\bar{S}\setminus S})$ cannot be a profitable deviation from $\sigma^*_{\bar{S}}$ at any $\tau \in T^*$. Therefore, there does not exist $\tau \in T^*$, such that

$$\min_{\pi_i(\tau_i)\in\Pi_i(\tau_i)} \sum_{t_{-i}\in T_{-i}} u_i\Big(h\big(\beta_{\bar{S}}(\tau_i, t_{\bar{S}\setminus\{i\}}), t_{-\bar{S}}\big), (\tau_i, t_{-i})\Big)\pi_i(\tau_i)[t_{-i}]$$
$$\geq \min_{\pi_i(\tau_i)\in\Pi_i(\tau_i)} \sum_{t_{-i}\in T_{-i}} u_i\big(f(\tau_i, t_{-i}), (\tau_i, t_{-i})\big)\pi_i(\tau_i)[t_{-i}].$$

This has established that $h \in H^{f,\bar{S}}$. When $I \in \mathcal{S}$, for any $f' \in F$, there exists an ambiguous coalitional equilibrium σ^{**} such that $g(\sigma^{**}(t)) = f'(t)$ for all $t \in T^*$. As $(\sigma'_S \circ \beta_S, \sigma^*_{-S} \circ$

β_{-S}) is not a profitable deviation for I, one can apply a similar argument to show that $h \in H^{f',I}$. As a result, we have $h \in H^f_S$.

Therefore, we have established the ambiguous coalitional monotonicity condition.

□

3.3.3 Efficiency

We begin with defining an ambiguous Pareto efficiency condition in the interim stage.

Definition 3.3.3: *A social choice function f is said to satisfy the **ambiguous Pareto efficiency** condition if there does not exist another social choice function $y : T \to A$ and a type profile $t^* \in T^*$ such that*

$$\min_{\pi_i(t_i^*) \in \Pi_i(t_i^*)} \sum_{t_{-i} \in T_{-i}} u_i(y(t_i^*, t_{-i}), (t_i^*, t_{-i}))\pi_i(t_i^*)[t_{-i}]$$
$$\geq \min_{\pi_i(t_i^*) \in \Pi_i(t_i^*)} \sum_{t_{-i} \in T_{-i}} u_i(f(t_i^*, t_{-i}), (t_i^*, t_{-i}))\pi_i(t_i^*)[t_{-i}]$$

for all $i \in I$ and the strict inequality holds for some $i \in I$. A social choice set F is said to satisfy the ambiguous Pareto efficiency condition if every social choice function $f \in F$ satisfies the ambiguous Pareto efficiency condition.

This condition is not necessary for implementation in general. For example, consider a constant social choice function that is not ambiguous Pareto efficient. The function is implementable by a mechanism with a constant outcome function.

However, when $I \in \mathcal{S}$, a "local" efficiency condition is necessary to prevent deviation of the grand coalition. The condition is "local" in the sense that any function in the social choice set cannot be dominated by another function in the social choice set.

Definition 3.3.4: *A social choice set F is said to satisfy the* ***local Pareto efficiency*** *condition if there does not exist $f, f' \in F$, $t^* \in T^*$, and a deception profile $\alpha : T \to T$ such that*

$$\min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i(f'(\alpha(t_i^*, t_{-i})), (t_i^*, t_{-i}))\pi_i(t_i^*)[t_{-i}]$$
$$\geq \min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i(f(t_i^*, t_{-i}), (t_i^*, t_{-i}))\pi_i(t_i^*)[t_{-i}]$$

for all $i \in I$ and the strict inequality holds for some $i \in I$.

It is easy to see that if a social choice set is ambiguous Pareto efficient, then it satisfies the local Pareto efficiency condition.

Proposition 3.3.3: *When $I \in \mathcal{S}$, if a social choice set F is implementable as an ambiguous coalitional equilibrium, then F satisfies the local Pareto efficiency condition.*

Proof. We prove by way of contradiction. Suppose F is implementable, but there exists $f, f' \in F$, $t^* \in T^*$, and a deception profile $\alpha : T \to T$ such that

$$\min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i(f'(\alpha(t_i^*, t_{-i})), (t_i^*, t_{-i}))\pi_i(t_i^*)[t_{-i}]$$
$$\geq \min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i(f(t_i^*, t_{-i}), (t_i^*, t_{-i}))\pi_i(t_i^*)[t_{-i}]$$

for all $i \in I$ and the strict inequality holds for some $i \in I$.

As F is implementable, there exists a mechanism (M, g) and ambiguous coalitional equilibria σ and σ' such that $g\big(\sigma(t)\big) = f(t)$ and $g\big(\sigma'(t)\big) = f'(t)$ for all $t \in T$. Then we have $\sigma' \circ \alpha$ is a profitable grand coalitional deviation from σ at state t^*. Hence, σ cannot be an ambiguous coalitional equilibrium, a contradiction. □

3.3.4 Closure

The closure condition for implementation as an ambiguous coalitional equilibrium is similar to those under the Bayesian implementation literature.

For any subset $E \subseteq T$, let the function $\mathbf{1}_E(\cdot) : T \to \{0, 1\}$ be the index function, which is equal to 1 when $t \in E$ and equal to 0 elsewhere.

Definition 3.3.5: *A social choice set F is said to satisfy the* ***closure*** *condition, if for any disjoint sequence of $(E^k \in \mathcal{F}_I)_{k\in K}$ such that $T^* = \cup_{k\in K} E^k$ and any sequence of social choice functions $(f^{E^k} \in F)_{k\in K}$, any function $f : T \to A$ satisfying $f(t) = \sum_{t\in T} \mathbf{1}_{E^k}(t) f^{E^k}(t)$ for all $t \in T^*$ is an element of F.*

Proposition 3.3.4: *If a social choice set F is implementable as an ambiguous coalitional equilibrium, then F satisfies the closure condition.*

Proof. As F is implementable, for each E^k and thus f^{E^k}, there exists an ambiguous coalitional equilibrium σ^{E^k} such that $g(\sigma^{E^k}(t)) = f^{E^k}(t)$ for all $t \in T^*$.

For each $i \in I$, define $\sigma_i : T_i \to M_i$ by $\sigma_i(t_i) = \sigma_i^{E^k}(t_i)$ for all $t_i \in T_i$ such that $\mathcal{F}_i(t_i) \subseteq E^k$. Then it is easy to prove that σ is an ambiguous coalitional equilibrium and leads to an outcome that is consistent with f in T^*. Since F is implementable, we have $f \in F$. □

3.4 Sufficient Conditions

The sufficient conditions to implement a social choice set F as an ambiguous coalitional equilibrium are usually slightly stronger than the necessary conditions. We will impose the additional condition, the bad outcome property, and construct a mechanism (M, g) to implement F.

Definition 3.4.1: *A social choice set F satisfies the* ***bad outcome property*** *if there exists $\underline{a} \in A$ and $\delta > 0$ such that $u_i(f(t'), t) - u_i(\underline{a}, t) \geq \delta$ for all $i \in I$, $f \in F$, and $t, t' \in T$.*

For example, consider a quasilinear environment where a social choice function has a non-linear part q and a monetary transfer part $(\xi_i)_{i \in I}$. Each agent i has a quasilinear utility function $u_i\big((q(t), \xi(t)), t\big) = v_i(q(t), t) + \xi_i(t)$. To bound the utilities, assume that q and each ξ_i are bounded functions. Taking a sufficiently large transfer from agents can usually serve as a bad outcome.

For full implementation of social choice sets, usually a full-support assumption is important. For example, Jackson (1991) focuses on implementation on T^*, where each type profile on T^* has positive probability. However, under the Wald-type maximin preferences, there are beliefs that impose all weights on the worst-cast events, and thus not all beliefs have full support over T^*. This brings in difficulties in the sufficiency proof. As a result, we impose the following assumption for Theorem 3.4.1.

Assumption 3.4.1: *For each $t \in T^*$, there exists $j \in I$ such that*

1. *there exists $\pi_j(t_j) \in \Pi_j(t_j)$ such that $\pi_j(t_j)[t_{-j}] = 1$, or*
2. *$\pi_j(t_j)[t'_{-j}] > 0$ for all t'_{-j} such that $(t_j, t'_{-j}) \in \mathcal{F}_j(t_j)$ and $\pi_j(t_j) \in \Pi_j(t_j)$.*

The assumption is satisfied, when for all $t \in T^*$, there exists an agent $j \in I$ such that $\Pi_j(t_j)$ includes all distributions over $\{t'_{-j} \in T_{-j} | (t_j, t'_{-j}) \in \mathcal{F}_j(t_j)\}$ or $\Pi_j(t_j)$ is a set of full-support distributions over $\{t'_{-j} \in T_{-j} | (t_j, t'_{-j}) \in \mathcal{F}_j(t_j)\}$. Hence, both the Bayesian preferences adopted by Jackson (1991) and the Wald-type maximin preferences of de Castro and Yannelis (2018) satisfy the above assumption.

Theorem 3.4.1: *Suppose Assumption 3.4.1 holds. A social choice set F is implementable as an ambiguous coalitional equilibrium if*

1. *$I \notin \mathcal{S}$ and F satisfies ambiguous coalitional incentive compatibility, ambiguous coalitional monotonicity, closure, and the bad outcome property;*
2. *$I \in \mathcal{S}$ and F satisfies ambiguous coalitional incentive compatibility, ambiguous coalitional monotonicity, closure, local Pareto efficiency, and the bad outcome property.*

Proof. We construct a mechanism (M, g) to implement F. Each agent i reports a message $m_i = (m_i^1, m_i^2, m_i^3, m_i^4, m_i^5)$, where $m_i^1 \in T_i$, $m_i^2 \in F$, $m_i^3 \in \mathbb{N}_+$, $m_i^4 \in \mathbb{N}_+$, $m_i^5 \in \{h : T \to A\}$. We partition the message space into M^1, M^2, and M^3 as follows:

$$M^1 = \{m | \exists f \in F \text{ s.t. } m_i = (\cdot, f, 0, \cdot, \cdot) \forall i \in I\},$$

$$M^2(S) = \{m | \exists f \in F, K_1 > 0, h \in H_S^f \text{ s.t. } m_i = (\cdot, f, K_1, \cdot, h) \forall i \in S, m_j = (\cdot, f, 0, \cdot, \cdot)$$
$$\forall j \notin S\},$$

$$M^2 = \bigcup_{S \in \mathcal{S}} M^2(S),$$

$$M^3 = M \backslash \{M^1 \cup M^2\}.$$

Let $\underline{a}$ be a "bad outcome" and $\delta > 0$ be a number described in the in the bad outcome property. Pick any $f^0 \in F$, and let $\underline{a}_\epsilon(\cdot)$ be $f^0(\cdot)$ with probability $\epsilon > 0$ and $\underline{a}$ with probability $1-\epsilon$, where ϵ is sufficiently small such that $u_i(\underline{a}_\epsilon(t'), t) = \epsilon u_i(f^0(t'), t) + (1-\epsilon)u_i(\underline{a}, t) < u_i(\underline{a}, t) + \delta \leq u_i(f'(t), t)$ for all $t, t' \in T$, $i \in I$, and $f' \in F$. Notice that the “ $=$ ” relies on the additivity of the utility function, which is a result of its integral form. The “ $<$ ” relies on the boundedness of u_i. By the bad outcome property, we also have $u_i(\underline{a}, t) + \epsilon\delta \leq u_i(\underline{a}_\epsilon(t'), t)$ for all $t, t' \in T$ and $i \in I$.

If $m \in M^1$, let the outcome allocation be $g(m) = f(m^1)$.

If $m \in M^2$, there exists $S \in \mathcal{S}$ such that $m \in M^2(S)$. Let $g(m)$ be a lottery $\tilde{h}(m^1)$, which has a realization of $h(m^1)$, with probability $K_1/(K_1+1)$, $\underline{a}_\epsilon(m^1)$ with probability $(1/(nK_1+n))\sum_{i\in I}(m_i^4/(m_i^4+1))$, and $\underline{a}$ with probability $(1/(nK_1+n))\sum_{i\in I}(1/(m_i^4+1))$.

If $m \in M^3$, let $g(m)$ be $\underline{a}_\epsilon(m^1)$ with probability $(1/n)\sum_{i\in I}(m_i^4/(m_i^4+1))$ and $\underline{a}$ with probability $(1/n)\sum_{i\in I}(1/(m_i^4+1))$.

The outcomes in M^2 and M^3 are compound lotteries of $\underline{a}$, $\underline{a}_\epsilon(m^1)$, and $h(m^1)$. The additivity of the utility function implies that the higher weight the lottery puts on $\underline{a}_\epsilon(m^1)$ as opposed to $\underline{a}$, the better the outcome is.

Claim 3.4.1: *For each $f \in F$, $\sigma_i^*(t_i) = (t_i, f, 0, \cdot, \cdot)$ for all $i \in I$ and $t_i \in T_i^*$ constitutes an ambiguous coalitional equilibrium of (M, g).*

Proof: We wish to show that for any $S \in \mathcal{S}$ and strategy profile σ'_S, σ'_S is not a profitable deviation from σ^*_S.

Suppose $I \in \mathcal{S}$. We consider the grand coalition's deviation at $t^* \in T^*$. Suppose there exists $f' \in F$ such that $g(\sigma'(t)) = f'(t)$ for all $t \in \cup_{i \in S}\mathcal{F}_i(t_i^*)$. By the local Pareto efficiency of F, the deviation is not profitable. Suppose there exists $\underline{S} \subseteq I$, $\underline{S} \in \mathcal{S}$, and $t \in \cup_{i \in S}\mathcal{F}_i(t_i^*)$ such that $\sigma'(t) \in M^2(\underline{S})$, then by the definition of $H^f_{\underline{S}}$, the deviation is not profitable. Suppose there exists $t \in \cup_{i \in S}\mathcal{F}_i(t_i^*)$ such that $\sigma'(t) \in M^3$, by the bad outcome property, the deviation is not profitable.

Now consider any non-grand coalition $S \in \mathcal{S}$. Suppose $(\sigma'_S(t_S), \sigma^*_{-S}(t_{-S})) \in M^1$ for all $t \in \cup_{i \in S}\mathcal{F}_i(t_i^*)$, by ambiguous coalitional incentive compatibility, S does not have the incentive to deviate with such a strategy profile σ'_S. Suppose there exists $\underline{S} \subseteq S$, $\underline{S} \in \mathcal{S}$, and $t \in \cup_{i \in S}\mathcal{F}_i(t_i^*)$ such that $\sigma'(t) \in M^2(\underline{S})$, then by the definition of $H^f_{\underline{S}}$, the deviation is not profitable. Suppose there exists $t \in \cup_{i \in S}\mathcal{F}_i(t_i^*)$ such that $\big(\sigma'_S(t_S), \sigma^*_{-S}(t_{-S})\big) \in M^3$, by the bad outcome property, the deviation is not profitable for S.

This completes the proof of the claim.

Claim 3.4.2: *If σ is an ambiguous coalitional equilibrium of the mechanism (M, g), then $\sigma(t) \in M^1$ for all $t \in T^*$.*

Proof: Decompose agent i's strategy $\sigma_i : T_i \to M_i$ by $\sigma_i = (\sigma_i^1, \sigma_i^2, \sigma_i^3, \sigma_i^4, \sigma_i^5)$. Suppose by way of contradiction that there exists $t \in T^*$ such that $\sigma(t) \notin M^1$. Below we show that there exists $j \in I$ who is strictly better off with the strategy σ'_j defined as $\sigma'_j(t_j) = \big(\sigma_j^1(t_j), \sigma_j^2(t_j), \sigma_j^3(t_j), 1 + \sigma_j^4(t_j), \sigma_j^5(t_j)\big)$ and $\sigma'_j(t'_j) = \sigma_j(t'_j)$ for $t'_j \neq t_j$. This would contradict the fact that σ is an ambiguous coalitional equilibrium.

Suppose that there exists $S \in \mathcal{S}$ and $t \in T^*$ such that $\sigma(t) \in M^2(S)$. This implies that there is an agent $j \in I$ with type t_j such that $\sigma_j^3(t_j) > 0$. Let agent j with type t_j deviates with strategy σ'_j. For all t'_{-j} such that $(t_j, t'_{-j}) \in \mathcal{F}_j(t_j)$, when $\sigma(t_j, t'_{-j})$ in M^2, $(\sigma'_j(t_j), \sigma_{-j}(t'_{-j}))$ leads to a strictly better lottery in M^2; when $\sigma(t_j, t'_{-j})$ in M^3, $(\sigma'_j(t_j), \sigma_{-j}(t'_{-j}))$ leads to a strictly better lottery in M^3. This means that the maximin expected utility of deviating is strictly higher, contradicting the fact that σ is an ambiguous coalitional equilibrium.

Suppose that there exists $t \in T^*$ such that $\sigma(t) \in M^3$. By Assumption 3.4.1, there exists an agent j such that part one or two of the assumption is satisfied. For all t'_{-j} such that $(t_j, t'_{-j}) \in \mathcal{F}_j(t_j)$ and $\sigma(t_j, t'_{-j}) \in M^1$, the message profile $(\sigma'_j(t_j), \sigma_{-j}(t'_{-j}))$ leads to the same outcome with $\sigma(t_j, t'_{-j})$. For all t'_{-j} such that $(t_j, t'_{-j}) \in \mathcal{F}_j(t_j)$ and $\sigma(t_j, t'_{-j}) \in M^2 \cup M^3$, the message $(\sigma'_j(t_j), \sigma_{-j}(t'_{-j}))$ leads to a strictly better outcome than $\sigma(t_j, t'_{-j})$. Then we want to show that this agent j with type t_j can deviate with strategy σ'_j and improve her maximin expected utility. Let $\hat{\pi}_j(t_j) \in \Pi_j(t_j)$ be the belief such that

$$\sum_{t'_{-j} \in T_{-j}} u_j\big(g(\sigma'_j(t_j), \sigma_{-j}(t'_{-j})), (t_j, t'_{-j})\big)\hat{\pi}_j(t_j)[t'_{-j}]$$

$$= \min_{\pi_j(t_j) \in \Pi_j(t_j)} \sum_{t'_{-j} \in T_{-j}} u_j\big(g(\sigma'_j(t_j), \sigma_{-j}(t'_{-j})), (t_j, t'_{-j})\big)\pi_j(t_j)[t'_{-j}].$$

Below we want to show that there exists t'_{-j} such $(t_j, t'_{-j}) \in \mathcal{F}_j(t_j)$, $(\sigma'_j(t_j), \sigma_{-j}(t'_{-j})) \in M^2 \cup M^3$, and $\hat{\pi}_j(t_j)[t'_{-j}] > 0$.

To see this, we discuss case by case. When part two of Assumption 3.4.1 in the last paragraph holds, we know $\hat{\pi}_j(t_j)[t_{-j}] > 0$ where $(\sigma'_j(t_j), \sigma_{-j}(t_{-j})) \in M^3$. Then by letting $t'_{-j} = t_{-j}$, we can fulfill the goal stated at the end of the last paragraph. When part one of the assumption holds, $\hat{\pi}_j(t_j)$ cannot put all weights on the set of t'_{-j} such that $(\sigma'_j(t_j), \sigma_{-j}(t'_{-j})) \in M^1$. Because otherwise, the belief $\pi_j(t_j) \in \Pi_j(t_j)$ such that $\pi_j(t_j)[t_{-j}] = 1$ would bring a strictly lower expected utility than $\hat{\pi}_j(t_j)$ by the bad lottery construction over M^3. This contradicts the definition of $\hat{\pi}_j(t_j)$. Thus, there exists t'_{-j} such $(t_j, t'_{-j}) \in \mathcal{F}_j(t_j)$, $(\sigma'_j(t_j), \sigma_{-j}(t'_{-j})) \in M^2 \cup M^3$, and $\hat{\pi}_j(t_j)[t'_{-j}] > 0$.

Hence, we have

$$\begin{aligned} & \sum_{t'_{-j} \in T_{-j}} u_j\big(g(\sigma'_j(t_j), \sigma_{-j}(t'_{-j})), (t_j, t'_{-j})\big)\hat{\pi}_j(t_j)[t'_{-j}] \\ > & \sum_{t'_{-j} \in T_{-j}} u_j\big(g(\sigma(t_j, t'_{-j})), (t_j, t'_{-j})\big)\hat{\pi}_j(t_j)[t'_{-j}] \\ \geq & \min_{\pi_j(t_j) \in \Pi_j(t_j)} \sum_{t'_{-j} \in T_{-j}} u_j\big(g(\sigma(t_j, t'_{-j}), (t_j, t'_{-j})\big)\pi_j(t_j)[t'_{-j}], \end{aligned}$$

which means the deviation is profitable for agent j, contradicting the fact that σ is an ambiguous coalitional equilibrium.

Claim 3.4.3: *If σ is an ambiguous coalitional equilibrium of (M, g), then there exists $f' \in F$ such that $g(\sigma(t)) = f'(t)$ for all $t \in T^*$.*

Proof: From the previous claim and the closure condition, there exists $f \in F$ such that $g(\sigma(t)) = f(\sigma^1(t))$ for all $t \in T^*$. Suppose there does not exist $f' \in F$ such that $g(\sigma(t)) = f'(t)$ for all $t \in T^*$. Define a deception profile $\alpha : T \to T$ by $\alpha_i(t_i) = \sigma_i^1(t_i)$ for all $i \in I$ and $t_i \in T_i$. Then we know that α is unacceptable for f.

By the ambiguous coalitional monotonicity condition, there exists $S \in \mathcal{S}$, $t^* \in T^*$, and $h \in H_S^f$ such that the inequality in Definition 3.3.2 is satisfied for all $i \in S$ and the strict inequality holds for some $i \in S$. Pick a large integer $K^* > 0$. For each $i \in S$, define σ''_i by $\sigma''_i(t_i) = (\sigma_i^1(t_i), \sigma_i^2(t_i), K^*, \cdot, h)$ for all $t_i = t_i^*$, and define $\sigma''_i(t_i) = \sigma_i(t_i)$

elsewhere. Then for all $(t_S^*, t_{-S}) \in \cup_{i \in S}\mathcal{F}_i(t_i^*)$, $\big(\sigma''_S(t_S^*), \sigma_{-S}(t_{-S})\big) \in M^2(S)$. When K^* is sufficiently large, by the ambiguous coalitional monotonicity condition, this deviation is strictly profitable for S at t^*, a contradiction.

In view of the three claims, we have established that (M, g) implements F. □

When F is a singleton, the local Pareto efficiency condition and the closure condition hold trivially, and thus we have the following corollary.

Corollary 3.4.1: *If a social choice function f satisfies ambiguous coalitional incentive compatibility, ambiguous coalitional monotonicity, and the bad outcome property, then it is implementable as an ambiguous coalitional equilibrium.*

Notice that in this special case, agents can never submit a message profile in M^3. Thus, we do not need Assumption 3.4.1.

3.5 Double Implementation

Suppose the mechanism designer does not know the coalition pattern, then it is of interest to study when and how a social choice set is implementable under all coalition patterns.

There exists a mechanism (M, g) to implement a social choice set as an ambiguous coalitional equilibrium under all coalition patterns, if and only if (M, g) implements a social choice set as an ambiguous Nash equilibrium and an ambiguous strong equilibrium simultaneously. Thus, this question is called a "double implementation" question.

The following strengthening of ambiguous coalitional incentive compatibility condition is necessary for implementation.

Definition 3.5.1: *Under Assumption 3.4.1, a social choice set F satisfies the* ***ambiguous strong incentive compatibility*** *condition if there exists no $f \in F$, $S \subseteq I$, $t^* \in T^*$, and $\alpha_S \colon T_S \to T_S$ such that*

$$\min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(f\big(\alpha_S(t_i^*, t_{S\setminus\{i\}}), t_{-S}\big), (t_i^*, t_{-i})\Big)\pi_i(t_i^*)[t_{-i}]$$
$$\geq \min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\big(f(t_i^*, t_{-i}), (t_i^*, t_{-i})\big)\pi_i(t_i^*)[t_{-i}]$$

for all $i \in S$ and the strict inequality holds for some $i \in S$.

The proof of the condition's necessity is omitted, as it is a natural extension of the ones in mechanism design theory.

Subsequently, we provide a strengthening of the ambiguous coalitional monotonicity condition.

Given a social choice set F, a social choice function $f \in F$, and a agent $i \in I$, define the **strong reward set** $\bar{H}_i^f = (\bigcap_{S\, s.t.\, i\in S\subsetneq I} H^{f,S}) \cap (\bigcap_{f'\in F} H^{f',I})$. The strong reward set $\bar{H}_i^f$ is equal to the reward set $H_{\{i\}}^f$ under the coalition pattern $\mathcal{S} = 2^I \backslash \emptyset$.

The ambiguous coalitional monotonicity condition is defined as follows.

Definition 3.5.2: *A social choice set F satisfies the* ***ambiguous strong monotonicity*** *condition, if for all $f \in F$, whenever the deception profile $\alpha : T \to T$ is unacceptable, there exists $i \in I$, $t_i \in T_i$, and $h \in \bar{H}_i^f$ such that*

$$\min_{\pi_i(t_i)\in\Pi_i(t_i)} \sum_{t_{-i}\in T_{-i}} u_i\Big(h\big(t_i, \alpha_i(t_{-i})\big), t\Big)\pi_i(t_i)[t_{-i}] > \min_{\pi_i(t_i)\in\Pi_i(t_i)} \sum_{t_{-i}\in T_{-i}} u_i\Big(f\big(\alpha(t)\big), t\Big)\pi_i(t_i)[t_{-i}].$$

We sketch the proof of its necessity below. Suppose F is doubly implemented by (M, g) and the deception profile $\alpha : T \to T$ is unacceptable for $f \in F$. There exists an ambiguous strong equilibrium σ^* such that $g\big(\sigma^*(t)\big) = f(t)$ for all $t \in T^*$ such that $\sigma^* \circ \alpha$ is not an ambiguous Nash equilibrium. Hence, there exists $i \in I$, $t_i^* \in T_i$, and a strategy $\sigma_i' : T_i \to M_i$ such that

$$\min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(g\big(\sigma_i'(t_i^*), \sigma_{-i}^*(\alpha_{-i}(t_{-i}))\big), (t_i^*, t_{-i})\Big)\pi_i(t_i^*)[t_{-i}]$$

$$> \min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(g\big(\sigma^*(\alpha(t_i^*, t_{-i}))\big), (t_i^*, t_{-i})\Big)\pi_i(t_i^*)[t_{-i}].$$

Define $h : T \to A$ by $h(t) = g\big(\sigma_i'(t_i), \sigma_{-i}^*(t_{-i})\big)$ for all $t \in T$. Then the above inequality becomes

$$\min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(h\big(t_i^*, \alpha_{-i}(t_{-i})\big), (t_i^*, t_{-i})\Big)\pi_i(t_i^*)[t_{-i}]$$

$$> \min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(f\big(\alpha(t_i^*, t_{-i})\big), (t_i^*, t_{-i})\Big)\pi_i(t_i^*)[t_{-i}].$$

Now we need to establish that $h \in \bar{H}_i^f$.

Since σ^* is an ambiguous strong equilibrium, for any deception profile β, coalition $S \subsetneq I$ with $S \ni i$, $(\sigma_i' \circ \beta_i, \sigma_{S\setminus\{i\}}^* \circ \beta_{S\setminus\{i\}})$ cannot be a profitable deviation from σ_S^*. Then one can establish $h \in H^{f,S}$.

For any $f' \in F$, there exists an ambiguous strong equilibrium σ^{**} such that $g(\sigma^{**}(t)) = f'(t)$ for all $t \in T^*$. As $(\sigma_i' \circ \beta_i, \sigma_{-i}^* \circ \beta_{-i})$ is not a profitable deviation for I, one can also show that $h \in H^{f',I}$.

The result below shows that the above two conditions, as well as local Pareto efficiency, closure, and the bad outcome property are sufficient for double implementation.

Theorem 3.5.1: *Under Assumption 3.4.1, if a social choice set F satisfies ambiguous strong incentive compatibility, ambiguous strong monotonicity, closure, local Pareto efficiency, and the bad outcome property, then it is doubly implementable as an ambiguous Nash equilibrium and an ambiguous strong equilibrium.*

The proof is similar to that of Theorem 3.4.1, except that $M^2 = \cup_{\emptyset \subsetneq S \subseteq I} M^2(S)$. Notice that the proofs of the claims need to be modified naturally for the purpose of double implementation. We omit the details.

When F is a singleton, we have the following corollary.

Corollary 3.5.1: *If a social choice function f satisfies ambiguous strong incentive compatibility, ambiguous strong monotonicity, and the bad outcome property, then it is doubly implementable as an ambiguous Nash equilibrium and an ambiguous strong equilibrium.*

3.6 Wald-type Maximin Preferences: Applications

We impose the following assumption throughout this section for insights beyond the Bayesian implementation literature.

Assumption 3.6.1: *Agents have private-value utility functions in the ex-post stage and the Wald-type maximin preferences in the interim stage.*

This assumption helps us to establish Lemmas 3.6.1 and 3.6.2, and thus we adopt the notation $u_i(a, t_i)$ and $\min_{t_{-i} \in T_{-i}} u_i(f(t), t_i)$ to represent the ex-post and interim utilities respectively.

de Castro and Yannelis (2018) have adopted a weaker version of Pareto efficiency and shown that under Assumption 3.6.1, every Pareto efficient social choice function is ambiguous (individual) incentive compatible. The following result shows that under the stronger version of ambiguous Pareto efficiency, every ambiguous efficient social choice function is also ambiguous strong incentive compatible, and thus is immune from any coalitional misreport.

Lemma 3.6.1: *Under Assumption 3.6.1, any ambiguous Pareto efficient social choice function f satisfies the ambiguous strong coalitonal incentive compatibility condition.*

Proof. Let f be an ambiguous Pareto efficient social choice function. Suppose by way of contradiction that F is not ambiguous coalitional incentive compatible. Then there exists $S \subseteq I$, $t^* \in T^*$, and $\alpha_S \colon T_S \to T_S$ such that

$$\min_{t_{-i}\in T_{-i}} u_i\Big(f\big(\alpha_S(t_i^*, t_{S\setminus\{i\}}), t_{-S}\big), t_i^*\Big) \geq \min_{t_{-i}\in T_{-i}} u_i\big(f(t_i^*, t_{-i}), t_i^*\big)$$

for all $i \in S$ and the strict inequality holds for some $i \in S$.

Define a new social choice function $y : T \to A$ by

$$y(t) = \begin{cases} f\big(\alpha_S(t_S), t_{-S}\big) & \text{if } t \in \cup_{i\in S}\mathcal{F}_i(t_i^*), \\ f(t) & \text{otherwise.} \end{cases}$$

Now we prove that y Pareto improves upon f.

From the previous paragraph, we know that

$$\min_{t_{-i}\in T_{-i}} u_i\big(y(t_i^*, t_{-i}), t_i^*\big) \geq \min_{t_{-i}\in T_{-i}} u_i\big(f(t_i^*, t_{-i}), t_i^*\big)$$

for all $i \in S$ and the strict inequality holds for some $i \in S$.

For all $j \notin S$, define $Y(t_j^*) = \{\bar{y} \in A | \exists t_{-j} \in T_{-j}\, s.t.\, \bar{y} = y(t_j^*, t_{-j})\}$ and $X(t_j^*) = \{\bar{x} \in A | \exists t_{-j} \in T_{-j}\, s.t.\, \bar{x} = f(t_j^*, t_{-j})\}$. We want to establish below that $Y(t_j^*) \subseteq X(t_j^*)$.

To see this, for any $\bar{y} \in Y(t_j^*)$, there exists $t_{-j} \in T_{-j}$ such that $\bar{y} = y(t_j^*, t_{-j})$. When $(t_j^*, t_{-j}) \in \cup_{i \in S} \mathcal{F}_i(t_i^*)$, $y(t_j^*, t_{-j}) = f(\alpha_S(t_S), t_j^*, t_{-S \cup \{j\}}) \in X(t_j^*)$. When $(t_j^*, t_{-j}) \notin \cup_{i \in S} \mathcal{F}_i(t_i^*)$, $y(t_j^*, t_{-j}) = f(t_j^*, t_{-j}) \in X(t_j^*)$. As a result, $Y(t_j^*) \subseteq X(t_j^*)$, which implies the following inequality,

$$\min_{t_{-i} \in T_{-i}} u_j\big(y(t_j^*, t_{-j}), t_j^*\big) = \min_{\bar{y} \in Y(t_j^*)} u_j\big(\bar{y}, t_j^*\big) \geq \min_{\bar{x} \in X(t_j^*)} u_j\big(\bar{x}, t_j^*\big) = \min_{t_{-j} \in T_{-j}} u_j\big(f(t_j^*, t_{-j}), t_j^*\big).$$

As a result, y Pareto dominates f, contradicting the supposition that f is ambiguous Pareto efficient. □

The following result establishes an easy sufficient condition for ambiguous coalitional monotonicity condition. If a social choice set is ambiguous Pareto efficient and that the unacceptable deception lowers the maximin expected utility of one agent compared to unanimous truthful report, then the ambiguous coalitional monotonicity condition is satisfied.

Lemma 3.6.2: *Suppose Assumption 3.6.1 holds. Let F be a social choice set in which every function is ambiguous Pareto efficient. The set F satisfies the ambiguous strong monotonicity condition, if for any function $f \in F$ and unacceptable deception profile $\alpha : T \to T$, there exists an agent $i \in I$ and type $t_i \in T_i$ such that* $\min_{t_{-i} \in T_{-i}} u_i\big(f(t), t_i\big) > \min_{t_{-i} \in T_{-i}} u_i\big(f(\alpha(t)), t_i\big)$.

Proof. Suppose for any $f \in F$ and an unacceptable deception $\alpha : T \to T$, there exists $i \in I$ and $t_i \in T_i$ such that

$$\min_{t_{-i} \in T_{-i}} u_i\big(f(t), t_i\big) > \min_{t_{-i} \in T_{-i}} u_i\big(f(\alpha(t)), t_i\big).$$

Define a new social choice function $h : T \to A$ by $h(t) = f(t)$ for all $t \in T$. Define $H(t_i) = \{\bar{h} \in A | \exists t_{-i} \in T_{-i}\, s.t.\, \bar{h} = h(t)\}$ and $H(t_i, \alpha) = \{\bar{h} \in A | \exists t_{-i} \in T_{-i}\, s.t.\, \bar{h} = h(t_i, \alpha_{-i}(t_{-i}))\}$. It is easy to see that $H(t_i, \alpha) \subseteq H(t_i)$, which implies the following inequality,

$$\begin{aligned} \min_{t_{-i}\in T_{-i}} u_i\big(h(t_i,\alpha_{-i}(t_{-i})),t_i\big) &= \min_{\bar{h}\in H(t_i,\alpha)} u_i\big(\bar{h},t_i\big) \geq \min_{\bar{h}\in H(t_i)} u_i\big(\bar{h},t_i\big) = \min_{t_{-i}\in T_{-i}} u_i\big(h(t),t_i\big) \\ &= \min_{t_{-i}\in T_{-i}} u_i\big(f(t),t_i\big) > \min_{t_{-i}\in T_{-i}} u_i\big(f(\alpha(t)),t_i\big). \end{aligned}$$

The argument below shows that $h \in \bar{H}_i^f = (\bigcap_{S\, s.t.\, i\in S \subsetneq I} H^{f,S}) \cap (\bigcap_{f'\in F} H^{f',I})$.

For any coalition S such that $i \in S \subsetneq I$, as f is ambiguous strong incentive compatible, we know there does not exist $t^* \in T^*$ and a deception profile $\beta : T \to T$ such that

$$\min_{t_{-i}\in T_{-i}} u_i\Big(f\big(\beta_S(t_i^*, t_{S\setminus\{i\}}), t_{-S}\big), t_i^*\Big)\pi_i(t_i^*)[t_{-i}] \geq \min_{t_{-i}\in T_{-i}} u_i\big(f(t_i^*, t_{-i}), t_i^*\big)\pi_i(t_i^*)[t_{-i}]$$

for all $i \in S$ and the strict inequality holds for some $i \in S$. Also, since $h = f$, we could replace the function f by h on the left-hand side. Hence, we have established that $h \in H^{f,S}$.

For the grand coalition I and any $f' \in F$, since f' is ambiguous Pareto efficient, we know there does not exist a deception profile $\beta : T \to T$ and $t^* \in T^*$ such that

$$\min_{t_{-i}\in T_{-i}} u_i\Big(f'\big(\beta(t_i^*, t_{-i})\big), t_i^*\Big)\pi_i(t_i^*)[t_{-i}] \geq \min_{t_{-i}\in T_{-i}} u_i\big(f(t_i^*, t_{-i}), t_i^*\big)\pi_i(t_i^*)[t_{-i}]$$

for all $i \in I$ and the strict inequality holds for some $i \in I$. Thus we have also established that $h \in H^{f',I}$.

Hence, we have proved the lemma. □

3.6.1 Ambiguous Pareto Efficient Social Choice Functions

Let F be the set of all ambiguous Pareto efficient social choice functions. If the set F also satisfies the bad outcome property, then it is doubly implementable.

Recall that Assumption 3.4.1 holds because we have Wald-type maximin preferences.

Corollary 3.6.1: *If the set of all ambiguous Pareto efficient allocations F satisfies the bad outcome property, then F is doubly implementable as an ambiguous Nash equilibrium and an ambiguous strong equilibrium.*

Proof. Lemma 3.6.1 has proved that F is ambiguous strong incentive compatible.

For any social choice function $f \in F$ and deception $\alpha : T \to T$. Suppose α is unacceptable, then there exists $f \circ \alpha$ that is not ambiguous Pareto efficient. We know there must exist an agent $i \in I$ and $t_i \in T_i$ such that

$$\min_{t_{-i} \in T_{-i}} u_i\big(f(t), t_i\big) > \min_{t_{-i} \in T_{-i}} u_i\big(f(\alpha(t)), t_i\big).$$

Otherwise, as $f \circ \alpha$ gives agents the same interim utility with f, which would contradict with the supposition that $f \circ \alpha$ is not ambiguous Pareto efficient. By Lemma 3.6.2, F satisfies the ambiguous strong monotonicity condition.

The local Pareto efficiency condition follows from the ambiguous Pareto efficiency of F.

Recall that with Wald-type maximin preferences, $T = T^*$, and thus the closure condition also holds trivially.

In view of Theorem 3.5.1, when F satisfies the bad outcome property, F is doubly implementable as an ambiguous Nash equilibrium and an ambiguous strong equilibrium.

□

3.6.2 Maximin Core

In this subsection, we doubly implement the set of all maximin core allocations of de Castro et al. (2011) under the Wald-type maximin preferences and the private value utility functions.

We define a feasible outcome for an economy below. Suppose in an economy, there are L goods and the total amount of each good l is a non-negative number $e^l \in \mathbb{R}_+$. Each agent i has a deterministic initial endowment of $(e_i^1, e_i^2, ..., e_i^L) \in \mathbb{R}_+^L \backslash \{\mathbf{0}\}$. The set of feasible pure outcomes is $X = \{(x_1, x_2, ..., x_n) | x_i : T \to \mathbb{R}_+^L \, \forall i \in I$, and $\sum_{i \in I} x_i^l(t) \leq e^l \, \forall t \in T, l = 1, 2, ..., L\}$. When defined on pure outcomes, agents' utility functions are strictly increasing in each dimension of her private consumption. The set of feasible outcomes is $A = \Delta(X)$. For any social choice function $f : T \to A$, type profile $t \in T$, and coalition S, if there exists a lottery with support $A' \subseteq A$ and each $x(t) \in A'$ satisfies $\sum_{i \in S} x_i(t) \leq \sum_{i \in S} e_i$, then we say $\sum_{i \in S} f_i(t) \leq \sum_{i \in S} e_i$.

Let **0** be a vector of zeros, which can serve as a bad outcome when every social choice function $f \in F$ satisfies $\min_{t_{-i} \in T_{-i}} u_i(f(t_i, t_{-i}), t_i) \geq \min_{t_{-i} \in T_{-i}} u_i(e, t_i)$.

The following notion is a modification of de Castro et al. (2011)'s maximin core allocation in their Definition 3.12. The major difference is that we only require a blocking coalition to strictly improve the interim preferences of one member.

Definition 3.6.1: *A social choice function f is said to be a* ***maximin core*** *allocation if there* *does not exist $S \subseteq I$, $t^* \in T^*$, and another social choice function $y : T \to A$, such that*

1. $\sum_{i \in S} y_i(t) \leq \sum_{i \in S} e_i$ *for all* $t \in T$,
2. $\min_{t_{-i} \in T_{-i}} u_i(y(t_i^*, t_{-i}), t_i^*) \geq \min_{t_{-i} \in T_{-i}} u_i(f(t_i^*, t_{-i}), t_i^*)$ *for all* $i \in S$ *and the strict inequality holds for some* $i \in S$.

Corollary 3.6.2: *The set of all maximin core allocations is doubly implementable as an ambiguous Nash equilibrium and an ambiguous strong equilibrium.*

Proof. By setting $S = I$, it is easy to see that F satisfies the ambiguous Pareto efficiency condition. By Lemma 3.6.1, F is ambiguous strong incentive compatible.

For any social choice function $f \in F$ and unacceptable deception $\alpha : T \to T$. As $f \circ \alpha$ is not a maximin core allocation, there exists $S \subseteq I$, $t^* \in T^*$, and $y : T \to A$ such that

1. $\sum_{i \in S} y_i(t) \leq \sum_{i \in S} e_i$ for all $t \in T$,
2. $\min_{t_{-i} \in T_{-i}} u_i(y(t_i^*, t_{-i}), t_i^*) \geq \min_{t_{-i} \in T_{-i}} u_i(f(\alpha(t_i^*, t_{-i})), t_i^*)$ for all $i \in S$ and the strict inequality holds for some $i \in S$.

We suppose by way of contradiction that

$$\min_{t_{-i} \in T_{-i}} u_i\big(f(\alpha(t)), t_i\big) \geq \min_{t_{-i} \in T_{-i}} u_i\big(f(t), t_i\big) \, \forall i \in I, t_i \in T_i.$$

Then we know there exists $S \subseteq I$, $t^* \in T^*$, and $y : T \to A$ such that

1. $\sum_{i \in S} y_i(t) \leq \sum_{i \in S} e_i$ for all $t \in T$,
2. $\min_{t_{-i} \in T_{-i}} u_i\big(y(t_i^*, t_{-i}), t_i^*\big) \geq \min_{t_{-i} \in T_{-i}} u_i\big(f(t_i^*, t_{-i}), t_i^*\big)$ for all $i \in S$ and the strict inequality holds for some $i \in S$.

This contradicts with the fact that $f \in F$, the maximin core. Hence, we know that there exists an agent $i \in I$ and a type $t_i \in T_i$ such that

$$\min_{t_{-i} \in T_{-i}} u_i\big(f(\alpha(t)), t_i\big) < \min_{t_{-i} \in T_{-i}} u_i\big(f(t), t_i\big).$$

By Lemma 3.6.2, F satisfies the ambiguous strong monotonicity condition.

Follow the argument of Corollary 3.6.1, we can also establish the conditions of local Pareto efficiency and closure.

Notice that by setting S to be singleton coalitions, for all $f \in F$, $i \in I$ and $t_i \in T_i$, $\min_{t_{-i} \in T_{-i}} u_i(f(t), t_i) \geq \min_{t_{-i} \in T_{-i}} u_i(e, t_i) > u_i(\mathbf{0}, t_i)$. As a result, the outcome that gives all agents zero consumption can serve as a "bad outcome". Hence, the bad outcome property holds as well.

In view of Theorem 3.5.1, F is doubly implementable as an ambiguous Nash equilibrium and an ambiguous strong equilibrium. □

3.6.3 Maximin Value

In this subsection, we show that the (interim) maximin value allocation of Angelopoulos and Koutsougeras (2015) is doubly implementable under the Wald-type maximin preferences and the private value utility functions.

For each $t \in T$ and weight profile $\lambda(t) \in \mathbb{R}^I_+ \backslash \{\mathbf{0}\}$, define the characteristic function by $V_{\lambda,t}(\emptyset) = 0$, and for any coalition $S \subseteq I$, define

$$V_{\lambda,t}(S) = \max\{\sum_{i \in S} \lambda_i(t) \min_{t'_{-i} \in T_{-i}} u_i(x(t_i, t'_{-i}), t_i) | \sum_{i \in S} x_i(t') \leq \sum_{i \in S} e_i \, \forall t' \in T\}.$$

Notice that for any disjoint coalitions $S^1, S^2 \subseteq I$, we have $V_{\lambda,t}(S^1 \cup S^2) \geq V_{\lambda,t}(S^1) +$

$V_{\lambda,t}(S^2)$. The Shapley value of agent i under type profile t is defined as

$$Sh_i(V_{\lambda,t}) = \sum_{S \ni i} \frac{(|S|-1)!(|I|-|S|)!}{|I|!}[V_{\lambda,t}(S) - V_{\lambda,t}(S \backslash \{i\})].$$

The notion below comes from Definition 2 of Angelopoulos and Koutsougeras (2015).

Definition 3.6.2: *A social choice function $f : T \to A$ is a **maximin value** allocation if for each $t \in T$, there exists a weight profile $\lambda(t) \in \mathbb{R}^n_+ \backslash \{\mathbf{0}\}$ such that*

$$\lambda_i(t) \min_{t'_{-i} \in T_{-i}} u_i(f(t_i, t'_{-i}), t_i) = Sh_i(V_{\lambda,t}) \, \forall i \in I.$$

We denote $\lambda(t) >> \mathbf{0}$ if every dimension of $\lambda(t)$ is strictly positive.

Corollary 3.6.3: *Let F be the set of all maximin value allocations. If for each $f \in F$, its weight profile $\lambda(t) >> \mathbf{0}$ for all $t \in T$, then F is doubly implementable as an ambiguous Nash equilibrium and an ambiguous strong equilibrium.*

Proof. We first establish that F satisfies the ambiguous Pareto efficiency condition. Suppose not, then there exists $f \in F$, a social choice function $y : T \to A$, and a type profile $t^* \in T^*$ such that

$$\min_{t_{-i} \in T_{-i}} u_i(y(t^*_i, t_{-i}), t^*_i) \geq \min_{t_{-i} \in T_{-i}} u_i(f(t^*_i, t_{-i}), t^*_i)$$

for all $i \in I$ and and the strict inequality holds for some $i \in I$. Since $\lambda(t^*) >> \mathbf{0}$,

$$\sum_{i \in I} \lambda_i(t^*) \min_{t_{-i} \in T_{-i}} u_i(y(t^*_i, t_{-i}), t^*_i) > \sum_{i \in I} \lambda_i(t^*) \min_{t_{-i} \in T_{-i}} u_i(f(t^*_i, t_{-i}), t^*_i)$$

$$= \sum_{i \in I} Sh_i(V_{\lambda,t^*}) = V_{\lambda,t^*}(I),$$

a contradiction with the definition of $V_{\lambda,t^*}(I)$. Hence, F is ambiguous Pareto efficient. By Lemma 3.6.1, F is ambiguous strong incentive compatible.

For a social choice function $f \in F$ and an unacceptable deception $\alpha : T \to T$. We know that there exists an agent $i \in I$ and a type $t_i \in T_i$ such that

$$\min_{t_{-i}\in T_{-i}} u_i\big(f(\alpha(t)), t_i\big) < \min_{t_{-i}\in T_{-i}} u_i\big(f(t), t_i\big).$$

If not, $f \circ \alpha$ would either bring the same interim utilities with f or Pareto dominates f. The former contradicts the fact that $f \circ \alpha$ is not a value allocation and the latter contradicts the ambiguous Pareto efficiency of F. By Lemma 3.6.2, F satisfies the ambiguous strong monotonicity condition.

Follow the argument of Corollary 3.6.2, we can also establish the conditions of local Pareto efficiency and closure.

As in Corollary 3.6.2, to establish the bad outcome property, it suffices to verify that $\min_{t'_{-i}\in T_{-i}} u_i(f(t_i, t'_{-i}), t_i) \geq u_i(e, t_i) > u_i(\mathbf{0}, t_i)$ for all $f \in F$, $i \in I$ and $t_i \in T_i$. We have that

$$\begin{aligned}
&\lambda_i(t) \min_{t'_{-i}\in T_{-i}} u_i(f(t_i, t'_{-i}), t_i) = Sh_i(V_{\lambda,t})\\
=&\sum_{S\ni i} \frac{(|S|-1)!(|I|-|S|)!}{|I|!}[V_{\lambda,t}(S) - V_{\lambda,t}(S\backslash\{i\})]\\
\geq&\sum_{S\ni i} \frac{(|S|-1)!(|I|-|S|)!}{|I|!}[V_{\lambda,t}(\{i\}) + V_{\lambda,t}(S\backslash\{i\}) - V_{\lambda,t}(S\backslash\{i\})]\\
=&\sum_{S\ni i} \frac{(|S|-1)!(|I|-|S|)!}{|I|!} V_{\lambda,t}(\{i\}) = V_{\lambda,t}(\{i\}) = \lambda_i(t) u_i(e, t_i).
\end{aligned}$$

Since $\lambda_i(t) > 0$, we know $\min_{t'_{-i}\in T_{-i}} u_i(f(t_i, t'_{-i}), t_i) \geq u_i(e, t_i) > u_i(\mathbf{0}, t_i)$.

In view of Theorem 3.5.1, F is doubly implementable as an ambiguous Nash equilibrium and an ambiguous strong equilibrium. □

3.7 Conclusion

This paper introduces the maximin expected utility framework into the problem of fully implementing a social choice set as an ambiguous coalitional equilibrium. We also identify conditions for a social choice set to be doubly implementable as an ambiguous Nash equilibrium and an ambiguous strong equilibrium. Under the Wald-type maximin preferences, we doubly implement the set of all ambiguous efficient social choice functions, the maximin core, and the maximin value, and thus provide insights beyond the Bayesian implementation literature.

APPENDIX A
APPENDIX TO CHAPTER 1

A.1 Proofs and Examples

Proof of Lemma 1.3.1. It is sufficient to prove the "**only if**" direction. For simplicity, we only prove the first statement. The second statement can be proved in a similar way.

Suppose a mechanism with ambiguous transfers $\tilde{\mathcal{M}} = (M, \tilde{q}, \tilde{\Phi})$ extracts the full surplus, then there exists an equilibrium σ such that

$$-\sum_{\theta\in\Theta}\sum_{i\in I}\tilde{\phi}_i(\sigma(\theta))p(\theta) = \max_{\hat{q}:\Theta\to A}\sum_{\theta\in\Theta}\sum_{i\in I}u_i\big(\hat{q}(\theta),\theta\big)p(\theta), \forall\tilde{\phi}\in\tilde{\Phi}.$$

Define $q(\theta) = \tilde{q}(\sigma(\theta))$ for all $\theta \in \Theta$. For each $\tilde{\phi} \in \tilde{\Phi}$, define $\phi : \Theta \to \mathbb{R}^N$ by $\phi = \tilde{\phi} \circ \sigma$, and denote the collection of all ϕ by Φ.

Now we prove that the direct mechanism with ambiguous transfers $\mathcal{M} = (q, \Phi)$ is incentive compatible. To see this, for all $i \in I$, $\theta_i, \theta_i' \in \Theta_i$,

$$\begin{aligned}
&\inf_{\phi\in\Phi}\sum_{\theta_{-i}\in\Theta_{-i}}[u_i\big(q(\theta_i',\theta_{-i}),(\theta_i,\theta_{-i})\big) + \phi_i(\theta_i',\theta_{-i})]p_i(\theta_{-i}|\theta_i)\\
=&\inf_{\tilde{\phi}\in\tilde{\Phi}}\sum_{\theta_{-i}\in\Theta_{-i}}[u_i\big(\tilde{q}(\sigma(\theta_i',\theta_{-i})),(\theta_i,\theta_{-i})\big) + \tilde{\phi}_i(\sigma(\theta_i',\theta_{-i}))]p_i(\theta_{-i}|\theta_i)\\
\le&\inf_{\tilde{\phi}\in\tilde{\Phi}}\sum_{\theta_{-i}\in\Theta_{-i}}[u_i\big(\tilde{q}(\sigma(\theta_i,\theta_{-i})),(\theta_i,\theta_{-i})\big) + \tilde{\phi}_i(\sigma(\theta_i,\theta_{-i}))]p_i(\theta_{-i}|\theta_i)\\
=&\inf_{\phi\in\Phi}\sum_{\theta_{-i}\in\Theta_{-i}}[u_i\big(q(\theta_i,\theta_{-i}),(\theta_i,\theta_{-i})\big) + \phi_i(\theta_i,\theta_{-i})]p_i(\theta_{-i}|\theta_i),
\end{aligned}$$

where the inequality comes from the fact that $\sigma_i(\theta_i') \in M_i$ can be viewed as a message sent by i under the constant strategy. Therefore, $\mathcal{M}$ is incentive compatible. □

Lemma A.1.1: *If the BDP property holds for agent i, then for all $\bar{\theta}_i, \hat{\theta}_i \in \Theta_i$ with $\bar{\theta}_i \neq \hat{\theta}_i$, there exists $\psi^{\bar{\theta}_i\hat{\theta}_i} : \Theta \to \mathbb{R}^N$ such that,*

1. $\sum_{j \in I} \psi_j^{\bar{\theta}_i\hat{\theta}_i}(\theta) = 0$ *for all* $\theta \in \Theta$*;*
2. $\sum_{\theta_{-j} \in \Theta_{-j}} \psi_j^{\bar{\theta}_i\hat{\theta}_i}(\theta_j, \theta_{-j}) p_j(\theta_{-j}|\theta_j) = 0$ *for all* $j \in I$ *and* $\theta_j \in \Theta_j$*;*
3. $\sum_{\theta_{-i} \in \Theta_{-i}} \psi_i^{\bar{\theta}_i\hat{\theta}_i}(\hat{\theta}_i, \theta_{-i}) p_i(\theta_{-i}|\bar{\theta}_i) < 0.$

Proof. We start with defining vectors e_θ for all $\theta \in \Theta$ and $p_{\theta_j\theta_j'}$ for all $j \in I, \theta_j, \theta_j' \in \Theta_j$. Each of the vectors has $N \times |\Theta|$ dimensions, and each dimension corresponds to an agent and a type profile. For each $\theta \in \Theta$, let all elements of e_θ that correspond to the type profile θ be 1 and everywhere else be 0. For each $j \in I$ and $\theta_j, \theta_j' \in \Theta_j$, let elements of $p_{\theta_j\theta_j'}$ that correspond to the agent j and some type profile (θ_j', θ_{-j}) be $p_j(\theta_{-j}|\theta_j)$ for all $\theta_{-j} \in \Theta_{-j}$. Everywhere else of $p_{\theta_j\theta_j'}$ is 0.[1]

Suppose by way of contradiction that the BDP property holds for agent i, but there exists $\bar{\theta}_i, \hat{\theta}_i \in \Theta_i$ with $\bar{\theta}_i \neq \hat{\theta}_i$, such that no $\psi^{\bar{\theta}_i\hat{\theta}_i}$ satisfies the three conditions. By Fredholm's theorem of the alternative, there exist coefficients $(a_{\theta_j})_{j \in I, \theta_j \in \Theta_j}$ and $(b_\theta)_{\theta \in \Theta}$ such that

$$p_{\bar{\theta}_i\hat{\theta}_i} = \sum_{j \in I} \sum_{\theta_j \in \Theta_j} a_{\theta_j} p_{\theta_j\theta_j} + \sum_{\theta \in \Theta} b_\theta e_\theta. \tag{A.1}$$

Fix any agent $j \neq i$. All elements of $p_{\bar{\theta}_i\hat{\theta}_i}$ that correspond to agent j are zero. All those corresponding to agent i and $\bar{\theta}_i$ are zero, too. Those corresponding to agent i and $\hat{\theta}_i$

[1] As an illustration, we look at a two-agent example with Θ being $((\theta_1^1, \theta_2^1), (\theta_1^1, \theta_2^2), (\theta_1^2, \theta_2^1), (\theta_1^2, \theta_2^2))$. For each e_θ or $p_{\theta_j\theta_j'}$, any of its first four dimensions corresponds to agent 1 and a type profile. Any of its last four dimensions corresponds to agent 2 and a type profile. Then for example, $e_{(\theta_1^2, \theta_2^1)} = (0, 0, 1, 0, 0, 0, 1, 0)$ and $p_{\theta_2^2\theta_2^1} = (0, 0, 0, 0, p_2(\theta_1^1|\theta_2^2), 0, p_2(\theta_1^2|\theta_2^2), 0)$.

may not be zero. The three observations, along with expression (A.1), imply that

$$0 = a_{\theta_j} p_j(\theta_i, \theta_{-i-j}|\theta_j) + b_{\theta_i,\theta_j,\theta_{-i-j}}, \forall \theta_i, \theta_j, \theta_{-i-j}, \tag{A.2}$$

$$0 = a_{\bar{\theta}_i} p_i(\theta_j, \theta_{-i-j}|\bar{\theta}_i) + b_{\bar{\theta}_i,\theta_j,\theta_{-i-j}}, \forall \theta_j, \theta_{-i-j}, \tag{A.3}$$

$$p_i(\theta_j, \theta_{-i-j}|\bar{\theta}_i) = a_{\hat{\theta}_i} p_i(\theta_j, \theta_{-i-j}|\hat{\theta}_i) + b_{\hat{\theta}_i,\theta_j,\theta_{-i-j}}, \forall \theta_j, \theta_{-i-j}. \tag{A.4}$$

By choosing $\theta_i = \bar{\theta}_i$ in expression (A.2) and cancelling $b_{\bar{\theta}_i,\theta_j,\theta_{-i-j}}$ in expressions (A.2) and (A.3), we have $a_{\theta_j} p_j(\bar{\theta}_i, \theta_{-i-j}|\theta_j) = a_{\bar{\theta}_i} p_i(\theta_j, \theta_{-i-j}|\bar{\theta}_i)$. Summing across all $\theta_{-i-j} \in \Theta_{-i-j}$ when $N \geq 3$ and ignoring any θ_{-i-j} when $N = 2$ yields $a_{\theta_j} p_j(\bar{\theta}_i|\theta_j) = a_{\bar{\theta}_i} p_i(\theta_j|\bar{\theta}_i)$. As $p(\cdot)$ is a common prior, we further know $a_{\theta_j} = a_{\bar{\theta}_i}\frac{p(\theta_j)}{p(\bar{\theta}_i)}$ for all $\theta_j \in \Theta_j$.

By choosing $\theta_i = \hat{\theta}_i$ in expression (A.2) and plugging in a_{θ_j} derived in the previous paragraph, we know $b_{\hat{\theta}_i,\theta_j,\theta_{-i-j}} = - a_{\bar{\theta}_i}\frac{p(\theta_j)}{p(\bar{\theta}_i)} p_j(\hat{\theta}_i, \theta_{-i-j}|\theta_j) = -a_{\bar{\theta}_i}\frac{p(\hat{\theta}_i)}{p(\bar{\theta}_i)} p_i(\theta_j, \theta_{-i-j}|\hat{\theta}_i)$ for all θ_j, θ_{-i-j}.

By plugging $b_{\hat{\theta}_i,\theta_j,\theta_{-i-j}}$ derived in the previous paragraph into expression (A.4), we obtain $p_i(\theta_j, \theta_{-i-j}|\bar{\theta}_i) = (a_{\hat{\theta}_i} - a_{\bar{\theta}_i}\frac{p(\hat{\theta}_i)}{p(\bar{\theta}_i)}) p_i(\theta_j, \theta_{-i-j}|\hat{\theta}_i)$ for all θ_j,θ_{-i-j}. Hence, $a_{\hat{\theta}_i} - a_{\bar{\theta}_i}\frac{p(\hat{\theta}_i)}{p(\bar{\theta}_i)} = 1$ and $p_i(\cdot|\bar{\theta}_i) = p_i(\cdot|\hat{\theta}_i)$, a contradiction. □

Lemma A.1.2: *For any positive integer K and any matrix $X_{K\times K}$ whose diagonal elements are all negative, there exists $\lambda \in \mathbb{R}_+^K \backslash \{\mathbf{0}\}$ such that $\sum_{\tilde{k}=1}^{K} x_{k\tilde{k}}\lambda_{\tilde{k}} \neq 0$ for all $k \in \{1, ..., K\}$.*

Proof. We prove the result by induction.

First, let $K = 1$. Pick an arbitrary $\lambda_1 > 0$. As $x_{11} < 0$, the statement holds for 1.

Suppose the statement holds for $K-1$, where $K \geq 2$. Now we consider an arbitrary $X_{K\times K}$ with negative diagonal elements. By the supposition for the northwest $K - 1$ by $K - 1$ block, there exists $(\lambda_1, ..., \lambda_{K-1}) \in \mathbb{R}_+^{K-1} \backslash \{\mathbf{0}\}$ such that $\sum_{\tilde{k}=1}^{K-1} x_{k\tilde{k}}\lambda_{\tilde{k}} \neq 0$ for all

$k \in \{1, ..., K-1\}$.

Case 1. Suppose $\sum_{\tilde{k}=1}^{K-1} x_{K\tilde{k}}\lambda_{\tilde{k}} \neq 0$. Let $\lambda_K = 0$, and thus the statement holds for K.

Case 2. Suppose $\sum\limits_{\tilde{k}=1}^{K-1} x_{K\tilde{k}}\lambda_{\tilde{k}} = 0$ and $x_{Kk_0}\lambda_{k_0} \neq 0$ for some $k_0 \in \{1, ..., K-1\}$. Let $(\lambda'_1, ..., \lambda'_{K-1}) = (\lambda_1, ..., \lambda_{k_0-1}, \lambda_{k_0}+\epsilon, \lambda_{k_0+1}, ..., \lambda_{K-1})$ for $\epsilon > 0$. Then $\sum_{\tilde{k}=1}^{K-1} x_{K\tilde{k}}\lambda'_{\tilde{k}} \neq 0$. When ϵ is sufficiently close to zero, $\sum_{\tilde{k}=1}^{K-1} x_{k\tilde{k}}\lambda'_{\tilde{k}} \neq 0$ for all $k \in \{1, ..., K-1\}$. Therefore, we can replace $(\lambda_1, ..., \lambda_{K-1})$ with $(\lambda'_1, ..., \lambda'_{K-1})$ and go back to Case 1.

Case 3. Suppose $x_{K\tilde{k}}\lambda_{\tilde{k}} = 0$ for all $\tilde{k} \in \{1, ..., K-1\}$. Let $\lambda_K > 0$ and $\lambda_K \neq -\frac{\sum_{\tilde{k}=1}^{K-1} x_{k\tilde{k}}\lambda_{\tilde{k}}}{x_{kK}}$ for all $k \in \{1, ..., K-1\}$ with $x_{kK} \neq 0$. Then the statement holds for K. □

Lemma A.1.3: *If the BDP property holds for all agents, then there exists $\psi : \Theta \to \mathbb{R}^N$ such that*

1. $\sum\limits_{i \in I} \psi_i(\theta) = 0$ *for all* $\theta \in \Theta$*;*
2. $\sum\limits_{\theta_{-i} \in \Theta_{-i}} \psi_i(\theta_i, \theta_{-i})p_i(\theta_{-i}|\theta_i) = 0$ *for all* $i \in I$ *and* $\theta_i \in \Theta_i$*;*
3. $\sum\limits_{\theta_{-i} \in \Theta_{-i}} \psi_i(\hat{\theta}_i, \theta_{-i})p_i(\theta_{-i}|\bar{\theta}_i) \neq 0$ *for all* $i \in I$ *and* $\bar{\theta}_i, \hat{\theta}_i \in \Theta_i$ *with* $\bar{\theta}_i \neq \hat{\theta}_i$*.*

Proof. Let K be the cardinality of $\mathcal{K} = \{(\bar{\theta}_i, \hat{\theta}_i)|i \in I, \bar{\theta}_i, \hat{\theta}_i \in \Theta_i, \bar{\theta}_i \neq \hat{\theta}_i\}$. Let $f : \mathcal{K} \to \{1, ..., K\}$ be a one-to-one mapping, which allows us to index the elements of $\mathcal{K}$.

For all $k, \tilde{k} \in \{1, ..., K\}$ ($k, \tilde{k}$ may be equal), where $f^{-1}(k) = (\bar{\theta}_i, \hat{\theta}_i)$ and $f^{-1}(\tilde{k}) = (\tilde{\bar{\theta}}_j, \tilde{\hat{\theta}}_j)$, we define $x_{k\tilde{k}} = \sum_{\theta_{-i} \in \Theta_{-i}} \psi_i^{\tilde{\bar{\theta}}_j\tilde{\hat{\theta}}_j}(\hat{\theta}_i, \theta_{-i})p_i(\theta_{-i}|\bar{\theta}_i)$, where each $\psi^{\tilde{\bar{\theta}}_j\tilde{\hat{\theta}}_j}$ is defined and proved to exist in Lemma A.1.1. By the third property of $\psi^{\tilde{\bar{\theta}}_j\tilde{\hat{\theta}}_j}$, we know $x_{\tilde{k}\tilde{k}} < 0$.

From Lemma A.1.2, there exists $\lambda \in \mathbb{R}_+^K \backslash \{\mathbf{0}\}$ such that $\sum_{\tilde{k}=1}^{K} x_{k\tilde{k}}\lambda_{\tilde{k}} \neq 0$ for all

$k \in \{1, ..., K\}$. This implies that for all $(\bar{\theta}_i, \hat{\theta}_i) \in \mathcal{K}$,

$$\sum_{\tilde{k}=1}^{K}[\sum_{\theta_{-i}\in\Theta_{-i}} \psi_i^{f^{-1}(\tilde{k})}(\hat{\theta}_i,\theta_{-i})p_i(\theta_{-i}|\bar{\theta}_i)]\lambda_{\tilde{k}} = \sum_{\theta_{-i}\in\Theta_{-i}}[\sum_{\tilde{k}=1}^{K}\lambda_{\tilde{k}}\psi_i^{f^{-1}(\tilde{k})}(\hat{\theta}_i,\theta_{-i})]p_i(\theta_{-i}|\bar{\theta}_i) \neq 0.$$

Define $\psi = \sum_{\tilde{k}=1}^{K}\lambda_{\tilde{k}}\psi^{f^{-1}(\tilde{k})}$. Then ψ satisfies the third requirement of this lemma. The other two requirements are trivial because ψ is a linear combination of transfer rules satisfying the two equations. □

Proof of Theorem 1.4.1. As it is without loss of generality to focus on incentive compatible direct mechanisms, full surplus extraction is equivalent to finding incentive compatible and interim individually rational direct mechanism with ambiguous transfers (q, Φ) such that

$$-\sum_{\theta\in\Theta}\sum_{i\in I}\phi_i(\theta)p(\theta) = \max_{\hat{q}:\Theta\to A}\sum_{\theta\in\Theta}\sum_{i\in I}u_i\big(\hat{q}(\theta),\theta\big)p(\theta), \forall\phi\in\Phi. \tag{A.5}$$

We first claim that an incentive compatible and interim individually rational direct mechanism with ambiguous transfers (q, Φ) extracts the full surplus if and only if q is ex-post efficient and $\sum_{\theta_{-i}\in\Theta_{-i}}[u_i\big(q(\theta_i,\theta_{-i}),(\theta_i,\theta_{-i})\big) + \phi_i(\theta_i,\theta_{-i})]p_i(\theta_{-i}|\theta_i) = 0$ for all $i \in I$, $\theta_i \in \Theta_i$, and $\phi_i \in \Phi_i$. The "if" direction is clear from expression (A.5). To see the "only if" direction, suppose q is inefficient or there exists $i \in I$, $\theta_i \in \Theta_i$, and $\phi \in \Phi$ such that $\sum_{\theta_{-i}\in\Theta_{-i}}[u_i\big(q(\theta_i,\theta_{-i}),(\theta_i,\theta_{-i})\big)+\phi_i(\theta_i,\theta_{-i})]p_i(\theta_{-i}|\theta_i) > 0$. By individual rationality and the fact that p is a common prior, we also have

$$-\sum_{\theta\in\Theta}\sum_{j\in I}\phi_j(\theta)p(\theta) \leq \sum_{\theta\in\Theta}\sum_{j\in I}u_j\big(q(\theta),\theta\big)p(\theta) \leq \max_{\hat{q}:\Theta\to A}\sum_{\theta\in\Theta}\sum_{j\in I}u_j\big(\hat{q}(\theta),\theta\big)p(\theta). \tag{A.6}$$

Combining the strict and weak inequalities and taking into account Assumption 1.2.1, we know at least one of the weak inequalities in expression (A.6) should be strict, a fact that contradicts expression (A.5).

Subsequently, we prove the **necessity** of the BDP property for full surplus extraction. Suppose by way of contradiction that there exists $i \in I$ and $\bar{\theta}_i, \hat{\theta}_i \in \Theta$ with $\bar{\theta}_i \neq \hat{\theta}_i$ such that $p_i(\cdot|\bar{\theta}_i) = p_i(\cdot|\hat{\theta}_i)$ and surplus extraction can be guaranteed. Consider a private value auction environment with one dimensional valuations satisfying $\bar{\theta}_i > \hat{\theta}_i > \theta_j$ for $(j, \theta_j) \neq (i, \bar{\theta}_i), (i, \hat{\theta}_i)$. Full surplus extraction requires i to obtain the good. The argument in the previous paragraph and interim incentive compatibility require that

$$\inf_{\phi\in\Phi} \sum_{\theta_{-i}\in\Theta_{-i}} \big(\bar{\theta}_i + \phi_i(\bar{\theta}_i, \theta_{-i})\big) p_i(\theta_{-i}|\bar{\theta}_i) = 0 \geq \inf_{\phi\in\Phi} \sum_{\theta_{-i}\in\Theta_{-i}} \big(\bar{\theta}_i + \phi_i(\hat{\theta}_i, \theta_{-i})\big) p_i(\theta_{-i}|\bar{\theta}_i).$$

The above inequality, $p_i(\cdot|\bar{\theta}_i) = p_i(\cdot|\hat{\theta}_i)$, and the fact that $\bar{\theta}_i > \hat{\theta}_i$ imply

$$\inf_{\phi\in\Phi} \sum_{\theta_{-i}\in\Theta_{-i}} \big(\hat{\theta}_i + \phi_i(\hat{\theta}_i, \theta_{-i})\big) p_i(\theta_{-i}|\hat{\theta}_i) < 0,$$

which contradicts interim individual rationality of type-$\hat{\theta}_i$ agent i.

To demonstrate the **sufficiency** of the BDP property, pick an arbitrary ex-post efficient allocation rule q. Define two transfer rules ϕ and ϕ' by $\phi_i = -\eta_i + c\psi_i$ and $\phi'_i = -\eta_i - c\psi_i$ for all $i \in I$, where ψ is defined and proved to exist in Lemma A.1.3, $\eta_i(\theta) = u_i(q(\theta), \theta)$ for all $\theta \in \Theta$, and c is no less than

$$\max_{\substack{i,\bar{\theta}_i,\hat{\theta}_i\in\Theta_i,\\ \bar{\theta}_i\neq\hat{\theta}_i}} \frac{\sum_{\theta_{-i}\in\Theta_{-i}} [u_i\big(q(\hat{\theta}_i, \theta_{-i}), (\bar{\theta}_i, \theta_{-i})\big) - u_i\big(q(\hat{\theta}_i, \theta_{-i}), (\hat{\theta}_i, \theta_{-i})\big)] p_i(\theta_{-i}|\bar{\theta}_i)}{|\sum_{\theta_{-i}\in\Theta_{-i}} \psi_i(\hat{\theta}_i, \theta_{-i}) p_i(\theta_{-i}|\bar{\theta}_i)|}.$$

Define $\Phi = \{\phi, \phi'\}$. All interim individual rationality constraints bind because when agents truthfully report, each $-\eta_i$ extracts agent i's full surplus, and $c\psi_i$ has zero interim expected value under agent i's belief. To check incentive compatibility, notice the choice of c gives agents non-positive worst-case expected payoffs when they misreport. Hence, (q, Φ) extracts the full surplus. □

Proof of Theorem 1.4.2. **Necessity**. Suppose there exists agent $i \in I$ and her different types $\bar{\theta}_i, \hat{\theta}_i \in \Theta$ such that $p_i(\cdot|\bar{\theta}_i) = p_i(\cdot|\hat{\theta}_i)$. We will establish the existence of a profile of utility functions and an efficient allocation rule q such that q cannot be implemented via an interim individually rational and ex-post budget-balanced mechanism with ambiguous transfers.

Consider an adaptation of the utility functions constructed by Kosenok and Severinov (2008). Let $A = \{x_0, x_1, x_2\}$, where all agents' payoffs of consuming the outside option x_0 are zero. The payoffs for agent i and all $j \neq i$ to consume x_1 and x_2 are given below with $0 < a < B$.

Table A.1.1: Payoffs of Feasible Outcomes in Proof of Theorem 1.4.2

	$u_i\big(x_1, (\theta_i, \theta_j)\big)$	$u_j\big(x_1, (\theta_i, \theta_j)\big)$	$u_i\big(x_2, (\theta_i, \theta_j)\big)$	$u_j\big(x_2, (\theta_i, \theta_j)\big)$
$\theta_i = \bar{\theta}_i$	a	a	a+B	a-2B
$\theta_i = \hat{\theta}_i$	0	a	a	a
$\theta_i \neq \bar{\theta}_i, \hat{\theta}_i$	a	a	0	a

The efficient allocation rule is $q(\theta) = x_2$ if $\theta_i = \hat{\theta}_i$ and $q(\theta) = x_1$ elsewhere.

Suppose by way of contradiction that there exists an interim individually rational and ex-post budget-balanced mechanism with ambiguous transfers implementing q. Denote the set of transfers by Φ. Then from $IC(\bar{\theta}_i\hat{\theta}_i)$ and $IC(\hat{\theta}_i\bar{\theta}_i)$,

$$\inf_{\phi \in \Phi}\{a + \sum_{\theta_{-i} \in \Theta_{-i}} \phi_i(\bar{\theta}_i, \theta_{-i}) p_i(\theta_{-i}|\bar{\theta}_i)\} \geq \inf_{\phi \in \Phi}\{a + B + \sum_{\theta_{-i} \in \Theta_{-i}} \phi_i(\hat{\theta}_i, \theta_{-i}) p_i(\theta_{-i}|\bar{\theta}_i)\},$$

$$\inf_{\phi\in\Phi}\{a+\sum_{\theta_{-i}\in\Theta_{-i}}\phi_i(\hat{\theta}_i,\theta_{-i})p_i(\theta_{-i}|\hat{\theta}_i)\}\geq\inf_{\phi\in\Phi}\{0+\sum_{\theta_{-i}\in\Theta_{-i}}\phi_i(\bar{\theta}_i,\theta_{-i})p_i(\theta_{-i}|\hat{\theta}_i)\}.$$

Recall that $p_i(\cdot|\bar{\theta}_i)=p_i(\cdot|\hat{\theta}_i)$. Adding the above two inequalities gives $2a\geq a+B$, a contradiction. Therefore, q is not implementable via an interim individually rational and ex-post budget-balanced mechanism with ambiguous transfers.

Sufficiency. Pick an arbitrary ex-post budget-balanced transfer rule $\eta:\Theta\to\mathbb{R}^N$ such that $\sum_{\theta_{-i}\in\Theta_{-i}}[u_i\big(q(\theta_i,\theta_{-i}),(\theta_i,\theta_{-i})\big)+\eta_i(\theta_i,\theta_{-i})]p_i(\theta_{-i}|\theta_i)\geq 0$ for all $i\in I$ and $\theta_i\in\Theta_i$. By Lemma A.1.3, there exists an ex-post budget-balanced transfer rule ψ which gives all agents zero expected values when they truthfully report and gives an agent i non-zero expected value when she is the only misreporting agent.

Pick any c that is no less than

$$\frac{\sum_{\theta_{-i}\in\Theta_{-i}}[u_i\big(q(\hat{\theta}_i,\theta_{-i}),(\bar{\theta}_i,\theta_{-i})\big)+\eta_i(\hat{\theta}_i,\theta_{-i})-u_i\big(q(\bar{\theta}_i,\theta_{-i}),(\bar{\theta}_i,\theta_{-i})\big)-\eta_i(\bar{\theta}_i,\theta_{-i})]p_i(\theta_{-i}|\bar{\theta}_i)}{|\sum_{\theta_{-i}\in\Theta_{-i}}\psi_i(\hat{\theta}_i,\theta_{-i})p_i(\theta_{-i}|\bar{\theta}_i)|}$$

for all $i\in I,\bar{\theta}_i,\hat{\theta}_i\in\Theta_i$, and $\bar{\theta}_i\neq\hat{\theta}_i$, where c exists because the denominator is positive. Let $\mathcal{M}$ be $(q,\{\eta+c\psi,\eta-c\psi\})$.

Interim individual rationality of $\mathcal{M}$ comes from the choice of η and the fact that ψ gives agents zero expected values when they truthfully report. For all $i\in I$ and $\bar{\theta}_i,\hat{\theta}_i\in\Theta_i$ with $\bar{\theta}_i\neq\hat{\theta}_i$, the choice of c indicates that

$$\sum_{\theta_{-i}\in\Theta_{-i}}[u_i\big(q(\bar{\theta}_i,\theta_{-i}),(\bar{\theta}_i,\theta_{-i})\big)+\eta_i(\bar{\theta}_i,\theta_{-i})]p_i(\theta_{-i}|\bar{\theta}_i)\geq$$

$$\min\{\sum_{\theta_{-i}\in\Theta_{-i}}[u_i\big(q(\hat{\theta}_i,\theta_{-i}),(\bar{\theta}_i,\theta_{-i})\big)+\eta_i(\hat{\theta}_i,\theta_{-i})\pm c\psi_i(\hat{\theta}_i,\theta_{-i})]p_i(\theta_{-i}|\bar{\theta}_i)\},$$

and thus we have interim incentive compatibility of $\mathcal{M}$. Ex-post budget balance of $\mathcal{M}$

follows from the property of η and ψ. Therefore, $\mathcal{M}$ is an interim individually rational and ex-post budget-balanced mechanism with ambiguous transfers that implements q. $\square$

Proof of Theorem 1.5.1. **Necessity**. By relabeling the indices, we assume without loss of generality that agent 1 has identical beliefs under θ_1^1 and θ_1^2, agent 2 has identical beliefs under θ_2^1 and θ_2^2, and that $L_2 \geq L_1$. For each agent i, let θ_i and θ_{-i} be generic elements of Θ_i and Θ_{-i}. For convenience, θ_1^m and θ_2^n are also used to represent generic elements of Θ_1 and Θ_2. We ignore θ_{-1-2} if $N = 2$. Now we construct a profile of private value utility functions such that an efficient outcome is not implementable. This would establish the necessity of the condition that at least $N - 1$ agents satisfy the BDP property.

Let agent 1 own a unit of private good and all others be potential buyers. Let θ_i represent agent i's private value of trading, where $\theta_2^1 > -\theta_1^1 > \theta_2^2 > -\theta_1^2 > ... > \theta_2^{L_1} > -\theta_1^{L_1} > \theta_i > 0$ for all other θ_i. No trade gives all agents zero payoffs. The efficient allocation rule q is that agent 1 should trade with 2 if and only if $\theta_1^m + \theta_2^n > 0$ (note that $\theta_1^m + \theta_2^n \neq 0$ by construction).

Suppose by way of contradiction that an interim individually rational and ex-post budget-balanced mechanism with ambiguous transfers, denoted by $\mathcal{M} = (q, \Phi)$, implements q. By individual rationality, for all $i \in I$ and θ_i, type-θ_i agent i's worst-case expected payoff from participation is $U_{\theta_i} \geq 0$. Hence, by fixing any $\phi \in \Phi$, we have

$$\sum_{\theta_{-i} \in \Theta_{-i}} \phi_i(\theta_i, \theta_{-i}) p_i(\theta_{-i}|\theta_i) \geq U_{\theta_i} - \sum_{\theta_{-i} \in \Theta_{-i}} u_i(q(\theta_i, \theta_{-i}), (\theta_i, \theta_{-i})) p_i(\theta_{-i}|\theta_i) \qquad \text{(A.7)}$$

for $i \in I$ and $\theta_i \in \Theta_i$. Multiply each of the inequalities by $p(\theta_i)$ and sum across all i and θ_i. By ex-post budget balance, the left-hand side of the aggregated inequality is zero and the right-hand side,

$$\sum_{i\in I}\sum_{\theta_i\in\Theta_i} p(\theta_i)U_{\theta_i} + \sum_m p(\theta_1^m)\big(-\theta_1^m \sum_{n\le m} p_1(\theta_2^n|\theta_1^m)\big) + \sum_n p(\theta_2^n)\big(-\theta_2^n \sum_{m\ge n} p(\theta_1^m|\theta_2^n)\big), \tag{A.8}$$

is non-positive. From $IC(\theta_1^2\theta_1^1)$ and $IC(\theta_2^1\theta_2^2)$, for all $\epsilon > 0$, there exists $\phi^1, \phi^2 \in \Phi$ satisfying

$$IC(\theta_1^2\theta_1^1) \quad - \sum_{n,\theta_{-1-2}} \phi_1^1(\theta_1^1,\theta_2^n,\theta_{-1,2})p_1(\theta_2^n,\theta_{-1-2}|\theta_1^2) + \epsilon \ge -U_1^2 + \theta_1^2 \sum_{n\le 1} p_1(\theta_2^n|\theta_1^2),$$

$$IC(\theta_2^1\theta_2^2) \quad - \sum_{m,\theta_{-1-2}} \phi_2^2(\theta_1^m,\theta_2^2,\theta_{-1,2})p_2(\theta_1^m,\theta_{-1-2}|\theta_2^1) + \epsilon \ge -U_2^1 + \theta_2^1 \sum_{m\ge 2} p_2(\theta_1^m|\theta_2^1).$$

Note $p_1(\cdot|\theta_1^1) = p_1(\cdot|\theta_1^2)$ and $p_2(\cdot|\theta_2^1) = p_2(\cdot|\theta_2^2)$. Add $IC(\theta_1^2\theta_1^1)$ and (A.7), where $(i,\theta_i) = (1,\theta_1^1)$ and $\phi = \phi^1$. Then, let ϵ go to zero. Similarly, add $IC(\theta_2^1\theta_2^2)$ and (A.7), where $(i,\theta_i) = (2,\theta_2^2)$ and $\phi = \phi^2$. Then, let ϵ go to zero. We obtain the following two equations.

$$U_{\theta_1^2} \ge U_{\theta_1^1} + (\theta_1^2 - \theta_1^1)\sum_{n\le 1} p_1(\theta_2^n|\theta_1^1), \qquad U_{\theta_2^1} \ge U_{\theta_2^2} + (\theta_2^1 - \theta_2^2)\sum_{m\ge 2} p_2(\theta_1^m|\theta_2^1).$$

By plugging the above two inequalities into expression (A.8), we have that (A.8) is no less than

$$\sum_m p(\theta_1^m)\big(-\theta_1^m \sum_{n\le m} p_1(\theta_2^n|\theta_1^m)\big) + p(\theta_1^2)(\theta_1^2-\theta_1^1)\sum_{n\le 1} p_1(\theta_2^n|\theta_1^1)$$
$$+ \sum_n p(\theta_2^n)\big(-\theta_2^n \sum_{m\ge n} p_2(\theta_1^m|\theta_2^n)\big) + p(\theta_2^1)(\theta_2^1-\theta_2^2)\sum_{m\ge 2} p_2(\theta_1^m|\theta_2^1). \tag{A.9}$$

In the above expression, the coefficients of θ_1^1 and θ_2^1 are

$$-p(\theta_1^1)\sum_{n\le 1} p_1(\theta_2^n|\theta_1^1) - p(\theta_1^2)\sum_{n\le 1} p_1(\theta_2^n|\theta_1^1) = -\big(p(\theta_1^1)+p(\theta_1^2)\big)\frac{p(\theta_1^1,\theta_2^1)}{p(\theta_1^1)} < -p(\theta_1^1,\theta_2^1),$$

$$-p(\theta_2^1)\sum_{m\ge 1} p_2(\theta_1^m|\theta_2^1) + p(\theta_2^1)\sum_{m\ge 2} p_2(\theta_1^m|\theta_2^1) = -p(\theta_2^1)\frac{p(\theta_1^1,\theta_2^1)}{p(\theta_2^1)} = -p(\theta_1^1,\theta_2^1),$$

where the strict inequality follows from Assumption 1.2.1. Let θ_1^1 and θ_2^1 be sufficiently close in absolute value and all other values θ_i be close to zero. Then expression (A.9) is positive, contradicting $0 \geq (A.8) \geq (A.9)$. Therefore, q cannot be implemented via an interim individually rational and ex-post budget-balanced mechanism with ambiguous transfers.

Sufficiency. When all agents satisfy the BDP property, the sufficiency part is proven by Theorem 1.4.2. When there is exactly one agent, i, whose BDP property fails, following Lemmas A.1.1 through A.1.3, one can prove that there exists $\psi : \Theta \to \mathbb{R}^N$ such that

1. $\sum\limits_{j \in I} \psi_j(\theta) = 0$ for all $\theta \in \Theta$;
2. $\sum\limits_{\theta_{-j} \in \Theta_{-j}} \psi_j(\theta_j, \theta_{-j}) p_j(\theta_{-j}|\theta_j) = 0$ for all $j \in I$ and $\theta_j \in \Theta_j$;
3. $\sum\limits_{\theta_{-j} \in \Theta_{-j}} \psi_j(\hat{\theta}_j, \theta_{-j}) p_j(\theta_{-j}|\bar{\theta}_j) \neq 0$ for all $j \neq i$ and $\bar{\theta}_j, \hat{\theta}_j \in \Theta_j$ satisfying $\bar{\theta}_j \neq \hat{\theta}_j$.

Notice that the third statement is different from the one in Lemma A.1.3, as agent i in this theorem has identical beliefs under different types.

We construct a mechanism where agent i obtains all the surplus by truthfully reporting. For all $\theta \in \Theta$ and $j \in I$ with $j \neq i$, let $\eta_j(\theta) = -u_j(q(\theta), \theta_j)$, and $\eta_i(\theta) = -\sum_{j \neq i} \eta_j(\theta)$.

Pick any c that is no less than

$$\max_{\substack{j \neq i, \bar{\theta}_j, \hat{\theta}_j \in \Theta_j, \\ \bar{\theta}_j \neq \hat{\theta}_j}} \frac{\sum_{\theta_{-j} \in \Theta_{-j}} [u_j\big(q(\hat{\theta}_j, \theta_{-j}), \bar{\theta}_j\big) - u_j\big(q(\hat{\theta}_j, \theta_{-j}), \hat{\theta}_j\big)] p_j(\theta_{-j}|\bar{\theta}_j)}{|\sum_{\theta_{-j} \in \Theta_{-j}} \psi_j(\hat{\theta}_j, \theta_{-j}) p_j(\theta_{-j}|\bar{\theta}_j)|}.$$

Let the set of ambiguous transfers be $\Phi = \{\eta + c\psi, \eta - c\psi\}$, which is interim individually rational and ex-post budget-balanced. The choice of η, ψ, and c implies that for any agent $j \neq i$ with type $\bar{\theta}_j$, truthfully reporting gives her zero worst-case expected payoffs while lying gives her non-positive ones. Therefore, j's incentive compatibility constraints are satisfied.

For type-$\bar{\theta}_i$ agent i, the argument below verifies her incentive compatibility constraints:

$$\begin{aligned}
&\min\{\sum_{\theta_{-i}\in\Theta_{-i}}[u_i\big(q(\bar{\theta}_i,\theta_{-i}),\bar{\theta}_i\big)+\sum_{j\neq i}u_j\big(q(\bar{\theta}_i,\theta_{-i}),\theta_j\big)\pm c\psi_i(\bar{\theta}_i,\theta_{-i})]p_i(\theta_{-i}|\bar{\theta}_i)\}\\
=&\sum_{\theta_{-i}\in\Theta_{-i}}[u_i\big(q(\bar{\theta}_i,\theta_{-i}),\bar{\theta}_i\big)+\sum_{j\neq i}u_j\big(q(\bar{\theta}_i,\theta_{-i}),\theta_j\big)]p_i(\theta_{-i}|\bar{\theta}_i)\\
\geq&\sum_{\theta_{-i}\in\Theta_{-i}}[u_i\big(q(\hat{\theta}_i,\theta_{-i}),\bar{\theta}_i\big)+\sum_{j\neq i}u_j\big(q(\hat{\theta}_i,\theta_{-i}),\theta_j\big)]p_i(\theta_{-i}|\bar{\theta}_i)\\
\geq&\min\{\sum_{\theta_{-i}\in\Theta_{-i}}[u_i\big(q(\hat{\theta}_i,\theta_{-i}),\bar{\theta}_i\big)+\sum_{j\neq i}u_j\big(q(\hat{\theta}_i,\theta_{-i}),\theta_j\big)\pm c\psi_i(\hat{\theta}_i,\theta_{-i})]p_i(\theta_{-i}|\bar{\theta}_i)\},
\end{aligned}$$

where the equality comes from the second property of ψ, the first inequality comes from ex-post efficiency of q, and the second inequality comes from the minimization operation.

Therefore, the interim individually rational and ex-post budget-balanced mechanism with ambiguous transfers implements q. □

Example A.1.1: *In this private value example, $N-1$ agents satisfy the BDP property. But an inefficient allocation rule q is not implementable via an interim individually rational and ex-post budget-balanced mechanism with ambiguous transfers.*

Define a common prior p by $p(\theta_1^3,\theta_2^2)=2/7$, and $p(\theta)=1/7$ for all other θ. Only agent 2 satisfies the BDP property. Let feasible allocations be $A=\{x_0,x_1,x_2\}$. Recall that x_0, the outside option, gives both agents zero payoffs. The payoffs of x_1 and x_2 are presented below.

Table A.1.2: Feasible Outcomes of Example A.1.1

x_1	θ_2^1	θ_2^2
θ_1^1	0,0	0,0
θ_1^2	2,0	2,0
θ_1^3	0,0	0,0

x_2	θ_2^1	θ_2^2
θ_1^1	2,0	2,0
θ_1^2	0,0	0,0
θ_1^3	0,0	0,0

Consider an allocation rule $q(\theta) = x_2$ if $\theta_1 = \theta_1^2$, and $q(\theta) = x_1$ elsewhere. Suppose by way of contradiction that q is implemented by $\mathcal{M} = (q, \Phi)$, where each $\phi \in \Phi$ is interpreted as a payment from agent 1 to 2. Let U_1^1 and U_1^2 denote type-θ_1^1 and type-θ_1^2 agent 1's worst-case expected payoff from participation.

As $IR(\theta_1^1)$ and $IC(\theta_1^2\theta_1^1)$ hold, for any $\epsilon > 0$, there exists $\phi^1 \in \Phi$ such that

$$IR(\theta_1^1) \qquad -0.5\phi^1(\theta_1^1, \theta_2^1) - 0.5\phi^1(\theta_1^1, \theta_2^2) \geq U_1^1,$$

$$IC(\theta_1^2\theta_1^1) \qquad U_1^2 + \epsilon \geq 2 - 0.5\phi^1(\theta_1^1, \theta_2^1) - 0.5\phi^1(\theta_1^1, \theta_2^2).$$

Similarly, by $IR(\theta_1^2)$ and $IC(\theta_1^1\theta_1^2)$, for any $\epsilon > 0$, there exists $\phi^2 \in \Phi$ such that

$$IR(\theta_1^2) \qquad -0.5\phi^2(\theta_1^2, \theta_2^1) - 0.5\phi^2(\theta_1^2, \theta_2^2) \geq U_1^2,$$

$$IC(\theta_1^1\theta_1^2) \qquad U_1^1 + \epsilon \geq 2 - 0.5\phi^2(\theta_1^2, \theta_2^1) - 0.5\phi^2(\theta_1^2, \theta_2^2).$$

We add the above inequalities pairwise and let ϵ go to zero. Thus we have $U_1^2 \geq 2 + U_1^1$ and $U_1^1 \geq 2 + U_1^2$. These two expressions imply $0 \geq 4$, which is a contradiction.

Proof of Proposition 1.5.1. For each $i \in I$, let θ_i be a generic element of Θ_i. By relabeling the indices, we assume without loss of generality there are $(\beta_{\theta_1})_{\theta_1 \neq \theta_1^1}, (\beta_{\theta_2})_{\theta_2 \neq \theta_2^2} \geq \mathbf{0}$ such that $p_1(\cdot|\theta_1^1) = \sum_{\theta_1 \neq \theta_1^1} \beta_{\theta_1} p_1(\cdot|\theta_1)$ and $p_2(\cdot|\theta_2^2) = \sum_{\theta_2 \neq \theta_2^2} \beta_{\theta_2} p_2(\cdot|\theta_2)$, $L_2 \geq L_1$, and

$$\frac{\beta_{\theta_2^1}}{p(\theta_2^1)} \geq \frac{\beta_{\theta_2}}{p(\theta_2)}, \forall \theta_2 \neq \theta_2^1, \theta_2^2. \tag{A.10}$$

Suppose agent 1 owns a unit of private good and all others are potential buyers. For each $i \in I$, let θ_i be agent i's private value of trading, where $\theta_2^1 > -\theta_1^1 > \theta_2^2 > -\theta_1^2 > ... > \theta_2^{L_1} > -\theta_1^{L_1} > \theta_i$ for all other θ_i. No trade gives all agents zero payoffs. The efficient allocation rule q is that agent 1 should trade with 2 if and only if $\theta_1 + \theta_2 > 0$. Subsequently, we will prove that q is not implementable, which proves the necessity of the condition.

Suppose by way of contradiction there exists an individually rational and budget-balanced Bayesian transfer ϕ that implements q. Then by individual rationality and incentive compatibility, for all $i \in I$, $\bar{\theta}_i \neq \hat{\theta}_i$, the following inequalities hold:

$$IR(\bar{\theta}_i) \qquad \sum_{\theta_{-i}} \phi_i(\bar{\theta}_i, \theta_{-i}) p_i(\theta_{-i}|\bar{\theta}_i) \geq -\sum_{\theta_{-i}} u_i(q(\bar{\theta}_i, \theta_{-i}), \bar{\theta}_i) p_i(\theta_{-i}|\bar{\theta}_i),$$

$$IC(\bar{\theta}_i \hat{\theta}_i) \qquad \sum_{\theta_{-i}} \phi_i(\bar{\theta}_i, \theta_{-i}) p_i(\theta_{-i}|\bar{\theta}_i) - \sum_{\theta_{-i}} \phi_i(\hat{\theta}_i, \theta_{-i}) p_i(\theta_{-i}|\bar{\theta}_i)$$

$$\geq -\sum_{\theta_{-i}} u_i(q(\bar{\theta}_i, \theta_{-i}), \bar{\theta}_i) p_i(\theta_{-i}|\bar{\theta}_i) + \sum_{\theta_{-i}} u_i(q(\hat{\theta}_i, \theta_{-i}), \bar{\theta}_i) p_i(\theta_{-i}|\bar{\theta}_i).$$

We choose a constant $\delta > 0$ sufficiently large such that

$$\frac{\delta \beta_{\theta_2^1} p(\theta_2^2)}{p(\theta_2^1)} \geq \frac{\beta_{\theta_1} p(\theta_1^1)}{p(\theta_1)}, \forall \theta_1 \neq \theta_1^1, \tag{A.11}$$

and then denote the left-hand-side term by γ. Now we compute the weighted sum of the above individual rationality and incentive compatibility constraints where (1) the weight of $IR(\theta_1^1)$ is $p(\theta_1^1)(\gamma + 1)$, (2) for each $\theta_1 \neq \theta_1^1$ the weight of $IR(\theta_1)$ is $p(\theta_1)\gamma - \beta_{\theta_1} p(\theta_1^1)$, (3) the weight of $IR(\theta_2^2)$ is $p(\theta_2^2)(\gamma + \delta)$, (4) for each $\theta_2 \neq \theta_2^2$ the weight of $IR(\theta_2)$ is

$p(\theta_2)\gamma - \delta\beta_{\theta_2}p(\theta_2^2)$, (5) for each $i \neq 1, 2$ and $\theta_i \in \Theta_i$ the weight of $IR(\theta_i)$ is $p(\theta_i)\gamma$, (6) for each $\theta_1 \neq \theta_1^1$ the weight of $IC(\theta_1\theta_1^1)$ is $p(\theta_1^1)\beta_{\theta_1}$, (7) for each $\theta_2 \neq \theta_2^2$ the weight of $IC(\theta_2\theta_2^2)$ is $\delta\beta_{\theta_2}p(\theta_2^2)$, and (8) every other inequality has weight zero. From expressions (A.10) and (A.11), we know all the weights are non-negative.

Ex-post budget balance cancels all terms containing transfers in the weighted sum, and thus the left-hand side is zero. On the right-hand side, the coefficients of θ_1^1 and θ_2^1 are $-(\gamma + 1)p(\theta_1^1, \theta_2^1)$ and $-\gamma p(\theta_1^1, \theta_2^1)$ respectively. Therefore, by choosing θ_1^1 and θ_2^1 sufficiently close in absolute value and all other θ_i close to zero, the right-hand side of the weighted sum is positive, a contradiction. □

Lemma A.1.4: *Given the belief system* $\left(p_i(\cdot|\theta_i)\right)_{i\in I, \theta_i \in \Theta_i}$*, if the BDP and NCP* properties hold for agent* i *, then for all* $\bar{\theta}_i, \hat{\theta}_i \in \Theta_i$ *with* $\bar{\theta}_i \neq \hat{\theta}_i$*, there exists* $\psi^{\bar{\theta}_i\hat{\theta}_i} : \Theta \to \mathbb{R}^N$ *such that,*

1. $\sum\limits_{j\in I} \psi_j^{\bar{\theta}_i\hat{\theta}_i}(\theta) = 0$ *for all* $\theta \in \Theta$*;*
2. $\sum\limits_{\theta_{-j}\in\Theta_{-j}} \psi_j^{\bar{\theta}_i\hat{\theta}_i}(\theta_j, \theta_{-j})p_j(\theta_{-j}|\theta_j) \geq 0$ *for all* $j \in I$*,* $\theta_j \in \Theta_j$*;*
3. $\sum\limits_{\theta_{-i}\in\Theta_{-i}} \psi_i^{\bar{\theta}_i\hat{\theta}_i}(\hat{\theta}_i, \theta_{-i})p_i(\theta_{-i}|\bar{\theta}_i) < 0.$

Proof. We prove by contraposition. Suppose there exists $\bar{\theta}_i \neq \hat{\theta}_i$ such that no $\psi^{\bar{\theta}_i\hat{\theta}_i}$ satisfies the above three requirements. By Motzkin's theorem of the alternative, there exist coefficients $(b_\theta)_{\theta\in\Theta}$ and non-negative coefficients $(a_{\theta_j})_{j\in I, \theta_j\in\Theta_j}$ such that

$$p_{\bar{\theta}_i\hat{\theta}_i} = \sum_{j\in I}\sum_{\theta_j\in\Theta_j} a_{\theta_j}p_{\theta_j\theta_j} - \sum_{\theta\in\Theta} b_\theta e_\theta. \tag{A.12}$$

We will subsequently establish that expression (A.12) holds **if and only if** either $p_i(\cdot|\bar{\theta}_i) =$

$p_i(\cdot|\hat{\theta}_i)$ or both groups of equations in the NCP* property are satisfied by $i, \bar{\theta}_i, \hat{\theta}_i$, a distribution $\mu \in \Delta(\Theta)$, constants $\bar{C} > 0$, and $\hat{C} > 1$. The only if part would imply either the BDP or NCP* property is violated for agent i.

We prove the "**only if**" direction first. Expression (A.12) implies

$$a_{\theta_i} p_i(\theta_j, \theta_{-i-j}|\theta_i) = b_{\theta_i,\theta_j,\theta_{-i-j}}, \forall \theta_i \neq \hat{\theta}_i, j \neq i, \theta_j, \theta_{-i-j}, \tag{A.13}$$

$$a_{\hat{\theta}_i} p_i(\theta_j, \theta_{-i-j}|\hat{\theta}_i) - p_i(\theta_j, \theta_{-i-j}|\bar{\theta}_i) = b_{\hat{\theta}_i,\theta_j,\theta_{-i-j}}, \forall j \neq i, \theta_j, \theta_{-i-j}, \tag{A.14}$$

$$a_{\theta_j} p_j(\theta_i, \theta_{-i-j}|\theta_j) = b_{\theta_i,\theta_j,\theta_{-i-j}}, \forall \theta_i, j \neq i, \theta_j, \theta_{-i-j}. \tag{A.15}$$

We remark that throughout the proof, if $N = 2$, we ignore any term θ_{-i-j} to avoid introducing additional notation. By canceling $b_{\bar{\theta}_i\theta_j\theta_{-i-j}}$ in (A.13) and (A.15), we also have $a_{\bar{\theta}_i} \geq 0$.

Case 1. Suppose $a_{\tilde{\theta}_i} = 0$ for some $\tilde{\theta}_i \neq \hat{\theta}_i$. The argument below shows that $a_{\hat{\theta}_i} = 1$, $a_{\theta_j} = 0$ for all $(j, \theta_j) \neq (i, \hat{\theta}_i)$, $b_{\theta_i,\theta_j,\theta_{-i-j}} = 0$ for all θ_i, θ_j, and θ_{-i-j}, and $p_i(\cdot|\bar{\theta}_i) = p_i(\cdot|\hat{\theta}_i)$.

Canceling $b_{\tilde{\theta}_i,\theta_j,\theta_{-i-j}}$ in (A.13) and (A.15) yields

$$0 = a_{\tilde{\theta}_i} p_i(\theta_j, \theta_{-i-j}|\tilde{\theta}_i) = a_{\theta_j} p_j(\tilde{\theta}_i, \theta_{-i-j}|\theta_j)$$

for all $j \neq i, \theta_j, \theta_{-i-j}$. From Assumption 1.5.1, it must be the case that $a_{\theta_j} = 0$ for all $j \neq i$ and θ_j.

By expression (A.15), the previous paragraph implies $b_{\theta_i,\theta_j,\theta_{-i-j}} = 0$ for all θ_i, θ_j, and θ_{-i-j}. From expression (A.13), we further know $a_{\theta_i} = 0$ for all $\theta_i \neq \hat{\theta}_i$.

By canceling $b_{\hat{\theta}_i,\theta_j,\theta_{-i-j}}$ in (A.14) and (A.15), we have

$$a_{\hat{\theta}_i} p_i(\theta_j, \theta_{-i-j}|\hat{\theta}_i) - p_i(\theta_j, \theta_{-i-j}|\bar{\theta}_i) = a_{\theta_j} p_j(\hat{\theta}_i, \theta_{-i-j}|\theta_j) = 0$$

for all θ_j and θ_{-i-j}. Summing the equation across all θ_j and θ_{-i-j}, we get $a_{\hat{\theta}_i} = 1$ and thus $p_i(\cdot|\bar{\theta}_i) = p_i(\cdot|\hat{\theta}_i)$.

Case 2. Suppose $a_{\theta_i} > 0$ for all $\theta_i \neq \hat{\theta}_i$. Similar to the argument of the previous case, we know $a_{\hat{\theta}_i} > 1$ and $a_{\theta_j} > 0$ for all $(j, \theta_j) \neq (i, \hat{\theta}_i)$. Subsequently, we will establish that $i, \bar{\theta}_i, \hat{\theta}_i$, a distribution $\mu \in \Delta(\Theta)$, constants $\bar{C} > 0$, and $\hat{C} > 1$ satisfy both groups of equations in the NCP* property so that the property fails.

Define $\mu \in \Delta(\Theta)$ by $\mu(\theta) = \frac{b_\theta}{\sum_{\tilde{\theta}\in\Theta} b_{\tilde{\theta}}}$ for all $\theta \in \Theta$. Then from expressions (A.13) and (A.15), we know $\mu(\cdot|\theta_j) = p_j(\cdot|\theta_j)$ and $\mu(\theta_j) = \frac{a_{\theta_j}}{\sum_{\tilde{\theta}\in\Theta} b_{\tilde{\theta}}} > 0$ for all $(j, \theta_j) \neq (i, \hat{\theta}_i)$. Hence, the first group of equations in the statement of the NCP* property holds. By canceling $b_{\hat{\theta}_i\theta_j\theta_{-i-j}}$ in expressions (A.14) and (A.15), we have $a_{\hat{\theta}_i} p_i(\theta_j, \cdot|\hat{\theta}_i) = p_i(\theta_j, \cdot|\bar{\theta}_i) + a_{\theta_j} p_j(\hat{\theta}_i, \cdot|\theta_j)$ for all $j \neq i$ and θ_j, where $a_{\theta_j} = \mu(\theta_j) \sum_{\tilde{\theta}\in\Theta} b_{\tilde{\theta}} = \mu(\bar{\theta}_i) \frac{\mu(\theta_j|\bar{\theta}_i)}{\mu(\bar{\theta}_i|\theta_j)} \sum_{\tilde{\theta}\in\Theta} b_{\tilde{\theta}} = a_{\bar{\theta}_i} \frac{p_i(\theta_j|\bar{\theta}_i)}{p_j(\bar{\theta}_i|\theta_j)}$. Recall $a_{\bar{\theta}_i} > 0$ and $a_{\hat{\theta}_i} > 1$. Thus by defining $\bar{C} = a_{\bar{\theta}_i}$ and $\hat{C} = a_{\hat{\theta}_i}$, we can see the second group of equations in the NCP* property also holds. Hence, the BDP property fails.

Now we prove the "**if**" direction, which will be used by Lemma A.1.5. When $p_i(\cdot|\bar{\theta}_i) = p_i(\cdot|\hat{\theta}_i)$, define (1) $a_{\hat{\theta}_i} = 1$, (2) $a_{\theta_j} = 0$ for all $(j, \theta_j) \neq (i, \hat{\theta}_i)$, and (3) $b_\theta = 0$ for all $\theta \in \Theta$. When the two groups of equations in the NCP* property hold for $i, \bar{\theta}_i, \hat{\theta}_i, \mu, \bar{C}$, and $\hat{C}$, define (1) $a_{\bar{\theta}_i} = \bar{C}$, $a_{\hat{\theta}_i} = \hat{C}$, (2) $b_{\theta_i\theta_{-i}} = \bar{C}\frac{\mu(\theta_i,\theta_{-i})}{\mu(\bar{\theta}_i)}$, $\forall\theta_i, \theta_{-i}$, and (3) $a_{\theta_k} = \bar{C}\frac{\mu(\theta_k)}{\mu(\bar{\theta}_i)}$, $\forall(k, \theta_k) \neq (i, \bar{\theta}_i), (i, \hat{\theta}_i)$. For both cases, it is easy to verify expression (A.12). □

Proof of Theorem 1.5.2. Suppose the BDP and NCP* properties hold for all agents. According to Lemma A.1.4, for all $i \in I$ and $\bar{\theta}_i, \hat{\theta}_i \in \Theta_i$ with $\bar{\theta}_i \neq \hat{\theta}_i$, there exists $\psi^{\bar{\theta}_i\hat{\theta}_i} : \Theta \to \mathbb{R}^N$, such that the three requirements are satisfied.

Let η be any interim individually rational and ex-post budget-balanced transfer rule. Define $\Phi = \{\eta, \eta + c\psi^{\bar{\theta}_j\hat{\theta}_j} : j \in I, \bar{\theta}_j, \hat{\theta}_j \in \Theta_j, \bar{\theta}_j \neq \hat{\theta}_j\}$, where c is sufficiently large such that for all $j \in I$ and $\bar{\theta}_j, \hat{\theta}_j \in \Theta_j$ with $\bar{\theta}_j \neq \hat{\theta}_j$, the expression $\sum_{\theta_{-j}\in\Theta_{-j}}[u_j(q(\hat{\theta}_j,\theta_{-j}),(\bar{\theta}_j,\theta_{-j})) - u_j(q(\bar{\theta}_j,\theta_{-j}),(\bar{\theta}_j,\theta_{-j})) + \eta_j(\hat{\theta}_j,\theta_{-j}) - \eta_j(\bar{\theta}_j,\theta_{-j}) + c\psi_j^{\bar{\theta}_j\hat{\theta}_j}(\hat{\theta}_j,\theta_{-j})]p_j(\theta_{-j}|\bar{\theta}_j)$ is negative.

For any type-$\bar{\theta}_i$ agent i, the inequality below shows that misreporting $\hat{\theta}_i$ is not profitable:

$$\begin{aligned}
&\min_{\phi\in\Phi} \sum_{\theta_{-i}\in\Theta_{-i}} [u_i(q(\bar{\theta}_i,\theta_{-i}),(\bar{\theta}_i,\theta_{-i})) + \phi_i(\bar{\theta}_i,\theta_{-i})]p_i(\theta_{-i}|\bar{\theta}_i)\\
=& \sum_{\theta_{-i}\in\Theta_{-i}} [u_i(q(\bar{\theta}_i,\theta_{-i}),(\bar{\theta}_i,\theta_{-i})) + \eta(\bar{\theta}_i,\theta_{-i})]p_i(\theta_{-i}|\bar{\theta}_i)\\
\geq& \sum_{\theta_{-i}\in\Theta_{-i}} [u_i(q(\hat{\theta}_i,\theta_{-i}),(\bar{\theta}_i,\theta_{-i})) + \eta(\hat{\theta}_i,\theta_{-i}) + c\psi_i^{\bar{\theta}_i\hat{\theta}_i}(\hat{\theta}_i,\theta_{-i})]p_i(\theta_{-i}|\bar{\theta}_i)\\
\geq& \min_{\phi\in\Phi} \sum_{\theta_{-i}\in\Theta_{-i}} [u_i(q(\hat{\theta}_i,\theta_{-i}),(\bar{\theta}_i,\theta_{-i})) + \phi_i(\hat{\theta}_i,\theta_{-i})]p_i(\theta_{-i}|\bar{\theta}_i),
\end{aligned}$$

where the equality follows from the second requirement of Lemma A.1.4 and the composition of ambiguous transfers, the first inequality comes from the choice of c, and the second inequality comes from the composition of ambiguous transfers again. Interim individual rationality and ex-post budget balance follow from corresponding properties of η and each $\phi \in \Phi$. □

Lemma A.1.5: *Given beliefs $\left(p_i(\cdot|\theta_i)\right)_{i\in I,\theta_i\in\Theta_i}$ that are not generated by a common prior, if the BDP property holds for all agents, then the NCP* property holds for at least $N-1$ agents.*

Proof. Let all agents satisfy the BDP property. Suppose by way of contradiction that there are agents $i \neq j$ for whom the NCP* property fails. Suppose types $\bar{\theta}_i \neq \hat{\theta}_i$ and $\bar{\theta}_j \neq \hat{\theta}_j$ are the two pairs that fail the NCP* property. By the two-case argument of Lemma A.1.4, there exist coefficients $(a_{\theta_k})_{k\in I,\theta_k\in\Theta_k} > \mathbf{0}$ where $a_{\hat{\theta}_i} > 1$, $(b_\theta)_{\theta\in\Theta}$, $(c_{\theta_k})_{k\in I,\theta_k\in\Theta_k} > \mathbf{0}$ where $c_{\hat{\theta}_j} > 1$, and $(d_\theta)_{\theta\in\Theta}$ such that $p_{\bar{\theta}_i\hat{\theta}_i} = \sum_{k\in I}\sum_{\theta_k\in\Theta_k} a_{\theta_k} p_{\theta_k\theta_k} - \sum_{\theta\in\Theta} b_\theta e_\theta$ and $p_{\bar{\theta}_j\hat{\theta}_j} = \sum_{k\in I}\sum_{\theta_k\in\Theta_k} c_{\theta_k} p_{\theta_k\theta_k} - \sum_{\theta\in\Theta} d_\theta e_\theta$. Thus, the following equations hold. Note that we ignore θ_{-i-j} if $N = 2$.

$$a_{\theta_i} p_i(\theta_j, \theta_{-i-j}|\theta_i) = b_{\theta_i,\theta_j,\theta_{-i-j}}, \forall \theta_i \neq \hat{\theta}_i, \forall \theta_j, \theta_{-i-j},$$

$$a_{\hat{\theta}_i} p_i(\theta_j, \theta_{-i-j}|\hat{\theta}_i) - p_i(\theta_j, \theta_{-i-j}|\bar{\theta}_i) = b_{\hat{\theta}_i,\theta_j,\theta_{-i-j}}, \forall \theta_j, \theta_{-i-j}$$

$$a_{\theta_j} p_j(\theta_i, \theta_{-i-j}|\theta_j) = b_{\theta_i,\theta_j,\theta_{-i-j}}, \forall \theta_i, \theta_j, \theta_{-i-j},$$

$$c_{\theta_i} p_i(\theta_j, \theta_{-i-j}|\theta_i) = d_{\theta_i,\theta_j,\theta_{-i-j}}, \forall \theta_i, \theta_j, \theta_{-i-j},$$

$$c_{\theta_j} p_j(\theta_i, \theta_{-i-j}|\theta_j) = d_{\theta_i,\theta_j,\theta_{-i-j}}, \forall \theta_j \neq \hat{\theta}_j, \forall \theta_i, \theta_{-i-j},$$

$$c_{\hat{\theta}_j} p_j(\theta_i, \theta_{-i-j}|\hat{\theta}_j) - p_j(\theta_i, \theta_{-i-j}|\bar{\theta}_j) = d_{\theta_i,\hat{\theta}_j,\theta_{-i-j}}, \forall \theta_i, \theta_{-i-j}.$$

Canceling all $b_{\theta_i,\theta_j,\theta_{-i-j}}$, $d_{\theta_i,\theta_j,\theta_{-i-j}}$, and $p_j(\theta_i, \theta_{-i-j}|\theta_j)$ in the above equations yields:

$$\frac{a_{\theta_i} p_i(\theta_j, \theta_{-i-j}|\theta_i)}{a_{\theta_j}} = \frac{c_{\theta_i} p_i(\theta_j, \theta_{-i-j}|\theta_i)}{c_{\theta_j}}, \forall \theta_i \neq \hat{\theta}_i, \forall \theta_j \neq \hat{\theta}_j, \forall \theta_{-i-j}, \tag{A.16}$$

$$\frac{a_{\hat{\theta}_i} p_i(\theta_j, \theta_{-i-j}|\hat{\theta}_i) - p_i(\theta_j, \theta_{-i-j}|\bar{\theta}_i)}{a_{\theta_j}} = \frac{c_{\hat{\theta}_i} p_i(\theta_j, \theta_{-i-j}|\hat{\theta}_i)}{c_{\theta_j}}, \forall \theta_j \neq \hat{\theta}_j, \forall \theta_{-i-j}, \tag{A.17}$$

$$\frac{a_{\theta_i} p_i(\hat{\theta}_j, \theta_{-i-j}|\theta_i)}{a_{\hat{\theta}_j}} = \frac{c_{\theta_i} p_i(\hat{\theta}_j, \theta_{-i-j}|\theta_i)}{c_{\hat{\theta}_j}} + \frac{c_{\theta_i} p_i(\bar{\theta}_j, \theta_{-i-j}|\theta_i)}{c_{\hat{\theta}_j} c_{\bar{\theta}_j}}, \forall \theta_i \neq \hat{\theta}_i, \forall \theta_{-i-j}, \tag{A.18}$$

$$\frac{a_{\hat{\theta}_i} p_i(\hat{\theta}_j, \theta_{-i-j}|\hat{\theta}_i) - p_i(\hat{\theta}_j, \theta_{-i-j}|\bar{\theta}_i)}{a_{\hat{\theta}_j}} = \frac{c_{\hat{\theta}_i} p_i(\hat{\theta}_j, \theta_{-i-j}|\hat{\theta}_i)}{c_{\hat{\theta}_j}} + \frac{c_{\hat{\theta}_i} p_i(\bar{\theta}_j, \theta_{-i-j}|\hat{\theta}_i)}{c_{\hat{\theta}_j} c_{\bar{\theta}_j}}, \forall \theta_{-i-j}. \tag{A.19}$$

Step 1. We want to prove for all $\theta_{-i-j} \in \Theta_{-i-j}$, either all the four numbers $p_i(\bar{\theta}_j, \theta_{-i-j}|\bar{\theta}_i)$, $p_i(\bar{\theta}_j, \theta_{-i-j}|\hat{\theta}_i)$, $p_i(\hat{\theta}_j, \theta_{-i-j}|\bar{\theta}_i)$, and $p_i(\hat{\theta}_j, \theta_{-i-j}|\hat{\theta}_i)$ are positive, or they are all equal to zero.

From Assumption 1.5.1, there exists $\tilde{\theta}_{-i-j}$ such that $p_i(\bar{\theta}_j, \tilde{\theta}_{-i-j}|\bar{\theta}_i) > 0$. Hence, expressions (A.17) and (A.18) imply $\frac{a_{\hat{\theta}_i}}{a_{\bar{\theta}_j}} - \frac{c_{\hat{\theta}_i}}{c_{\bar{\theta}_j}}, \frac{a_{\bar{\theta}_i}}{a_{\hat{\theta}_j}} - \frac{c_{\bar{\theta}_i}}{c_{\hat{\theta}_j}} > 0$. Thus for each θ_{-i-j}, either (1) $p_i(\bar{\theta}_j, \theta_{-i-j}|\bar{\theta}_i)$, $p_i(\bar{\theta}_j, \theta_{-i-j}|\hat{\theta}_i)$, and $p_i(\hat{\theta}_j, \theta_{-i-j}|\bar{\theta}_i) > 0$, or (2) $p_i(\bar{\theta}_j, \theta_{-i-j}|\bar{\theta}_i) = p_i(\bar{\theta}_j, \theta_{-i-j}|\hat{\theta}_i) = p_i(\hat{\theta}_j, \theta_{-i-j}|\bar{\theta}_i) = 0$.

If $p_i(\bar{\theta}_j, \theta_{-i-j}|\bar{\theta}_i)$, $p_i(\bar{\theta}_j, \theta_{-i-j}|\hat{\theta}_i)$, and $p_i(\hat{\theta}_j, \theta_{-i-j}|\bar{\theta}_i) > 0$, expression (A.19) implies that $p_i(\hat{\theta}_j, \theta_{-i-j}|\hat{\theta}_i) > 0$.

If $p_i(\bar{\theta}_j, \theta_{-i-j}|\bar{\theta}_i) = p_i(\bar{\theta}_j, \theta_{-i-j}|\hat{\theta}_i) = p_i(\hat{\theta}_j, \theta_{-i-j}|\bar{\theta}_i) = 0$, we must also have $p_i(\hat{\theta}_j, \theta_{-i-j}|\hat{\theta}_i) = 0$. Because otherwise expression (A.19) would imply $\frac{a_{\hat{\theta}_i}}{a_{\hat{\theta}_j}} = \frac{c_{\hat{\theta}_i}}{c_{\hat{\theta}_j}}$, which further means that $p_i(\hat{\theta}_j, \cdot|\bar{\theta}_i) = p_i(\bar{\theta}_j, \cdot|\hat{\theta}_i) = \mathbf{0}$, a contradiction.

Step 2. We want to prove that for all $\theta_{-i-j} \in \Theta_{-i-j}$ such that $p_i(\bar{\theta}_j, \theta_{-i-j}|\bar{\theta}_i) > 0$,

$$\frac{p_i(\bar{\theta}_j, \theta_{-i-j}|\bar{\theta}_i)}{p_i(\hat{\theta}_j, \theta_{-i-j}|\bar{\theta}_i)} = \frac{p_i(\bar{\theta}_j, \theta_{-i-j}|\hat{\theta}_i)}{p_i(\hat{\theta}_j, \theta_{-i-j}|\hat{\theta}_i)}.$$

When $p_i(\bar{\theta}_j, \theta_{-i-j}|\bar{\theta}_i) > 0$, canceling $a_{\hat{\theta}_j}$, $c_{\hat{\theta}_j}$, and $c_{\hat{\theta}_i}$ in expressions (A.16) through (A.19) yields

$$\frac{c_{\bar{\theta}_j} p_i(\hat{\theta}_j, \theta_{-i-j}|\bar{\theta}_i) + p_i(\bar{\theta}_j, \theta_{-i-j}|\bar{\theta}_i)}{c_{\bar{\theta}_j} p_i(\hat{\theta}_j, \theta_{-i-j}|\hat{\theta}_i) + p_i(\bar{\theta}_j, \theta_{-i-j}|\hat{\theta}_i)} = \frac{a_{\hat{\theta}_i} - \frac{p_i(\bar{\theta}_j,\theta_{-i-j}|\bar{\theta}_i)}{p_i(\bar{\theta}_j,\theta_{-i-j}|\hat{\theta}_i)}}{a_{\hat{\theta}_i} - \frac{p_i(\hat{\theta}_j,\theta_{-i-j}|\bar{\theta}_i)}{p_i(\hat{\theta}_j,\theta_{-i-j}|\hat{\theta}_i)}} \times \frac{p_i(\hat{\theta}_j, \theta_{-i-j}|\bar{\theta}_i)}{p_i(\hat{\theta}_j, \theta_{-i-j}|\hat{\theta}_i)}.$$

Suppose $\frac{p_i(\bar{\theta}_j,\theta_{-i-j}|\bar{\theta}_i)}{p_i(\bar{\theta}_j,\theta_{-i-j}|\hat{\theta}_i)} > (<) \frac{p_i(\hat{\theta}_j,\theta_{-i-j}|\bar{\theta}_i)}{p_i(\hat{\theta}_j,\theta_{-i-j}|\hat{\theta}_i)}$. The left-hand side of the above equation is greater (less) than $\frac{p_i(\hat{\theta}_j,\theta_{-i-j}|\bar{\theta}_i)}{p_i(\hat{\theta}_j,\theta_{-i-j}|\hat{\theta}_i)}$ and the right-hand side is less (greater) than $\frac{p_i(\hat{\theta}_j,\theta_{-i-j}|\bar{\theta}_i)}{p_i(\hat{\theta}_j,\theta_{-i-j}|\hat{\theta}_i)}$, a contradiction. Hence, $\frac{p_i(\bar{\theta}_j,\theta_{-i-j}|\bar{\theta}_i)}{p_i(\bar{\theta}_j,\theta_{-i-j}|\hat{\theta}_i)} = \frac{p_i(\hat{\theta}_j,\theta_{-i-j}|\bar{\theta}_i)}{p_i(\hat{\theta}_j,\theta_{-i-j}|\hat{\theta}_i)}$. Rearranging terms yields the desired result.

Step 3. We want to prove that $p_i(\cdot|\bar{\theta}_i) = p_i(\cdot|\hat{\theta}_i)$, which contradicts the BDP property.

Expression (A.16) implies that $\frac{a_{\bar{\theta}_i}}{a_{\theta_j}} = \frac{c_{\bar{\theta}_i}}{c_{\theta_j}}$ for all $\theta_j \neq \hat{\theta}_j$. Plugging it into expression (A.17) yields $(\frac{a_{\hat{\theta}_i}}{a_{\bar{\theta}_i}} - \frac{c_{\hat{\theta}_i}}{c_{\bar{\theta}_i}})p_i(\theta_j, \theta_{-i-j}|\hat{\theta}_i) = \frac{1}{a_{\bar{\theta}_i}}p_i(\theta_j, \theta_{-i-j}|\bar{\theta}_i)$ for all $\theta_j \neq \hat{\theta}_j$ and θ_{-i-j}. Hence,

$$\frac{p_i(\theta_j, \tilde{\theta}_{-i-j}|\bar{\theta}_i)}{p_i(\bar{\theta}_j, \theta_{-i-j}|\bar{\theta}_i)} = \frac{p_i(\theta_j, \tilde{\theta}_{-i-j}|\hat{\theta}_i)}{p_i(\bar{\theta}_j, \theta_{-i-j}|\hat{\theta}_i)}, \forall \theta_j \neq \hat{\theta}_j, \theta_{-i-j} \text{ s.t. } p_i(\bar{\theta}_j, \theta_{-i-j}|\bar{\theta}_i) > 0, \text{ and } \tilde{\theta}_{-i-j}.$$

Combining this expression with Step 1 and Step 2, we have established the desired result.

□

Lemma A.1.6: *Let q be an efficient allocation rule under a private value environment. For any $i \in I$, $\tilde{\Theta}_i \subseteq \Theta_i$ with $|\tilde{\Theta}_i| \geq 2$, and distribution $\pi \in \Delta(\Theta_{-i})$, there exist values $(U_{\theta_i})_{\theta_i \in \tilde{\Theta}_i} \geq \mathbf{0}$ such that $U_{\theta_i} - U_{\theta'_i} \geq \sum_{\theta_{-i} \in \Theta_{-i}}[u_i(q(\theta'_i, \theta_{-i}), \theta_i) - u_i(q(\theta'_i, \theta_{-i}), \theta'_i)]\pi(\theta_{-i})$ for all $\theta_i, \theta'_i \in \tilde{\Theta}_i$.*

Proof. Let a loop be a sequence $(\theta_i^1, \theta_i^2, ..., \theta_i^K)$ in $\tilde{\Theta}_i$ with length $K \geq 2$ and $\theta_i^1 = \theta_i^K$. As q is ex-post efficient, $u_i(q(\theta_i^{k+1}, \theta_{-i}), \theta_i^{k+1}) + \sum_{j \neq i} u_j(q(\theta_i^{k+1}, \theta_{-i}), \theta_j) \geq u_i(q(\theta_i^k, \theta_{-i}), \theta_i^{k+1}) + \sum_{j \neq i} u_j(q(\theta_i^k, \theta_{-i}), \theta_j)$ for all $k = 1, ..., K-1$ and $\theta_{-j} \in \Theta_{-j}$. Summing the inequalities across $k = 1, ..., K-1$, we obtain that $\sum_{k=1}^{K-1}[u_i(q(\theta_i^k, \theta_{-i}), \theta_i^{k+1}) - u_i(q(\theta_i^k, \theta_{-i}), \theta_i^k)] \leq 0$. This is the "cyclical monotonicity" condition is the literature.

Fix an arbitrary $\tilde{\theta}_i \in \tilde{\Theta}_i$. For each $(\theta_i, \theta_{-i}) \in \tilde{\Theta}_i \times \Theta_{-i}$, define the function $V_i(\cdot) : \tilde{\Theta}_i \times \Theta_{-i} \to \mathbb{R}$ by:

$$V_i(\theta_i, \theta_{-i}) \equiv \sup_{\substack{(\theta_i^1, \ldots, \theta_i^k) \text{ is any finite sequence} \\ \text{starting with } \tilde{\theta}_i \text{ and ending with } \theta_i}} \sum_{k=1}^{K-1}[u_i(q(\theta_i^k, \theta_{-i}), \theta_i^{k+1}) - u_i(q(\theta_i^k, \theta_{-i}), \theta_i^k)].$$

Then by Theorem 1 of Rochet (1987) or Proposition 5.2 of Börgers et al. (2015), $V_i(\cdot)$ is a well-defined function satisfying

$$V_i(\theta_i,\theta_{-i}) - V_i(\theta_i',\theta_{-i}) \geq \sum_{\theta_{-i}\in\Theta_{-i}} u_i(q(\theta_i',\theta_{-i}),\theta_i) - u_i(q(\theta_i',\theta_{-i}),\theta_i'), \forall \theta_i,\theta_i' \in \tilde{\Theta}_i.$$

When we choose $C > 0$ sufficiently large, $U_{\theta_i} \equiv \sum_{\theta_{-i}\in\Theta_{-i}} V_i(\theta_i,\theta_{-i})\pi_i(\theta_{-i}) + C \geq 0$ for all $\theta_i \in \tilde{\Theta}_i$. Hence, we have established the desired result. □

Proof of Theorem 1.5.3. Suppose there do not exist agents $i \neq j$ such that the BDP property fails for i and the NCP* property fails for j. Then either of the following is true. Case 1: there are at least $N-1$ agents satisfying both the BDP and NCP* properties. Note by Lemma A.1.5, a special situation in this case is that all agents satisfy the BDP property. Case 2: all agents satisfy the NCP* property.

Case 1. Suppose there are at least $N-1$ agents satisfying both the BDP and NCP* properties. By Lemma A.1.4, there exists $I' \subseteq I$ with $|I'| \geq N-1$ such that for all $i \in I'$ and $\bar{\theta}_i \neq \hat{\theta}_i$, there exists $\psi^{\bar{\theta}_i\hat{\theta}_i} : \Theta \rightarrow \mathbb{R}^N$, such that the three requirements in the lemma are satisfied.

Pick an agent $i \in I$, where $\{i\} = I\backslash I'$ if $I\backslash I'$ is a singleton and $i \in I$ is arbitrary if $I\backslash I' = \emptyset$. As in Theorem 1.5.1, let η be an interim individually rational and ex-post budget-balanced transfer rule such that agent i obtains all the surplus. Define $\Phi = \{\eta\} \cup \{\eta + c\psi^{\bar{\theta}_j\hat{\theta}_j} : j \in I, j \neq i, \bar{\theta}_j, \hat{\theta}_j \in \Theta_j, \bar{\theta}_j \neq \hat{\theta}_j\}$, where c is sufficiently large such that for all $j \neq i$ and $\bar{\theta}_j \neq \hat{\theta}_j$,

$$0 \geq \sum_{\theta_{-j}\in\Theta_{-j}} [u_j(q(\hat{\theta}_j,\theta_{-j}),\bar{\theta}_j) - u_j(q(\hat{\theta}_j,\theta_{-j}),\hat{\theta}_j) + c\psi_j^{\bar{\theta}_j\hat{\theta}_j}(\hat{\theta}_j,\theta_{-j})]p_j(\theta_{-j}|\bar{\theta}_j).$$

For agent $j \neq i$ with type θ_j, truthfully reporting gives her a worst-case expected utility level of zero because the worst transfer rule, η, extracts all her surplus. Thus, j's interim individual rationality condition binds. The choice of c makes misreporting unprofitable. Therefore, her incentive compatibility condition holds.

When all agents truthfully report, a type-$\bar{\theta}_i$ agent i obtains a worst-case expected payoff of

$$\min_{\phi\in\Phi} \sum_{\theta_{-i}\in\Theta_{-i}} [u_i(q(\bar{\theta}_i,\theta_{-i}),\bar{\theta}_i)+\phi_i(\bar{\theta}_i,\theta_{-i})]p_i(\theta_{-i}|\bar{\theta}_i)$$

$$= \sum_{\theta_{-i}\in\Theta_{-i}} [u_i(q(\bar{\theta}_i,\theta_{-i}),\bar{\theta}_i)+\sum_{j\neq i} u_j(q(\bar{\theta}_i,\theta_{-i}),\theta_j)]p_i(\theta_{-i}|\bar{\theta}_i) \geq 0$$

. Hence, agent i's interim individual rationality condition holds. By efficiency of q, this term is weakly higher than $\sum_{\theta_{-i}\in\Theta_{-i}}[u_i(q(\hat{\theta}_i,\theta_{-i}),\bar{\theta}_i)+\sum_{j\neq i} u_j(q(\hat{\theta}_i,\theta_{-i}),\theta_j)]p_i(\theta_{-i}|\bar{\theta}_i)$ for all $\hat{\theta}_i \neq \bar{\theta}_i$. Note the latter expression is weakly higher than the worst-case expected payoff of misreporting $\hat{\theta}_i$, $\min_{\phi\in\Phi}\sum_{\theta_{-i}\in\Theta_{-i}}[u_i(q(\hat{\theta}_i,\theta_{-i}),\bar{\theta}_i)+\phi_i(\hat{\theta}_i,\theta_{-i})]p_i(\theta_{-i}|\bar{\theta}_i)$. Hence, we have also verified agent i's incentive compatibility.

Ex-post budget balance is easy to see. Therefore, the individually rational and budget-balanced mechanism with ambiguous transfers implements q.

Case 2. Suppose all agents satisfy the NCP* property. For any $j \in I$, let $\mathcal{P}_j$ be the partition of Θ_j such that $p_j(\cdot|\theta_j) = p_j(\cdot|\theta'_j)$ if and only if θ_j and θ'_j are in the same $\tilde{\Theta}_j \in \mathcal{P}_j$. For each $\tilde{\Theta}_j$ with $|\tilde{\Theta}_j| \geq 2$ and $\theta_j \in \tilde{\Theta}_j$, define U_{θ_j} according to Lemma A.1.6. For a singleton $\tilde{\Theta}_j \in \mathcal{P}_j$ and $\{\theta_j\} = \tilde{\Theta}_j$, define $U_{\theta_j} = 0$.

We will demonstrate that for each i and $\bar{\theta}_i \neq \hat{\theta}_i$, the following system has a solution

$\phi^{\bar{\theta}_i\hat{\theta}_i}$.

$$\sum_{\theta_{-j}\in\Theta_{-j}} \phi_i^{\bar{\theta}_i\hat{\theta}_i}(\bar{\theta}_i,\theta_{-i})p_i(\theta_{-i}|\bar{\theta}_i) = U_{\bar{\theta}_i} - \sum_{\theta_{-i}\in\Theta_{-i}} u_i(q(\bar{\theta}_i,\theta_{-i}),\theta_i)p_i(\theta_{-i}|\bar{\theta}_i),$$

$$\sum_{\theta_{-j}\in\Theta_{-j}} \phi_j^{\bar{\theta}_i\hat{\theta}_i}(\theta_j,\theta_{-j})p_j(\theta_{-j}|\theta_j) \geq U_{\theta_j} - \sum_{\theta_{-j}\in\Theta_{-j}} u_j(q(\theta_j,\theta_{-j}),\theta_j)p_j(\theta_{-j}|\theta_j),$$

$$\forall (j,\theta_j) \neq (i,\bar{\theta}_i),$$

$$-\sum_{j\in I}\phi_j^{\bar{\theta}_i\hat{\theta}_i}(\theta) = 0, \forall\theta\in\Theta,$$

$$-\sum_{\theta_{-i}\in\Theta_{-i}} \phi_i^{\bar{\theta}_i\hat{\theta}_i}(\hat{\theta}_i,\theta_{-i})p_i(\theta_{-i}|\bar{\theta}_i) \geq -U_{\bar{\theta}_i} + \sum_{\theta_{-i}\in\Theta_{-i}} u_i(q(\hat{\theta}_i,\theta_{-i}),\bar{\theta}_i)p_i(\theta_{-i}|\bar{\theta}_i).$$

Suppose by way of contradiction that the system does not have a solution. By a theorem of the alternative, there exist coefficients $a_{\bar{\theta}_i}$, $(a_{\theta_j})_{(j,\theta_j)\neq(i,\bar{\theta}_i)} \geq \mathbf{0}$, $(b_\theta)_{\theta\in\Theta}$, and $\gamma_{\bar{\theta}_i\hat{\theta}_i} \geq 0$ that are not all zero, such that the weighted sum of the left-hand sides of the expressions is cancelled and the weighted sum of the right-hand sides is positive.

Suppose $\gamma_{\bar{\theta}_i\hat{\theta}_i} = 0$. Following Lemma A.1.4, we know $(a_{\theta_j})_{j\in I,\theta_j\in\Theta_j} > \mathbf{0}$ and $(b_\theta)_{\theta\in\Theta} \gneqq \mathbf{0}$. Define $\mu(\theta) = \frac{b_\theta}{\sum_{\tilde{\theta}\in\Theta} b_{\tilde{\theta}}}$ for all θ, which is a common prior, contradicting the assumption that beliefs are not generated from a common prior.

Suppose $\gamma_{\bar{\theta}_i\hat{\theta}_i} > 0$. From Lemma A.1.4 and that the NCP* property holds for all agents, we know: (1) $p_i(\cdot|\bar{\theta}_i) = p_i(\cdot|\hat{\theta}_i)$, and (2) among all the coefficients, $a_{\hat{\theta}_i} = \gamma_{\bar{\theta}_i\hat{\theta}_i} > 0$ and everything else is zero. According to Lemma A.1.6, the choice of $U_{\bar{\theta}_i}$ and $U_{\hat{\theta}_i}$ satisfies $U_{\hat{\theta}_i} - U_{\bar{\theta}_i} + \sum_{\theta_{-i}\in\Theta_{-i}}[u_i(q(\hat{\theta}_i,\theta_{-i}),\bar{\theta}_i) - u_i(q(\hat{\theta}_i,\theta_{-i}),\hat{\theta}_i)]p_i(\theta_{-i}|\bar{\theta}_i) \leq 0$. Hence, the weighted sum of the right-hand sides is non-positive, a contradiction.

Therefore, for each i, $\bar{\theta}_i \neq \hat{\theta}_i$, the system has a solution $\phi^{\bar{\theta}_i\hat{\theta}_i}$. Let the set of ambiguous transfers be $\Phi = \{\phi^{\bar{\theta}_i\hat{\theta}_i}, \forall i, \bar{\theta}_i, \hat{\theta}_i \in \Theta_i, \bar{\theta}_i \neq \hat{\theta}_i\}$. The interim individually rational and ex-post budget-balanced mechanism with ambiguous transfers implements q. □

A.2 Including Agents without Private Information

In this section, we relax the assumption that $|\Theta_i| \geq 2$ for all $i \in I$. Denote the set of all agents with at least two types by $\tilde{I}$, which has a cardinality of $\tilde{N}$. As an agent in $I \backslash \tilde{I}$ has only one type, she cannot lie. We claim that all theorems of this paper hold if $\tilde{N} \geq 2$, i.e., at least two agents have private information.

To see why including agents without private information may be interesting, consider two consumers with unknown values paying for producing a costly public project. In this example $\tilde{I} = \{1, 2\}$ and $I = \{1, 2, 3\}$, where 3 is interpreted as a producer whose payoff (profit) is the payments of 1 and 2 minus the cost of production. By efficiency and budget balance, two consumers' aggregated utility from the project minus the cost of production should be maximized.

We demonstrate the modification needed for Theorem 1.4.2 as an example. In Lemmas A.1.1 through A.1.3, we replace all I with $\tilde{I}$ and all N with $\tilde{N}$. Then we extend the transfer rule ψ to include agents $I \backslash \tilde{I}$ by letting $\psi_i(\theta) = 0$ for all $i \in I \backslash \tilde{I}$ and $\theta \in \Theta$. Let η be a transfer rule that is interim individually rational and ex-post budget balanced across every agent $i \in I$. Then one can follow Theorem 1.4.2 to construct ambiguous transfers. Incentive compatibility of agents in $\tilde{I}$ is achieved in the same way as the original proof. We obtain incentive compatibility of all other agents for free as each of them has only one type. Individual rationality and budget balance follow from the respective properties of η and ψ.

APPENDIX B
APPENDIX TO CHAPTER 2

B.1 Proofs

This Appendix establishes the relationship between the conditions for interim coalitional implementation and those for robust coalitional implementation.

Proposition B.1.1: *Given a payoff environment Θ, a social choice function f is robust coalitional incentive compatible if and only if it is interim coalitional incentive compatible in all type spaces with payoff environment Θ.*

Proof. We prove the "only if" part of the proposition first. Let f be a robust coalitional incentive compatible social choice function. Suppose by way of contradiction that there exists a type space $\mathcal{T}$, $S \in \mathcal{S}$, $t_S^* \in T_S$, and $\alpha_S \colon T_S \to T_S$ such that for all $i \in S$,

$$\sum_{t_{-i}\in T_{-i}} u_i\Big(f\big(\hat{\theta}(\alpha_S(t_S^*), t_{-S})\big), \hat{\theta}(t_S^*, t_{-S})\Big)\pi_i(t_i^*)[t_{-i}|t^*_{S\setminus\{i\}}]$$

$$> \sum_{t_{-i}\in T_{-i}} u_i\Big(f\big(\hat{\theta}(t_S^*, t_{-S})\big), \hat{\theta}(t_S^*, t_{-S})\Big)\pi_i(t_S^*)[t_{-i}|t^*_{S\setminus\{i\}}].$$

For all $i \in S$, let $\theta_i^* = \hat{\theta}_i(t_i^*)$ and $\theta_i' = \hat{\theta}_i(\alpha_i(t_i^*))$. The above inequality shows that for all $i \in S$, there exists θ^i_{-S} such that $u_i\big(f(\theta'_S, \theta^i_{-S}), (\theta^*_S, \theta^i_{-S})\big) > u_i\big(f(\theta^*_S, \theta^i_{-S}), (\theta^*_S, \theta^i_{-S})\big)$, contradicting the robust coalitional incentive compatibility condition.

To prove the "if" part, suppose that f does not satisfy the robust coalitional incentive compatibility condition, i.e., there exists $S \in \mathcal{S}$, and $\theta^*_S, \theta'_S \in \Theta_S$ such that for all $i \in S$, there exists $\theta^i_{-S} \in \Theta_{-S}$ such that $u_i\big(f(\theta'_S, \theta^i_{-S}), (\theta^*_S, \theta^i_{-S})\big) > u_i\big(f(\theta^*_S, \theta^i_{-S}), (\theta^*_S, \theta^i_{-S})\big)$. Then we let $\mathcal{T}$ be any payoff type space satisfying the following restriction: for all $i \in S$

and $t_i^* \in T_i$ satisfying $\hat{\theta}_i(t_i^*) = \theta_i^*$, $\pi_i(t_i^*)[\cdot]$ puts weight 1 on the type profile t_{-i} with payoff type profile $(\theta^*_{S\backslash\{i\}}, \theta^i_{-S})$. For each $i \in S$, let t_i' be the type with payoff type θ_i', and $\alpha_i : T_i \rightarrow T_i$ be the identical mapping except that $\alpha_i(t_i^*) = t_i'$. In the type space $\mathcal{T}$, for all $i \in S$,

$$\sum_{t_{-i}\in T_{-i}} u_i\Big(f\big(\hat{\theta}(t'_S, t_{-S})\big), \hat{\theta}(t^*_S, t_{-S})\Big)\pi_i(t_i^*)[t_{-i}|t^*_{S\backslash\{i\}}]$$

$$> \sum_{t_{-i}\in T_{-i}} u_i\Big(f\big(\hat{\theta}(t^*_S, t_{-S})\big), \hat{\theta}(t^*_S, t_{-S})\Big)\pi_i(t_i^*)[t_{-i}|t^*_{S\backslash\{i\}}].$$

Therefore, f is not interim coalitional incentive compatible in $\mathcal{T}$, a contradiction. □

In order to establish the equivalence between the interim coalitional monotonicity condition under all type spaces and the robust coalitional monotonicity condition, we begin with several auxiliary definitions and as well as one auxiliary equivalence relationship.

Definition B.1.1: *Given a type space $\mathcal{T}$ and a coalition $S \in \mathcal{S}$, the social choice function f satisfies the S* ***interim coalitional monotonicity condition*** *if whenever α is unacceptable at t^*, there exists $h \in H_S^{f,\alpha(t^*)}$ such that for all $i \in S$,*

$$\sum_{t_{-i}\in T_{-i}} u_i\Big(h\big(\alpha(t^*_S, t_{-S})\big), \hat{\theta}(t^*_S, t_{-S})\Big)\pi_i(t_i^*)[t_{-i}|t^*_{S\backslash\{i\}}]$$

$$> \sum_{t_{-i}\in T_{-i}} u_i\Big(f\big(\hat{\theta}\big(\alpha(t^*_S, t_{-S})\big)\big), \hat{\theta}(t^*_S, t_{-S})\Big)\pi_i(t_i^*)[t_{-i}|t^*_{S\backslash\{i\}}].$$

Definition B.1.2: *Given a coalition $S \in \mathcal{S}$, a social choice function f satisfies the S* ***robust coalitional monotonicity*** *condition if whenever the deception profile $\boldsymbol{\beta}$ is unacceptable at the pair (θ^*, θ'), for any conjectures and distributions $\big(\theta'^i_{-S} \in \boldsymbol{\beta}_{-S}(\Theta_{-S}), \psi_i(\cdot) \in \Delta(\boldsymbol{\beta}_{-S}(\theta'^i_{-S}))\big)_{i\in S}$, there exists $y \in Y_S^{f,\theta'}$ such that for all $i \in S$,*

$$\sum_{\theta_{-S}\in\boldsymbol{\beta}_{-S}(\theta'^{i}_{-S})} u_i\big(y(\theta'_S,\theta'^{i}_{-S}),(\theta^*_S,\theta_{-S})\big)\psi_i(\theta_{-S})$$

$$> \sum_{\theta_{-S}\in\boldsymbol{\beta}_{-S}(\theta'^{i}_{-S})} u_i\big(f(\theta'_S,\theta'^{i}_{-S}),(\theta^*_S,\theta_{-S})\big)\psi_i(\theta_{-S}).$$

Lemma B.1.1: *Given a coalition $S \in \mathcal{S}$ and a payoff environment Θ, a social choice function f is S robust coalitional monotonic if and only if it is S interim coalitional monotonic in all type spaces with payoff environment Θ.*

Proof. We begin with proving the "if" part of the equivalence relation. Let f satisfy the S interim coalitional monotonicity condition in all type spaces, but suppose by way of contradiction that the S robust coalitional monotonicity condition fails. Then there exists an unacceptable deception profile $\boldsymbol{\beta}$ at the pair (θ^*,θ') and $\big(\theta'^{i}_{-S} \in \boldsymbol{\beta}_{-S}(\Theta_{-S}), \psi_i(\cdot) \in \Delta(\boldsymbol{\beta}^{-1}_{-S}(\theta'^{i}_{-S}))\big)_{i\in S}$, such that whenever $y \in Y^f_{\theta'_S}$, there exists some $i \in S$ such that

$$\sum_{\theta_{-S}\in\boldsymbol{\beta}^{-1}_{-S}(\theta'^{i}_{-S})} u_i\big(y(\theta'_S,\theta'^{i}_{-S}),(\theta^*_S,\theta_{-S})\big)\psi_i(\theta_{-S})$$

$$\leq \sum_{\theta_{-S}\in\boldsymbol{\beta}^{-1}_{-S}(\theta'^{i}_{-S})} u_i\big(f(\theta'_S,\theta'^{i}_{-S}),(\theta^*_S,\theta_{-S})\big)\psi_i(\theta_{-S}). \quad \text{(B.1)}$$

The proof proceeds as follows. Firstly, we construct a type space $\mathcal{T}$, where $T_i = T^1_i \cup T^2_i$ for all $i \in I$. Secondly, we define an unacceptable deception profile $\alpha : T \to T$. Thirdly, a contradiction is reached.

For the above-mentioned $\theta^* \in \Theta$, $\theta' \in \boldsymbol{\beta}(\theta)$, and $(\theta'^{i}_{-S}, \psi_i(\cdot))_{i\in S}$ such that the above statement is satisfied. For each $i \in I$, we will construct a type set T^1_i.

For all $j \notin S$, there is a bijection $\xi^1_j : T^1_j \to \{(\theta_j, \theta''_j) | \theta_j \in \Theta_j, \theta''_j \in \boldsymbol{\beta}_j(\theta_j)\}$. For all $i \in S$, T^1_i has only one element. For each $j \notin S$ and t_j with $\xi^1_j(t_j) = (\theta_j, \theta''_j)$, let t_j

has payoff type θ_j and a full support belief type over T^1_{-j}. For all $i \in S$, let $t_i \in T^1_i$ has payoff type $\hat{\theta}_i(t_i) = \theta^*_i$. Its belief type satisfies $\pi_i(t_i)[t_{-i}] = \psi_i(\theta_{-S})$ if $t_{-i} \in T^1_{-i}$ and $\xi^1_{-S}(t_{-S}) = (\theta_{-S}, \theta'^i_{-S})$.

For each $i \in I$, construct another type set T^2_i below.

For all $i \in I$, a type set T^2_i is bijection to Θ under $\xi^2_i : T^2_i \to \Theta_i \times \prod_{j \notin S} \Theta_j$. For each $t_i \in T^2_i$ with $\xi^2_i(t_i) = (\theta_i, (\theta_j)_{j \notin S})$, let $\hat{\theta}_i(t_i) = \theta_i$ and $\pi_i(t_i)[t_{-i}]$ be any distribution over T^2_{-i} such that the margin on the event that $(t_j)_{j \notin S}$ has payoff type profile $(\theta_j)_{j \notin S}$ equals 1, and the margin has full support over $T^2_{S \setminus \{i\}}$.

Fix any $\bar{\theta} \in \Theta$. Let a deception profile $\alpha : T \to T$ be:

$$\alpha_i(t_i) = \begin{cases} [\xi^2_i]^{-1}(\theta'_i, \bar{\theta}_{-i}) & \text{if } i \in S \text{ and } t_i \in T^1_i, \\ [\xi^2_i]^{-1}(\theta''_i, \bar{\theta}_{-i}) & \text{if } i \notin S \text{ and } t_i = [\xi^1_i]^{-1}(\theta_i, \theta''_i) \in T^1_i, \\ t_i & \text{elsewhere.} \end{cases}$$

It is easy to see the deception profile α is unacceptable at the type profile t^* such that $\xi^1_i(t^*_i) = (\theta^*_i, \theta'_i)$ for all $i \in S$ and $\xi^1_j(t_j) = (\theta^*_j, \theta'_j)$ for all $j \in S$.

Therefore, there exists $h \in H^{f,\alpha(t^*)}_S$ such that the strict inequality in Definition B.1.1 holds for all $i \in S$. For all $\theta \in \Theta$, define $y(\theta) = h\Big(\alpha_S(t^*_S), \big([\xi^2_j]^{-1}(\theta_j, \bar{\theta}_{-j})\big)_{j \notin S}\Big)$. By letting t''_S go over all type profiles in T^2_S, one can verify that $y \in Y^{f,\theta'}_S$. The strict inequalities in Definition B.1.1 imply that for all $i \in S$,

$$\sum_{\theta_{-S} \in \boldsymbol{\beta}^{-1}_{-S}(\theta'^i_{-S})} u_i\big(y(\theta'_S, \theta'^i_{-S}), (\theta_S, \theta_{-S})\big)\psi_i(\theta_{-S})$$

$$> \sum_{\theta_{-S} \in \boldsymbol{\beta}^{-1}_{-S}(\theta'^i_{-S})} u_i\big(f(\theta'_S, \theta'^i_{-S}), (\theta_S, \theta_{-S})\big)\psi_i(\theta_{-S}),$$

a contradiction to expression (B.1).

Now we prove the “only if” half of the equivalence relation. Let $\mathcal{T}$ be an arbitrary type space with payoff environment Θ. Let $\alpha : T \to T$ be an unacceptable deception profile at t^*. Define a correspondence $\boldsymbol{\beta} : \Theta \to 2^{\Theta} \backslash \emptyset$ by $\boldsymbol{\beta}(\theta) = \cup_{\{t \in T | \hat{\theta}(t) = \theta\}} \alpha(t)$ for all $\theta \in \Theta$. Let $\theta^* = \hat{\theta}(t^*)$ and $\theta' = \hat{\theta}(\alpha(t^*))$. From the supposition, $\boldsymbol{\beta}$ is not acceptable at the pair (θ^*, θ').

Suppose the social choice function f satisfies the S robust coalitional monotonicity condition. Then we know for any conjectures and distributions $\big(\theta'^{i}_{-S} \in \boldsymbol{\beta}_{-S}(\Theta_{-S}), \psi_i(\cdot) \in \Delta(\boldsymbol{\beta}_{-S}(\theta'^{i}_{-S}))\big)_{i \in S}$, there exists $y \in Y_S^{f,\theta'}$ such that for all $i \in S$,

$$\sum_{\theta_{-S} \in \boldsymbol{\beta}_{-S}(\theta'^{i}_{-S})} u_i\big(y(\theta'_S, \theta'^{i}_{-S}), (\theta^*_S, \theta_{-S})\big)\psi_i(\theta_{-S})$$

$$> \sum_{\theta_{-S} \in \boldsymbol{\beta}_{-S}(\theta'^{i}_{-S})} u_i\big(f(\theta'_S, \theta'^{i}_{-S}), (\theta^*_S, \theta_{-S})\big)\psi_i(\theta_{-S}).$$

For all $i \in S$ and $\theta'^{i}_{-S} \in \Theta_{-S}$, let $\psi_i(\cdot) \in \Delta(\boldsymbol{\beta}_{-S}(\theta'^{i}_{-S}))$ be

$$\psi_i(\theta_{-S}) = \sum_{\substack{\{t_{-S}: \hat{\theta}_{-S}(t_{-S}) = \theta_{-S}, \\ \hat{\theta}_{-S}\left(\alpha_{-S}(t_{-S})\right) = \theta'^{i}_{-S}\}}} \pi_i(t_i^*)[t_{-i} | t^*_{S \backslash \{i\}}] \quad / \quad \sum_{\{t_{-S}: \hat{\theta}_{-S}\left(\alpha_{-S}(t_{-S})\right) = \theta'^{i}_{-S}\}} \pi_i(t_i^*)[t_{-i} | t^*_{S \backslash \{i\}}]$$

for all $\theta_{-S} \in \Theta_{-S}$ when the denominator is nonzero, and let ψ_i be any distribution $\psi_i(\cdot) \in \Delta(\boldsymbol{\beta}_{-S}(\theta'^{i}_{-S}))$ when the denominator is zero. For all $t \in T$, let $h(t) = y\big(\theta'_S, \hat{\theta}_{-S}(t_{-S})\big)$. Then,

$$\sum_{t_{-i} \in T_{-i}} u_i\Big(h\big(\alpha(t^*_S, t_{-S})\big), \hat{\theta}(t^*_S, t_{-S})\Big)\pi_i(t_i^*)[t_{-i} | t^*_{S \backslash \{i\}}]$$

$$= \sum_{\substack{\theta'^{i}_{-S} \in \\ \boldsymbol{\beta}_{-S}(\Theta_{-S})}} \Bigg(\sum_{\substack{\theta_{-S} \in \\ \boldsymbol{\beta}^{-1}_{-S}(\theta'^{i}_{-S})}} u_i\big(y(\theta'_S, \theta'^{i}_{-S}), (\theta^*_S, \theta_{-S})\big)\psi_i(\theta_{-S})\Bigg)\Bigg(\sum_{\{t_{-S}: \hat{\theta}_{-S}\left(\alpha_{-S}(t_{-S})\right) = \theta'^{i}_{-S}\}} \pi_i(t_i^*)[t_{-i} | t^*_{S \backslash \{i\}}]\Bigg)$$

$$> \sum_{\substack{\theta'^{i}_{-S} \in \\ \boldsymbol{\beta}_{-S}(\Theta_{-S})}} \Bigg(\sum_{\substack{\theta_{-S} \in \\ \boldsymbol{\beta}^{-1}_{-S}(\theta'^{i}_{-S})}} u_i\big(f(\theta'_S, \theta'^{i}_{-S}), (\theta^*_S, \theta_{-S})\big)\psi_i(\theta_{-S})\Bigg)\Bigg(\sum_{\{t_{-S}: \hat{\theta}_{-S}\left(\alpha_{-S}(t_{-S})\right) = \theta'^{i}_{-S}\}} \pi_i(t_i^*)[t_{-i} | t^*_{S \backslash \{i\}}]\Bigg)$$

$$= \sum_{t_{-i} \in T_{-i}} u_i\Bigg(f\Big(\hat{\theta}\big(\alpha(t^*_S, t_{-S})\big)\Big), \hat{\theta}(t^*_S, t_{-S})\Bigg)\pi_i(t_i^*)[t_{-i} | t^*_{S \backslash \{i\}}]$$

for all $i \in S$, where the inequality follows from the S robust coalitional monotonicity condition. It is straightforward to see $h \in H_S^{f,\alpha(t^*)}$ from $y \in Y_S^{f,\theta'}$. Hence, we have established the S interim coalitional monotonicity condition in $\mathcal{T}$. As $\mathcal{T}$ is arbitrary, we have proved that the S interim coalitional monotonicity condition holds in all type spaces with payoff environment Θ. □

Proposition B.1.2: *A social choice function f is robust coalitional monotonic if and only if it is interim coalitional monotonic in all type spaces with payoff environment Θ.*

Proof. The social choice function f satisfies the interim (robust) coalitional monotonicity condition if and only if there exists $S \in \mathcal{S}$ such that f satisfies the S interim (robust) coalitional monotonicity condition. Then by applying Lemma B.1.1, we can prove the desired result. □

REFERENCES

Angelopoulos, Angelos and Leonidas C Koutsougeras, "Value allocation under ambiguity," *Economic Theory*, 2015, *59* (1), 147–167.

Aoyagi, Masaki, "Correlated types and Bayesian incentive compatible mechanisms with budget balance," *Journal of Economic Theory*, 1998, *79* (1), 142–151.

Aumann, Robert J, "Acceptable points in general cooperative n-person games," *Contributions to the Theory of Games (AM-40)*, 1959, *4*, 287–324.

_ , "Agreeing to disagree," *The Annals of Statistics*, 1976, *4* (6), 1236–1239.

Ausubel, Lawrence M, Paul Milgrom et al., "The lovely but lonely Vickrey auction," *Combinatorial auctions*, 2006, *17*, 22–26.

Barelli, Paulo, "On the genericity of full surplus extraction in mechanism design," *Journal of Economic Theory*, 2009, *144* (3), 1320–1332.

Bergemann, Dirk and Stephen Morris, "Robust implementation: The role of large type spaces," 2005. Working paper.

_ and _ , "Robust mechanism design," *Econometrica*, 2005, *73* (6), 1771–1813.

_ and _ , "Robust implementation in direct mechanisms," *The Review of Economic Studies*, 2009, *76* (4), 1175–1204.

_ and _ , "Robust implementation in general mechanisms," *Games and Economic Behavior*, 2011, *71* (2), 261–281.

_ , _ , and Satoru Takahashi, "Efficient auctions and interdependent types," *American Economic Review: Papers and Preceedings*, 2012, *102* (3), 319–324.

Bodoh-Creed, Aaron L, "Ambiguous beliefs and mechanism design," *Games and Economic Behavior*, 2012, *75* (2), 518–537.

Börgers, Tilman, Daniel Krähmer, and Roland Strausz, *An introduction to the theory of mechanism design*, Oxford University Press, USA, 2015.

Borghans, Lex, James J Heckman, Bart HH Golsteyn, and Huub Meijers, "Gender differences in risk aversion and ambiguity aversion," *Journal of the European Economic Association*, 2009, *7* (2-3), 649–658.

Bose, Subir and Arup Daripa, "A dynamic mechanism and surplus extraction under ambiguity," *Journal of Economic Theory*, 2009, *144* (5), 2084–2114.

_ and Ludovic Renou, "Mechanism design with ambiguous communication devices," *Econometrica*, 2014, *82* (5), 1853–1872.

_ **, Emre Ozdenoren, and Andreas Pape**, "Optimal auctions with ambiguity," *Theoretical Economics*, 2006, *1* (4), 411–438.

Chen, Yi-Chun and Siyang Xiong, "The genericity of beliefs-determine-preferences models revisited," *Journal of Economic Theory*, 2011, *146* (2), 751–761.

_ **and** _ , "Genericity and robustness of full surplus extraction," *Econometrica*, 2013, *81* (2), 825–847.

Chen, Zengjing and Larry Epstein, "Ambiguity, risk, and asset returns in continuous time," *Econometrica*, 2002, *70* (4), 1403–1443.

Chung, Kim-Sau, "A note on Matsushima's regularity condition," *Journal of Economic Theory*, 1999, *87* (2), 429–433.

Clarke, Edward H, "Multipart pricing of public goods," *Public Choice*, 1971, *11* (1), 17–33.

Crémer, J and Richard P McLean, "Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent," *Econometrica*, 1985, *53* (2), 345–361.

Crémer, Jacques and Richard P McLean, "Full extraction of the surplus in Bayesian and dominant strategy auctions," *Econometrica*, 1988, *56* (6), 1247–1257.

Dasgupta, Partha and Eric Maskin, "Efficient auctions," *The Quarterly Journal of Economics*, 2000, *115* (2), 341–388.

d'Aspremont, Claude and Louis-André Gérard-Varet, "Incentives and incomplete information," *Journal of Public Economics*, 1979, *11* (1), 25–45.

_ **, Jacques Crémer, and Louis-André Gérard-Varet**, "Balanced Bayesian mechanisms," *Journal of Economic Theory*, 2004, *115* (2), 385–396.

de Castro, Luciano I, Marialaura Pesce and Nicholas C Yannelis, "Core and equilibria under ambiguity," *Economic Theory*, 2011 *48*, 519—548.

_ **and Nicholas C Yannelis**, "Uncertainty, Efficiency and Incentive Compatibility," *Journal of Economic Theory*, 2018. Forthcoming.

_ **, Zhiwei Liu, and Nicholas C Yannelis**, "Ambiguous implementation: the partition model," *Economic Theory*, 2017, *63* (1), 233–261.

_ **,** _ **, and** _ , "Implementation under ambiguity," *Games and Economic Behavior*, 2017, *101*, 20–33.

Di Tillio, Alfredo, Nenad Kos, and Matthias Messner, "The design of ambiguous mechanisms," *The Review of Economic Studies*, 2017, *84* (1), 237–274.

Dutta, Bhaskar and Arunava Sen, "Implementation under strong equilibrium: A complete characterization," *Journal of Mathematical Economics*, 1991, *20* (1), 49–67.

Ellsberg, Daniel, "Risk, ambiguity, and the Savage axioms," *The Quarterly Journal of Economics*, 1961, pp. 643–669.

Epstein, Larry G and Tan Wang, ""Beliefs about beliefs" without probabilities," *Econometrica*, 1996, pp. 1343–1373.

Fox, Craig R and Amos Tversky, "Ambiguity aversion and comparative ignorance," *The Quarterly Journal of Economics*, 1995, *110* (3), 585–603.

Garlappi, Lorenzo, Raman Uppal, and Tan Wang, "Portfolio selection with parameter and model uncertainty: A multi-prior approach," *The Review of Financial Studies*, 2006, *20* (1), 41–81.

Ghirardato, Paolo and Massimo Marinacci, "Ambiguity made precise: A comparative foundation," *Journal of Economic Theory*, 2002, *102* (2), 251–289.

Gilboa, Itzhak and David Schmeidler, "Maxmin expected utility with non-unique prior," *Journal of Mathematical Economics*, 1989, *18* (2), 141–153.

Gizatulina, Alia and Martin Hellwig, "Informational smallness and the scope for limiting information rents," *Journal of Economic Theory*, 2010, *145* (6), 2260–2281.

_ **and** _ , "Beliefs, payoffs, information: On the robustness of the BDP property in models with endogenous beliefs," *Journal of Mathematical Economics*, 2014, *51*, 136–153.

_ **and** _ , "The generic possibility of full surplus extraction in models with large type spaces," *Journal of Economic Theory*, 2017, *170* (7), 385-416.

Groves, Theodore, "Incentives in teams," *Econometrica*, 1973, *41* (4), 617–631.

Hahn, Guangsug and Nicholas C Yannelis, "Coalitional Bayesian Nash implementation in differential information economies," *Economic Theory*, 2001, *18* (2), 485–509.

Hansen, Lars Peter and Thomas J Sargent, "Robust control and model uncertainty," *The American Economic Review*, 2001, *91* (2), 60–66.

_ **and** _ , *Robustness*, Princeton university press, 2008.

Harsanyi, John C, "Games with incomplete information played by "Bayesian" players, I–III Part I. The basic model," *Management Science*, 1967, *14* (3), 159—182.

Heifetz, Aviad and Zvika Neeman, "On the generic (im) possibility of full surplus extraction in mechanism design," *Econometrica*, 2006, *74* (1), 213–233.

Jackson, Mathew O, "Bayesian implementation," *Econometrica*, 1991, pp. 461–477.

Jehiel, Philippe and Benny Moldovanu, "Efficient design with interdependent valuations," *Econometrica*, 2001, *69* (5), 1237–1259.

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji, "A smooth model of decision making under ambiguity," *Econometrica*, 2005, *73* (6), 1849–1892.

Korpela, Ville, "A simple sufficient condition for strong implementation," *Journal of Economic Theory*, 2013, *148* (5), 2183–2193.

Kosenok, Grigory and Sergei Severinov, "Individually rational, budget-balanced mechanisms and allocation of surplus," *Journal of Economic Theory*, 2008, *140* (1), 126–161.

Liu, Heng, "Efficient dynamic mechanisms in environments with interdependent valuations: the role of contingent transfers," *Theoretical Economics*, 2018. Forthcoming.

Maskin, Eric, "Implementation and strong Nash equilibrium," 1978. Working paper.

— , "Incentive schemes immune to group manipulation," Technical Report, Working paper 1979.

— , "Nash equilibrium and welfare optimality," *The Review of Economic Studies*, 1999, *66* (1), 23–38.

Matsushima, Hitoshi, "Incentive compatible mechanisms with full transferability," *Journal of Economic Theory*, 1991, *54* (1), 198–203.

— , "Mechanism design with side payments: Individual rationality and iterative dominance," *Journal of Economic Theory*, 2007, *133* (1), 1–30.

McAfee, R Preston and Philip J Reny, "Correlated information and mechanism design," *Econometrica*, 1992, *60* (2), 395–421.

McLean, Richard P and Andrew Postlewaite, "Informational size and incentive compatibility," *Econometrica*, 2002, *70* (6), 2421–2453.

— **and** — , "Informational size and incentive compatibility with aggregate uncertainty," *Games and Economic Behavior*, 2003, *45* (2), 410–433.

— **and** — , "Informational size, incentive compatibility, and the core of a game with incomplete information," *Games and Economic Behavior*, 2003, *45* (1), 222–241.

— **and** — , "Informational size and efficient auctions," *The Review of Economic Studies*, 2004, *71* (3), 809–827.

— **and** — , "Implementation with interdependent valuations," *Theoretical Economics*, 2015, *10* (3), 923–952.

Miller, Nolan H, John W Pratt, Richard J Zeckhauser, and Scott Johnson, "Mechanism design with multidimensional, continuous types and interdependent valuations," *Journal of Economic Theory*, 2007, *136* (1), 476–496.

Morris, Stephen, "The common prior assumption in economic theory," *Economics and Philosophy*, 1995, *11* (2), 227–253.

Moulin, Herve and Bezalel Peleg, "Cores of effectivity functions and implementation theory," *Journal of Mathematical Economics*, 1982, *10* (1), 115–145.

Müller, Christoph, "Robust virtual implementation under common strong belief in rationality," *Journal of Economic Theory*, 2016, *162*, 407–450.

Myerson, Roger B and Mark A Satterthwaite, "Efficient mechanisms for bilateral trading," *Journal of Economic Theory*, 1983, *29* (2), 265–281.

Neeman, Zvika, "The relevance of private information in mechanism design," *Journal of Economic Theory*, 2004, *117* (1), 55–77.

Noda, Shunya, "Full surplus extraction and costless information revelation in dynamic environments," *Theoretical Economics*, 2018. Forthcoming.

Ollár, Mariann and Antonio Penta, "Full implementation and belief restrictions," *American Economic Review*, 2017, *107* (8), 2243–77.

Palfrey, Thomas R and Sanjay Srivastava, "On Bayesian implementable allocations," *The Review of Economic Studies*, 1987, *54* (2), 193–208.

_ and _ , "Implementation with incomplete information in exchange economies," *Econometrica*, 1989, pp. 115–134.

Pasin, Pelin, "Essays on implementability and monotonicity." PhD dissertation, Bilkent University 2009.

Penta, Antonio, "Robust dynamic implementation," *Journal of Economic Theory*, 2015, *160*, 280–316.

Postlewaite, Andrew and David Schmeidler, "Implementation in differential information economies," *Journal of Economic Theory*, 1986, *39* (1), 14–33.

Repullo, Rafael, "A simple proof of Maskin's theorem on Nash implementation," *Social Choice and Welfare*, 1987, *4* (1), 39–41.

Rochet, Jean-Charles, "A necessary and sufficient condition for rationalizability in a quasi-linear context," *Journal of Mathematical Economics*, 1987, *16* (2), 191–200.

Rothkopf, Michael H, "Thirteen reasons why the Vickrey-Clarke-Groves process is not practical," *Operations Research*, 2007, *55* (2), 191–197.

Saijo, Tatsuyoshi, "Strategy space reduction in Maskin's theorem: sufficient conditions for Nash implementation," *Econometrica*, 1988, pp. 693–700.

Smith, Doug, “A prior free efficiency comparison of mechanisms for the public good problem,” 2010. Working paper.

Song, Yangwei, “Efficient implementation with interdependent valuations and maxmin agents,” *Journal of Economic Theory*, 2018. Forthcoming.

Suh, Sang-Chul, “Implementation with coalition formation: A complete characterization,” *Journal of Mathematical Economics*, 1996, *26* (4), 409–428.

_ , “Double implementation in Nash and strong Nash equilibria,” *Social Choice and Welfare*, 1997, *14* (3), 439–447.

Sun, Yeneng and Nicholas C Yannelis, “Perfect competition in asymmetric information economies: compatibility of efficiency and incentives,” *Journal of Economic Theory*, 2007, *134* (1), 175–194.

_ **and** _ , “Ex ante efficiency implies incentive compatibility,” *Economic Theory*, 2008, *36* (1), 35–55.

Vickrey, William, “Counterspeculation, auctions, and competitive sealed tenders,” *Journal of Finance*, 1961, *16* (1), 8–37.

Wilson, Robert, “Game-Theoretic Analysis of Trading Processes.,” Technical Report, Stanford University, 1985.

Wolitzky, Alexander, “Mechanism design with maxmin agents: theory and an application to bilateral trade,” *Theoretical Economics*, 2016, *11*, 971–1004.