



**HAL**  
open science

## Incentives in Cooperation and Communication

Andrés Salamanca Lugo

► **To cite this version:**

Andrés Salamanca Lugo. Incentives in Cooperation and Communication. Economics and Finance. Université Toulouse 1 Capitole (UT1 Capitole), 2017. English. NNT: . tel-01720051

**HAL Id: tel-01720051**

**<https://hal.science/tel-01720051>**

Submitted on 28 Feb 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITÉ TOULOUSE 1 CAPITOLE  
ÉCOLE DOCTORAL “TOULOUSE SCHOOL OF ECONOMICS - TSE”  
UNITÉ MIXTE DE RECHERCHE “TSE-R”

---

**INCENTIVES IN COOPERATION AND COMMUNICATION**

---

THÈSE

Présentée et soutenue le 24 Novembre 2017

en vue de l'obtention du

DOCTORAT EN SCIENCES ÉCONOMIQUES

par

**Andrés SALAMANCA LUGO**

devant le jury composé de

<b>Françoise</b>	<b>FORGES</b>	Professeur des Universités Université Paris Dauphine	<i>Directeur</i>
<b>Thomas</b>	<b>MARIOTTI</b>	Directeur de Recherche CNRS Toulouse School of Economics	<i>Directeur</i>
<b>Frédéric</b>	<b>KOESSLER</b>	Directeur de Recherche CNRS Paris School of Economics	<i>Rapporteur</i>
<b>David</b>	<b>WETTSTEIN</b>	Professor Ben-Gurion University of the Negev	<i>Rapporteur</i>
<b>Peter</b>	<b>SUDHÖLTER</b>	Professor University of Southern Denmark	<i>Suffragant</i>
<b>Jérôme</b>	<b>RENAULT</b>	Professeur Université Toulouse 1 Capitole	<i>Suffragant</i>



*À ma chère épouse Audrey*  
*En mémoire de mon cher père Orlando*



# Remerciements

Françoise Forges, vous avez excellé dans le rôle de directrice de cette thèse et mes remerciements s'adressent tout naturellement à vous dans un premier temps. Merci d'avoir partagé votre expertise qui a contribué grandement à l'avancée de ce travail. Votre regard critique, votre patience, votre disponibilité et vos conseils avisés m'ont permis de mener à bien cette thèse avec succès. Je vous transmets mes respects et mon admiration la plus profonde.

Je tiens aussi à remercier mon co-directeur Thomas Mariotti qui a toujours apporté un oeil nouveau et encourageant sur mes recherches. Ses remarques aux nombreux documents que je lui ai adressés ont été d'une grande valeur scientifique et pédagogique.

Je n'oublie pas de remercier Michel Le Breton qui fût mon co-encadrant pendant ma première année de thèse. Merci pour son suivi et son soutien durant cette première étape de mon doctorat.

Je suis très honoré que David Wettstein et Frédéric Koessler aient accepté d'écrire le rapport de cette thèse. Je les remercie pour tout l'intérêt qu'ils ont porté à mes travaux. C'est aussi un honneur que m'ont fait Peter Sudhölter et Jérôme Renault en acceptant d'être membres de ce jury.

Je remercie les laboratoires TSE-R (anciennement GREMAQ, Groupe de Recherche en Économie Mathématique et Quantitative) et LEDa (Laboratoire d'Économie de l'Université Paris Dauphine) pour leur soutien financier qui m'a permis de participer à de nombreuses conférences. Mes remerciements s'adressent aussi à tous les chercheurs que j'ai rencontrés à l'occasion de conférences, et dont les discussions ont été très stimulantes et agréables. Une mention spéciale à Marie Laclau, qui m'a invité au Séminaire Parisien de Théorie des Jeux et au Séminaire de Théorie, Organisation et Marchés à Paris School of Economics et qui, par ailleurs, a soutenu ma candidature à l'École d'été en Théorie des Réseaux (2017) à Roscoff.

Je remercie également la Banque Centrale colombienne (Banco de la República de Colombia) ainsi que le Gouvernement Français pour leur soutien financier à travers d'une bourse d'études supérieures et d'un contrat doctoral. Ces aides économiques m'ont permis de finir ce doctorat dans les meilleures conditions possibles.

J'ai eu la chance de faire de belles collaborations en tant que professeur assistant avec d'autres enseignant-chercheurs. Je tiens donc à les remercier pour cette expérience. J'ai une pensée particulière pour François Salanié, Bertrand Gobillard, Thierry Magnac et Élodie Alet.

Mes remerciements vont également au personnel administratif de la TSE, en particulier Mme. Laurance Delorme, Céline Lemesle, Louise Strachan, Caroline Tédor et Aude Schloesing pour leur sympathie et leur assistance dans toutes les procédures administratives.

J'aimerais aussi remercier tous mes proches qui ont vécu cette thèse en même temps que moi. Un grand merci tout spécialement à ma chère épouse Audrey pour son soutien quotidien sans faille. Ce travail n'aurait jamais vu le jour sans toi. Je pense aussi à mes parents pour leur encouragement constant malgré la distance. Je veux également exprimer ma gratitude envers tous les membres de ma belle-famille. Si ces années ont été si agréables, c'est surtout grâce à vous.

En conclusion de ces remerciements, je souhaiterais mettre en avant l'excellente qualité de vie que j'ai rencontrée dans ce beau pays qui est la France. Je me suis réjoui de sa gastronomie, de sa nature et ses paysages, de sa langue, de sa culture et de la sympathie de ses citoyens. Enfin, merci de m'accueillir et de me permettre devenir l'un des vôtres.

# Abstract

This dissertation consists of three self-contained papers in which we analyze cooperation and strategic information transmission in situations of asymmetric information where communication is subject to incentive constraints. Chapter 1 proposes a new solution concept for cooperative games with incomplete information. Chapter 2 compares this solution and other cooperative solutions in various classes of games (two-player games, games with transferable utility, games with verifiable information). Chapter 3 proposes a new approach to Bayesian persuasion by characterizing the ex-ante optimal communication equilibrium for the sender in the class of sender-receiver games.

Myerson [Cooperative games with incomplete information. *Int. J. Game Theory*, 13, 1984, pp. 69- 96] has made significant progress towards a general concept of value for cooperative games with asymmetric information. His cooperative solution, called the *M-value* (short for Myerson value), generalizes the Shapley non-transferable utility (NTU) value to games with incomplete information. In Chapter 1, we show that Myerson's theory exhibits some "difficulties" for recognizing certain informational externalities. To do this, we construct a three-player cooperative game in which the M-value does not capture some "negative" externality generated by the adverse selection. We then introduce a new solution concept, which we call the *H-value*. Our theory generalizes the Harsanyi NTU value to cooperative games with incomplete information. When we explicitly compute the H-value in our game, it turns out that it prescribes a more intuitive outcome taking into account the informational externalities not captured by the M-value.

In Chapter 2 we explore the relationship between the following value like solution concepts for cooperative games with incomplete information: the M-value, the H-value and A. Kalai and E. Kalai's [Cooperation in strategic games revisited. *Q. J. Econ.*, 128, (2013), 917-966] *cooperative-competitive (or "coco") value*. We consider a model in which utility transfers in the form of sidepayments are allowed. In our model, however, state-contingent contracts are required to be incentive compatible, thus utility might not be not fully transferable (as it would be in the complete information case). Restricting attention to games with orthogonal coalitions, which do not involve strategic externalities, we show that the M-value and the H-value coincide. Allowing for arbitrary informational and strategic externalities, we show that the ex-ante evaluation of the M-value equals the coco value in two-player games with verifiable information.



In Chapter 3 we provide an analytical framework for studying Bayesian persuasion problems. We consider a model of strategic information transmission in which a sender chooses a communication system for signaling his information to an uninformed receiver, who then takes an action that affects the welfare of both individuals. Our main concern in this chapter is the question, what kinds of communication systems are the best ones for the informed party? By a use of a general form of the *revelation principle*, we can restrict attention to communication equilibria (mediated communication protocols). Using a geometric approach based on Duality Theory, we are able to characterize the optimal communication equilibrium from the *concavification* of a (modified) non-revealing payoff function as in Aumann and Maschler [*Repeated Games with Incomplete Information*. (1995). Cambridge, MIT Press.].

# Résumé

Dans cette dissertation, composée de trois articles, nous étudions la coopération et la transmission stratégique d'information dans des situations d'asymétrie d'information où la communication est soumise à des contraintes d'incitation. Le chapitre 1 propose un nouveau concept de solution pour les jeux coopératifs en information incomplète. Le chapitre 2 compare cette solution et d'autres solutions de coopération dans différentes classes de jeux (jeux à 2 joueurs, jeux à utilité transférable, jeux en information vérifiable). Le chapitre 3 propose une nouvelle approche pour la "persuasion Bayésienne" en caractérisant l'équilibre en communication optimale (ex-ante) pour l'émetteur dans les jeux émetteur-récepteur.

Myerson [Cooperative games with incomplete information. *Int. J. Game Theory*, 13, 1984, pp. 69- 96] a fait des progrès considérables vers un concept de valeur s'appliquant aux jeux coopératifs en information asymétrique. Sa solution coopérative, désignée *valeur M*, étend la valeur de Shapley à utilité non-transférable (UNT) aux jeux en information incomplète. Dans le Chapitre 1, nous montrons que la théorie développée par Myerson présente des "difficultés" pour reconnaître certaines externalités d'information. Pour ce faire, nous construisons un jeu coopératif à trois joueurs dans lequel la valeur M ne capture pas une externalité négative engendrée par la sélection adverse. Nous introduisons ensuite un nouveau concept de solution, que nous appelons la *valeur H*. La théorie que nous proposons généralise la valeur d'Harsanyi UNT aux jeux coopératifs en information incomplète. Lorsque nous calculons explicitement la valeur H dans notre jeu, il s'avère qu'elle prescrit un résultat plus intuitif tenant compte des externalités non reconnues par la valeur M.

Dans le Chapitre 2, nous examinons la relation entre les concepts de solution suivants: la valeur M, la valeur H et la *valeur coco* introduite par A. Kalai and E. Kalai's [Cooperation in strategic games revisited. *Q. J. Econ.*, 128, (2013), 917-966]. Nous considérons un modèle dans lequel des transferts d'utilité sous forme de paiements latéraux sont autorisés. Cependant, dans notre modèle, les contrats prennent en compte les possibilités de communication des individus et les contraintes d'incitation qui en résultent, de sorte que l'utilité pourrait ne pas être entièrement transférable (comme c'est le cas dans le modèle en information complète). En se concentrant sur les jeux à coalitions orthogonales, c'est à dire qui ne comportent pas d'externalités stratégiques, nous montrons que la valeur M et la valeur H coïncident. En permettant simultanément des externalités stratégiques et d'information, nous montrons que l'évaluation ex-ante (c.à.d. avant que tout joueur ne reçoive son information privée) de la valeur M coïncide avec la valeur coco

dans les jeux à deux joueurs dont l'information est ex-post vérifiable.

Dans le chapitre 3, nous proposons un cadre analytique pour l'étude des problèmes dits de "persuasion Bayésienne". Nous étudions un modèle de transmission stratégique d'information dans lequel un agent informé, l'émetteur, choisit un système de communication pour relayer son information privée à un agent non informé, le récepteur, qui prend une décision affectant les utilités des deux individus. Notre principale préoccupation dans ce chapitre est de déterminer les meilleurs systèmes de communication pour l'émetteur. En utilisant une forme générale du *principe de révélation*, on ne perd rien à se limiter aux équilibres en communication (protocoles de communication assistés d'un médiateur). Grâce à une approche géométrique basée sur la Théorie de la Dualité, nous sommes en mesure de caractériser l'équilibre en communication optimale pour l'émetteur à partir de la *concavification* d'une fonction de paiements d'équilibre non révélateur (modifiée), comme dans la méthode d'Aumann et Maschler [*Repeated Games with Incomplete Information*. (1995). Cambridge, MIT Press.].

# Contents

<b>Introduction</b>	<b>1</b>
<b>A Generalization of the Harsanyi NTU Value to Games with Incomplete Information</b>	<b>7</b>
1 Introduction . . . . .	7
2 Formulation . . . . .	10
2.1 Bayesian Cooperative Game . . . . .	10
2.2 Incentive Efficiency and The Virtual Utility Approach . . . . .	11
2.3 The M-solution . . . . .	12
3 Motivating Examples . . . . .	13
3.1 Example 1: A Collective Choice Problem . . . . .	13
3.2 Example 2: A Bilateral Trade Problem . . . . .	16
4 Equity Principles for Bayesian Cooperative Games . . . . .	17
5 Optimal Threats . . . . .	21
6 The H-Solution . . . . .	23
6.1 Example 1 . . . . .	24
6.2 Example 2 . . . . .	26
7 Some Comments About the (Non-)Existence of the H-solution . . . . .	27
7.1 Example 3: Non-existence of the H-solution . . . . .	27
7.2 Free Disposal and the Structure of Incentives . . . . .	29
8 Appendix: Proof of Proposition 4 . . . . .	30
<b>On the Values for Bayesian Cooperative Games with Sidepayments</b>	<b>35</b>
1 Introduction . . . . .	35
2 Bayesian Cooperative Game . . . . .	37
3 Incentive Efficiency and Virtual Utility . . . . .	38
4 Values for Bayesian Cooperative Games with Orthogonal Coalitions . . . . .	39
4.1 The M-Solution . . . . .	40
4.2 The H-Solution . . . . .	41
4.3 Reconciling the Differences . . . . .	42
5 Values for Two-person Bayesian Games . . . . .	47

<b>The Value of Mediated Communication</b>	<b>55</b>
1 Introduction . . . . .	55
2 Motivating Example . . . . .	58
3 Basic Game . . . . .	62
4 Mediated Persuasion . . . . .	64
4.1 Mediated Persuasion Under Verifiable Information . . . . .	65
4.2 The Virtual Persuasion Game . . . . .	66
4.3 Optimal Mediators . . . . .	71
4.4 Extreme Communication Equilibria and the Number of Signals . . . . .	73
5 Discussions . . . . .	76
5.1 Cheap-Talk Implementation . . . . .	76
5.2 Information Design Problems . . . . .	76
<b>Bibliography</b>	<b>79</b>

# Introduction

Communication is often limited by the conflict of interest between the different parties. The reason is that, although an individual may find it advantageous to disclose her private information, the uninformed parties cannot always believe in such disclosure, since in alternative circumstances the informed individual would have released the same information in order to reap the benefits from cheating. Communication is therefore subject to *incentive constraints*, which take into consideration the fact that anything an individual says about her private information or any promise she does to make some decision is not believable whenever it contradicts the individual's incentives. Incentive constraints might thus limit the ability of individuals to coordinate joint decisions, which explains some of the difficulties that individual parties in dispute face for reaching mutually beneficial cooperative agreements. This matter is treated in Chapters 1 and 2, where we study the extent to which the concept of *value* from Cooperative Game Theory can be extended to analyze cooperative agreements in situations of asymmetric information. Incentive constraints also rise important conceptual issues for the optimal design of communication systems. Our main concern in Chapter 3 is the analysis of incentives in situations where an informed individual wants to signal information to some uninformed party who retains the ultimate right of making the final decision. We consider a model in which a player can send messages about his private information to an uninformed receiver, through some communication system, to try to influence the receiver's decision. We are interested in answering the question, what kind of communication systems are the best ones for the sender? Although the model adopted in Chapter 3 differs conceptually from the one studied in the preceding chapters, there are some similarities in the mathematical description of both interaction situations that allow us to adapt some techniques and results introduced in Chapters 1 and 2 in order to provide an answer to the question motivating Chapter 3.

The concept of value from Cooperative Game Theory starts with the assumption that players will agree on some *feasible* utility (payoff) allocation that is *efficient*, in the sense that there is no feasible reallocation of the utilities making all individuals better off. Yet, in most game situations there are multiple efficient allocations, each of them being preferred by different individuals. *Equity* principles are then used in order to identify an efficient allocation to be, in some sense, a "fair" compromise between the interests of the different individuals. The selected allocation is called a value. Various values may be defined according to different equity principles.

Notable efforts have already been done to understand the implications of asymmetric information for the notions of feasibility, efficiency and equity<sup>1</sup>. In particular, Myerson (1983, 1984a,b) has made significant progress towards a general concept of value for cooperative games with asymmetric information. However, in the same vein as de Clippel (2005), we exhibit in Chapter 1 some “difficulties” with Myerson’s theory for recognizing certain informational externalities. In the same chapter we provide an alternative notion of value dealing with these difficulties. Even though there are important conceptual differences between Myerson’s theory and ours, in Chapter 2 we identify instances in which both theories can be reconciled.

Cooperative Game Theory generally represents an interaction situation as a collection of utility sets describing the feasible outcomes that the players can achieve when they cooperate in different coalitions (groups of players). Any explicit mention of the decisions generating the utilities is suppressed. This *characteristic function form*, by definition, rules out *strategic externalities* –situations in which the utilities of the players inside a coalition depend on what the other coalitions are doing–. Coalitions are then said to be *orthogonal*. The characteristic function is defined according to several hypothesis on the geometric properties of the utility sets. A widely accepted assumption is that of *comprehensiveness*, which means that players can dispose of their utility at will (*free disposal*)<sup>2</sup>. This assumption has been considered to be innocuous, since it can only enlarge the utility sets by adding Pareto dominated allocations<sup>3</sup>. Nevertheless, its relevance extends beyond the fact that it increases the strategic possibilities of the players. Indeed, comprehensiveness is of utmost importance for the existence and axiomatic characterizations of the value.

In Chapters 1 and 2 we study game situations with orthogonal coalitions where, at the time in which coalitional agreements are made, each player may have private information that the other players do not know<sup>4</sup>. We use the concept of *type* developed by Harsanyi (1967-8) for modeling asymmetric information. Each player can be of several types where a type is an encapsulated description of the whole hierarchy of beliefs that a player holds about the unknown parameters of the game, about the other players’ beliefs about these parameters, about the other players’ beliefs about her beliefs about the parameters, and so on ad infinitum. More commonly, a player’s type is thought as a full description of the player’s private information. In a situation of asymmetric information, an allocation should now be seen as a *mechanism* (type-contingent decision plan). The enforcement of any such mechanisms is subject to incentive constraints guaranteeing that the cooperative agreement does not create the incentives for any player to lie about her actual type. Incentive constraints restrict what is feasible in a way that makes it impossible to describe a cooperative game as a collection of utility sets. Indeed, the fact that a utility allocation is feasible at some type profile does not allow us to determine what would

---

<sup>1</sup>See Myerson (1979, 1983, 1984a,b), Holmström and Myerson (1983), Nehring (2004), de Clippel (2008, 2012), de Clippel and Minelli (2004) and de Clippel, Pérez-Castrillo and Wettstein (2012). Forges and Serrano (2013) present a brief survey describing some of the recent progress and open problems in the area of cooperative games with asymmetric information.

<sup>2</sup>If a utility set fails to be comprehensive, then it may be bounded below.

<sup>3</sup>An allocation is Pareto dominated if it is possible to reallocate the utilities so as to make one individual better off without making any other individual worse off.

<sup>4</sup>This situation is known as the *interim* case. Other forms of incomplete information include the *ex-ante* case, in which players enter into coalitional agreements before they receive any private information.

be the utility an individual would get by reporting a different type. Thus, decisions should be explicitly included in the (primitive) description of the game. As a consequence, the geometric properties assumed in the characteristic function cannot be easily adapted to situations with asymmetric information. In particular, free disposal activities may have an important effect on the incentives structure of the game, thus they cannot be considered innocuous anymore. This issue is discussed in Chapter 1 (see below).

Holmström and Myerson (1983) characterize *incentive-efficient mechanisms* through parametric linear programming problems. The incentive constraints of these optimization problems yield “shadow prices” that are employed by Myerson (1984a,b) to define the *virtual utility* of players. These virtual utility scales incorporate into the player’s utility function the signaling costs (informational inefficiencies) that may be incurred due to the incentive constraints. By considering virtual utilities, rather than real utilities (*virtual utility hypothesis*), Myerson (1984a,b) defines some principles for an equitable compromise. Such a compromise is proportional with the contributions of every individual to all coalitions to which she belongs. Moreover, it is inscrutable, in the sense that it achieves a balance between the interests of the actual type of a player and the goals of the other possible types that she pretends to be. Elaborating on this equity principle, Myerson (1984b) develops a theory of values for selecting a mechanism among the many incentive efficient mechanisms. The so-defined *M-value* (short for Myerson value) extends Shapley’s (1969) non-transferable utility (NTU) value to games with asymmetric information.

Aware of the difficulties that incentive constraints may entail for the feasibility of equitable agreements, Myerson (1984b) imposes a series of assumptions to guarantee the existence of his value<sup>5</sup>. The implied simplified nature of the M-value has led to it being insensitive to some “positive” information externalities. This is evidenced by an eloquent example introduced by de Clippel (2005). In Chapter 1, we provide another intuitive example in which the M-value does not capture some “negative” externality generated by the adverse selection. In the hope to provide more appealing outcomes for these examples, we elaborate on the virtual utility approach in order to develop alternative equitable compromises based on the principle of equal gains (*egalitarian criterion*), according to which cooperating players within a coalition should have equal compensations for their cooperation. These egalitarian principles are then used to define a new value, which we call the *H-value*. Our theory generalizes Harsanyi’s (1963) NTU value to games with asymmetric information.

The H-value is a more sophisticated adaptation of the Myerson’s (1984b) theory. Indeed, our egalitarian criterion implies that the H-value is also equitable in the sense of Myerson (1984a,b). However, both theories differ in the way they evaluate the strength of the different coalitions. When applied to the examples mentioned above, the H-value prescribes outcomes for which there is more agreement with what we intuitively expect these outcomes to be. Unfortunately, the H-value may fail to exist for some games. The reason is that, when utility cannot be transferred between the players, equal compensations might require some of the players to forego a portion of their proceeds<sup>6</sup>, which might be impossible due to the way incentive constraints

---

<sup>5</sup>Indeed, in the absence of appropriate assumptions on the feasible utility sets, it may not exist a (feasible) efficient allocation which is also equitable. This problem is exacerbated by the presence of incentive constraints.

<sup>6</sup>To illustrate the situation, imagine that two kids have to agree on a way to distribute 3 candies. Assume that each bonbon cannot be splitted into smaller pieces (lack of transferable utility). If both children are constrained to



restrict what is feasible. In the absence of information asymmetries, the same difficulty is ruled out by the comprehensiveness assumption<sup>7</sup>. Then, one is tempted to accommodate free disposal by introducing decisions specifying how much utility a player may discard. However, adding new decisions may change the incentive structure of the game: free disposal can be used for signaling purposes, i.e., for weakening incentive compatibility<sup>8</sup>. As a result, for any (interim incentive) efficient utility allocation, we cannot generally extend the original game by introducing additional decisions allowing players to discard utility, while leaving the original utility allocation efficient in the expanded game. This issue prevents us from obtaining an existence result of the H-value. Incentive constraints interconnect the decisions in different states in an intricate way, which complicates matters for identifying (nontrivial) conditions on the primitives of the game allowing to introduce free disposal activities without having to be concerned about the issue described above. Despite the recognized difficulties, our results suggest that the H-value is the most suitable way to extend the Harsanyi NTU value to games with information asymmetries.

In Chapter 2 we look for ways to reconcile the differences between the M-value and the H-value. A simple situation in which both values coincide is when there are only two players. This is so because, in two-player games, the only possible threats come from “singleton” coalitions (consisting of a single player) which are treated the same by both solution concepts. When information is complete, it is well known that both the Shapley NTU value and the Harsanyi NTU value coincide in games with *unrestricted* utility transfers. Moreover, their common formula is given by Shapley’s (1953) transferable utility (TU) value. We thus explore the possibility of extending this result to games with asymmetric information. For that, we consider a broad class of mechanisms allowing the players to transfer utility in the form of state-contingent sidepayments. Any of these mechanisms is required to satisfy the incentive constraints, and thus our model may exhibit *restricted* utility transfers, i.e., not all state-contingent sidepayments are feasible. In spite of this, our first main result in Chapter 2 states that, in our model with sidepayment, any M-value is an H-value, and viceversa<sup>9</sup>. Their common definition, however, cannot be described by a simple closed form expression as the Shapley TU value. The reason is that, due to the incentive constraints, the set of (interim incentive) efficient utility allocations is not generally described by a hyperplane as it would be in a game with complete information.

In the second part of Chapter 2 we consider game situations in which the players face strategic externalities. However, we simplify the coalitional analysis by focusing only on two-player games. We first extend the definition of the M/H-value to allow for strategic externalities<sup>10</sup>. To do this, we follow the method developed by Nash (1953) in defining his bargaining solution

---

have the same number of candies, they have to forego one of the candies.

<sup>7</sup>Comprehensiveness of the characteristic function is sometimes replaced by the more stringent assumption that feasible sets are *compactly generated*, i.e., they are the (closed) comprehensive hull of a compact set.

<sup>8</sup>“I forego a portion of my proceeds for the sake of a fair agreement, but also I want to demonstrate that I am the type that loses very little by agreeing to discard utility”.

<sup>9</sup>Remarkably, this result is not the consequence of the fact that sidepayments can be used for weakening incentive compatibility (see d’Aspremont and Gérard-Varet (1979, 1982)), but rather it follows from the fact that coalitional agreements can be made equitable by means of an appropriate transfer scheme.

<sup>10</sup>Recall that in the two-player case both the M-value and the H-value coincide regardless of whether sidepayments are allowed or not.

with *variable threats*<sup>11</sup>. Recently, A. Kalai and E. Kalai (2013) proposed a value for two-player games with incomplete information and transferable utility. Their semi-cooperative solution, called the cooperative-competitive (or “coco”) value, is based on a decomposition of the game into cooperative and competitive component games. The coco value conceptually differs from the M-value in that the former ignores potential incentive compatibility issues. Also, it is defined only at the ex-ante stage, i.e., before any player receives her private information. Our second main result in Chapter 2 asserts that, when sidepayments are allowed and private information is ex-post verifiable<sup>12</sup> (so that incentive constraints are unnecessary), the M-value and the coco value are ex-ante utility equivalent, that is, if the players evaluate their welfare as if they were uninformed, both values prescribe the same utility allocation.

The analysis presented in Chapters 1 and 2 leads us through the basic conceptual issues arising for cooperation in situations of asymmetric information where communication is subject to incentive constraints. On the other hand, incentive compatibility also gives rise to important issues for the problem of information design. In Chapter 3 we study optimal design of communication systems in a particular model of information transmission. We consider a model in which, before learning his type (ex-ante stage), an individual, called the *sender*, chooses a public *communication device* for signaling his information to an uninformed *receiver*, who then takes an action that affects the welfare of both individuals. Here a communication device is meant to be any procedure helping the players to transmit information and to coordinate decisions. It may include plain conversation schemes (direct communication), consisting of several rounds in which the players exchange messages. It also may consist of noisy communication channels (indirect communication) in which an intermediary individual (or a machine) sends signals to every player over multiple stages but also receives messages from them. These two kinds of communication devices are examples of interplay communication systems in which communication occurs after the sender has learnt his type. However, a communication device may also involve preplay communication i.e., before the sender learns his type (correlation device).

Mediated communication is a particular communication device in which the sender reports a type to a neutral trustworthy mediator who then recommends an action to the receiver. The sender’s report is not verifiable either by the mediator or by the receiver, which allows the sender to strategically manipulate his private information. The mediator’s recommendation is not binding, that is, the receiver is free to choose any action different from the recommended one. A mediation protocol is then constrained by the necessity to provide the appropriate incentives for the sender to reveal his private information (informational incentive constraints) and for the receiver to follow the prescribed recommendation (strategic incentive constraints). A mediation protocol in which the sender always reports the truth and the receiver always follows the recommendation is called a *communication equilibrium*. In our setting, a very broad form of the *revelation principle* applies and, thus, without loss of generality, we can focus on communication equilibria (see Forges (1985, 1986)). The problem of the sender is thus to select a communication equilibrium maximizing his ex-ante expected payoff.

---

<sup>11</sup>The generalization here presented is not new. Indeed, Myerson (1984b) offers a more general approach to the case  $n \geq 2$  without sidepayments.

<sup>12</sup>Private information is verifiable ex-post if when agreements are implemented all private information becomes public.

This model of information transmission is mathematically analogous to a two-person cooperative game with asymmetric information in which contracting takes place at the ex-ante stage<sup>13</sup> and the unique informed individual has all the bargaining ability (as in Myerson (1983)). In this context, having all the bargaining ability does not mean that the sender can force the receiver to behave in a certain way. It means that the sender has the power to manipulate the receiver's incentives by controlling the communication channels (persuasion ability) and that the receiver does not have any threatening ability. As a consequence, here a mechanism is a mediation protocol in which the mediator merely suggests an action but has no power to impose decisions, which makes necessary the presence of additional strategic incentive constraints. This is in contrast with the mechanisms considered in Chapters 1 and 2, which must be understood as *arbitration* protocols in which the "mediator" imposes a settlement once it is agreed upon. Under this alternative interpretation of the model, the value can be assimilated to the sender's optimal (ex-ante) expected payoff.

The previous interpretation of the model gives us some insights into the use of the virtual utility approach for characterizing the optimal mediated communication protocol. Specifically, the truth-telling incentive constraints of the sender's problem yield shadow prices that can be used to define the virtual utility of the sender exactly in the same way as in Chapters 1 and 2. These virtual utility scales are characterized by the signaling costs measured by the shadow prices and the prior beliefs. Relying on the virtual utility hypothesis, we construct a *fictitious game* in which there are no truth-telling incentive constraints and the sender's payoffs are in the virtual utility scales. We then proceed to construct a non-revealing virtual payoff function over prior beliefs describing the (ex-ante) expected virtual payoffs that the sender can achieve when there is no communication in the fictitious game. As in Kamenica and Gentzkow (2011), the sender's optimal expected payoff (*value of persuasion*) in the fictitious game equals the *concavification*<sup>14</sup> of the non-revealing virtual payoff function evaluated at the initial prior beliefs. Our main result in this chapter says that the value of persuasion in the original problem equals the value of persuasion in the fictitious game with virtual scales defined by the optimal shadow prices, i.e., those minimizing the signaling costs.

We characterize the optimal mediation protocol through a constrained splitting of the prior belief into a distribution over posterior beliefs. Given the optimal signaling costs, the posterior beliefs induced by any optimal mediation protocol correspond to the points on the domain of the non-revealing virtual payoffs function for which the convex combination of their images yields the value of persuasion. The corresponding distribution of posteriors is constrained by the Bayes plausibility (martingale property) together with *complementary slackness* conditions from Duality Theory. Complementary slackness says that if a shadow price is positive, then the associated informational incentive constraint must be binding.

---

<sup>13</sup>By assuming that the choice of a communication device is made at the ex-ante stage, we exclude the possibility that the mere fact of choosing it signals information. This allows us to focus on the problem of information transmission without having to worry about the need to consider the sender's inscrutable inter-type compromises.

<sup>14</sup>The concavification of a function  $f$  is the smallest concave function that is larger or equal to  $f$ .

# A Generalization of the Harsanyi NTU Value to Games with Incomplete Information

## 1 Introduction

The *value* is a central solution concept in the theory of cooperative games<sup>1</sup>. Introduced by Shapley (1953) for the study of games with transferable utility (TU), the value has been extended in different ways to general games with nontransferable utility (NTU); some of the most notable NTU values are due to Harsanyi (1963) and Shapley (1969)<sup>2</sup>. The value has proved to be a surprisingly useful solution concept for the analysis of cooperative outcomes in economic models under complete information (see Aumann (1994) for a partial bibliography of applications). However, many interesting economic situations are characterized by information asymmetries, such as adverse selection and moral hazard problems. Then, the question of examining the value in more realistic environments with incomplete information naturally arises.

Under incomplete information, an agreement should be seen as a mechanism (state contingent decision plan). The enforcement of any such mechanisms relies on the players' claims about their private information. As a consequence, the final agreement may be subject to strategic manipulation. A cooperative agreement must then be *incentive compatible*, in the sense that it provides the appropriate incentives for every individual to reveal honestly his private information.

Myerson (1984a,b) developed a method in which the incentive constraints are used to define the *virtual utility* of the players. Virtual utility generalizes the weighted-utility scales in the Harsanyi-Shapley fictitious transfer procedure<sup>3</sup>. Elaborating on this approach, Myerson (1984b)

---

<sup>1</sup>This chapter corresponds to the paper: *A Generalization of the Harsanyi NTU Value to Games with Incomplete Information*, HAL Working paper 01579898, 2016.

<sup>2</sup>The Shapley NTU value is sometimes referred as the “ $\lambda$ -transfer value”. The Harsanyi NTU value, being less tractable, has received less attention. Indeed, the Shapley NTU value was introduced as a simplification of the Harsanyi NTU value. Both values are compared in Hart (2004) by means of a simple example. The reader is referred to Peleg and Sudhölter (2007, ch. 13) and Myerson (1991, ch. 9) for further details and formal definitions of these two solution concepts.

<sup>3</sup>Myerson (1992) provides a detailed explanation of the fictitious transfer procedure.

defines a bargaining solution which extends the Shapley NTU value to games with incomplete information. The *M-solution* (short for Myerson's solution) takes into account not only the signaling costs associated to incentive compatibility, but also the fact that individuals negotiate at the interim stage. It also involves the identification of "rational threat" mechanisms for each coalition. Rational threats determine how much credit each (type of a) player can claim from the proceeds of cooperation.

In order to keep a tractable mathematical formalization allowing for general existence of the M-solutions, Myerson (1984b) imposes various assumptions on the commitment structure of the underlying bargaining situation (see Section 6 in Myerson (1984b) for a detailed discussion). These simplifying assumptions entail, however, a reduced sensitivity of the M-solution to some informational externalities. This is evidenced by a prominent example introduced by de Clippel (2005).<sup>4</sup> In this paper we provide another intuitive example in which the M-solution does not capture some "negative" externality generated by the adverse selection. Starting from the two-person bargaining problem studied in Section 10 of Myerson (1984a), we construct a three-player game in which the uninformed individuals (players 1 and 2) can overcome the potential adverse selection problem they face by ignoring the informed individual (player 3) and agreeing on an outcome that is equitable and efficient for both of them. As we will argue in Section 3, a reasonable outcome for this game should leave the informed player with a low expected payoff. Nevertheless, under the M-solution the informed player extracts a considerable amount of utility. Our example shares features with a complete information NTU game previously proposed by Roth (1980). In that game, the Shapley NTU value exhibits some difficulties of a similar nature to that of the M-solution in our example. Hart (1985a) showed, however, that the Harsanyi NTU value prescribes a more appealing outcome in Roth's game. Our main goal in this paper is to generalize the Harsanyi NTU value to games with incomplete information. Therewith we hope to provide an alternative outcome for our three-player game.

Harsanyi (1963) introduced his NTU value using a model of bargaining in which players inside each coalition negotiate a vector of *dividends*. This dividend allocation procedure is rather intractable and difficult to extend to games with incomplete information. In this work, we shall generalize a simpler (yet equivalent) definition of the Harsanyi NTU value introduced by Myerson (1980). This definition, which dispenses with the notion of dividends, is characterized by an equity condition called *balanced contributions* (see also Myerson (1992) for a detailed explanation). While there might be several appealing ways to extend the balanced contributions to games with incomplete information, here we adopt a method that preserves a conceptual coherence with Imai's (1983) equivalent *subgame value* characterization of the Harsanyi NTU value. In Section 4, we build on Myerson's (1984a,b) virtual utility approach to formulate a "natural" extended version of the subgame value equity condition. We then define an *egalitarian criterion* to be the unique extension of the balanced contributions that is consistent with our generalized subgame value condition (cf. Proposition 2). These equity principles are then used in Section 5 to define optimal threat mechanisms for all coalitions. In Section 6 we formally define our cooperative solution concept, which we call the *H-solution*. We also exhibit its properties.

---

<sup>4</sup>De Clippel's example is an incomplete information version of a NTU game introduced by Owen (1972).

We construct the H-solution to be a more sophisticated adaptation of Myerson's (1984b) theory. Indeed, the H-solution requires all threat mechanisms to be equitable, whereas the M-solution only requires equity in the case of the grand coalition. This difference between both solution concepts is a matter of "credibility" of the threats. As a result, when we explicitly compute the H-solution in our example, it turns out that it prescribes an outcome for which there is more agreement with what we intuitively expect the outcome of this game to be. Unfortunately, and as it might be expected, extending equity to all coalitions makes a significant difference to the analysis, and the H-solution may fail to exist. In Section 7 we provide a simple example of a game in which there is no H-solution. Under complete information, the same difficulty for the Harsanyi NTU value is ruled out by considering games whose characteristic function is comprehensive<sup>5</sup>. This amounts to assuming free disposal of utility. The same approach does not immediately extend to games with incomplete information. Indeed, when cooperative agreements are made at the interim stage, it is not clear how to derive an analog of the characteristic function under incomplete information<sup>6</sup>. On the other hand, while in a setting with complete information free disposal is usually taken as an innocuous assumption, the same cannot be done under asymmetric information. In fact, allowing players to discard utility at the interim stage may alter the incentives structure of the game, as it will be illustrated in Section 7. The previous difficulties prevent us from obtaining an existence result of the H-solution. This is not specific of our approach. Indeed, de Clippel (2012) encounters similar difficulties for the existence of an alternative (interim) egalitarian criterion in the context of mechanism design. The techniques used in the special case of complete information to achieve positive results cannot generally be extended to games with incomplete information. The reason is that incentive compatibility makes arguments significantly more complicated. This is also the case for the non-emptiness of the core of an exchange economy with incomplete information (see Forges, Minelli and Vohra (2002) for a detailed discussion on this issue).

As described above, the paper is organized as follows: Section 2 is devoted to specifying formally the model of a cooperative game with incomplete information and the notations used, including the basic assumptions on the class of games considered. We also present a summary of the facts one needs to know about Myerson's (1984b) virtual utility approach. Our motivating example is analyzed in Section 3. The virtual utility approach is used in Section 4 to define our egalitarian criterion. In Section 5, the ideas of Section 4 are applied to define optimal egalitarian threats. In Section 6 we introduce the H-solution. We then compute the H-solution of the example of Section 3. Non-existence of the H-solution is discussed in Section 7.

---

<sup>5</sup>A characteristic function game  $V$  is *comprehensive* if, for every (nonempty) coalition  $S$ , whenever  $V(S) \subseteq \mathbb{R}^S$  contains an allocation  $u$ , it also contains all allocations  $v$  satisfying  $v \leq u$ . Further assumptions are also required for the existence of the Harsanyi NTU value: (i) if  $u, v \in \partial V(N)$  (i.e.,  $u$  and  $v$  are efficient for the grand coalition) and  $u \geq v$ , then  $u = v$  (*non-levelness*); (ii)  $V(N) = K + C$ , where  $K \subseteq \mathbb{R}^N$  is a compact set and  $C \subseteq \mathbb{R}^N$  is a convex cone (see Peleg and Sudhölter (2007, Theorem 13.3.5)). Assumption (i) excludes vanishing utility weights, while (ii) is a technical assumption guaranteeing that the set of utility weights  $\lambda \in \mathbb{R}_{++}^N$  for which the "primal problem"  $\max_{v \in V(N)} \lambda \cdot v$  has a finite optimum is compact and convex. It is worth noticing that these assumptions are not necessary for the definition of the Harsanyi NTU value. They express restrictions on the space of games for which a Harsanyi NTU value can be computed. Similar hypothesis are also required for the axiomatic characterization of the Harsanyi NTU value (see Hart (1985b)).

<sup>6</sup>See Forges and Serrano (2013) for a discussion about this issue.

## 2 Formulation

### 2.1 Bayesian Cooperative Game

The model of a cooperative game with incomplete information is as follows. Let  $N = \{1, 2, \dots, n\}$  denote the set of players. For each (non-empty) coalition  $S \subseteq N$ ,  $D_S$  denotes the set of feasible joint actions for coalition  $S$ . We assume that the sets of joint actions are finite and *superadditive*, that is, for any two disjoint coalitions<sup>7</sup>  $S$  and  $R$ ,

$$D_R \times D_S \subseteq D_{R \cup S}.$$

For any player  $i \in N$ , we let  $T_i$  denote the (finite) set of possible types for player  $i$ . The interpretation is that  $t_i \in T_i$  denotes the private information possessed by player  $i$ . We use the notations<sup>8</sup>  $t_S = (t_i)_{i \in S} \in T_S = \prod_{i \in S} T_i$ ,  $t_{-i} = t_{N \setminus i} \in T_{-i} = T_{N \setminus i}$  and  $t_{-S} = t_{N \setminus S} \in T_{-S} = T_{N \setminus S}$ . For simplicity, we drop the subscript  $N$  in the case of the grand coalition, so we define  $D = D_N$  and  $T = T_N$ . We assume that players have a common prior belief  $p$  defined on  $T$ , and that all types have positive marginal probability, i.e.,  $p(t_i) > 0$  for all  $t_i \in T_i$  and all  $i \in N$ . At the interim stage each player knows his type  $t_i \in T_i$ , and hence, we let  $p(t_{-i} | t_i)$  denote the conditional probability of  $t_{-i} \in T_{-i}$  that player  $i$  infers given his type  $t_i$ .

The utility function of player  $i \in N$  is  $u_i : D \times T \rightarrow \mathbb{R}$ . As in most of the literature in cooperative game theory, we assume that coalitions are *orthogonal*, namely, when coalition  $S \subseteq N$  chooses an action which is feasible for it, the payoffs to the members of  $S$  do not depend on the actions of the complementary coalition  $N \setminus S$ . Formally,

$$u_i((d_S, d_{N \setminus S}), t) = u_i((d_S, d'_{N \setminus S}), t),$$

for every  $S \subset N$ ,  $i \in S$ ,  $d_S \in D_S$ ,  $d_{N \setminus S}, d'_{N \setminus S} \in D_{N \setminus S}$  and  $t \in T$ . Then we can let  $u_i(d_S, t)$  denote the utility for player  $i \in S$  if  $d_S \in D_S$  is carried out. That is,  $u_i(d_S, t) = u_i((d_S, d_{N \setminus S}), t)$  for any  $d_{N \setminus S} \in D_{N \setminus S}$  (recall that  $D_S \times D_{N \setminus S} \subseteq D$ ).

A *cooperative game with incomplete information* is defined by

$$\Gamma = \{N, (D_S)_{S \subseteq N}, (T_i, u_i)_{i \in N}, p\}.$$

A (*direct*) *mechanism* for the grand coalition  $N$  is a mapping  $\mu_N : T \rightarrow \Delta(D)$ , where  $\Delta(D)$  denotes the set of probability distributions over  $D$ . The interpretation is that if  $N$  forms, it makes a decision randomly as a function of its members' information. Let the set of mechanisms for  $N$  be denoted  $\mathcal{M}_N$ .

The (interim) expected utility of player  $i$  of type  $t_i$  under the mechanism  $\mu_N$  when he pretends to be of type  $\tau_i$  (while all other players are truthful) is

$$U_i(\mu_N, \tau_i | t_i) = \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) \sum_{d \in D} \mu_N(d | \tau_i, t_{-i}) u_i(d, (t_i, t_{-i})).$$

<sup>7</sup>For any two sets  $A$  and  $B$ ,  $A \subseteq B$  denotes *weak* inclusion (i.e., possibly  $A = B$ ), and  $A \subset B$  denotes strict inclusion.

<sup>8</sup>For simplicity we write  $S \setminus i$ ,  $S \cup i$  and  $D_i$  instead of the more cumbersome  $S \setminus \{i\}$ ,  $S \cup \{i\}$  and  $D_{\{i\}}$ .

As is standard, we denote  $U_i(\mu_N | t_i) = U_i(\mu_N, t_i | t_i)$ .

Players can use any communication mechanism to implement a state-contingent contract. Because information is not verifiable, the only feasible contracts are those which are induced by Bayesian Nash equilibria of the corresponding communication game. By the revelation principle (see Myerson (1991)), we can restrict attention to (Bayesian) incentive compatible direct mechanisms. Formally, a mechanism  $\mu_N$  is *incentive compatible* (for the grand coalition) if and only if

$$U_i(\mu_N | t_i) \geq U_i(\mu_N, \tau_i | t_i), \quad \forall t_i, \tau_i \in T_i, \quad \forall i \in N.$$

We denote as  $\mathcal{M}_N^*$  the set of incentive compatible mechanisms for coalition  $N$  (“\*” stands for incentive compatible as in Holmström and Myerson (1983)).

A mechanism  $\mu_N$  is (*interim*) *individually rational* if and only if

$$U_i(\mu_N | t_i) \geq \max_{d_i \in D_i} \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) u_i(d_i, t), \quad \forall t_i \in T_i, \quad \forall i \in N.$$

## 2.2 Incentive Efficiency and The Virtual Utility Approach

Following Holmström and Myerson (1983) we say that a mechanism  $\bar{\mu}_N$  for the grand coalition is (*interim*) *incentive efficient* if and only if  $\bar{\mu}_N$  is incentive compatible and there does not exist any other incentive compatible mechanism giving a strictly higher expected utility to all types  $t_i$  of all players  $i \in N$ .<sup>9</sup> Because the set of incentive-compatible mechanisms is a compact and convex polyhedron, (by the supporting hyperplane theorem) the mechanism  $\bar{\mu}_N$  is incentive efficient if and only if there exist non-negative numbers  $\lambda = (\lambda_i(t_i))_{i \in N, t_i \in T_i}$ , not all zero, such that  $\bar{\mu}_N$  is a solution to

$$\max_{\mu_N \in \mathcal{M}_N^*} \sum_{i \in N} \sum_{t_i \in T_i} \lambda_i(t_i) U_i(\mu_N | t_i) \quad (2.1)$$

We shall refer to this linear-programming problem as the *primal problem for  $\lambda$* . Let  $\alpha_i(\tau_i | t_i) \geq 0$  be the Lagrange multiplier (or dual variable) for the constraint that the type  $t_i$  of player  $i$  should not gain by reporting  $\tau_i$ . Then the Lagrangian for this optimization problem can be written as

$$\mathcal{L}(\mu_N, \lambda, \alpha) = \sum_{i \in N} \sum_{t_i \in T_i} \left( \lambda_i(t_i) U_i(\mu_N | t_i) + \sum_{\tau_i \in T_i} \alpha_i(\tau_i | t_i) [U_i(\mu_N | t_i) - U_i(\mu_N, \tau_i | t_i)] \right),$$

where  $\mu_N \in \mathcal{M}_N$ . To simplify this expression, let

$$v_i(d, t, \lambda, \alpha) = \frac{1}{p(t_i)} \left[ \left( \lambda_i(t_i) + \sum_{\tau_i \in T_i} \alpha_i(\tau_i | t_i) \right) u_i(d, t) - \sum_{\tau_i \in T_i} \alpha_i(\tau_i | t_i) \frac{p(t_{-i} | \tau_i)}{p(t_{-i} | t_i)} u_i(d, (\tau_i, t_{-i})) \right] \quad (2.2)$$

The quantity  $v_i(d, t, \lambda, \alpha)$  is called the *virtual utility* of player  $i \in N$  from the joint action  $d \in D$ , when the type profile is  $t \in T$ , w.r.t. the utility weights  $\lambda$  and the Lagrange multipliers  $\alpha$ . Then, the above Lagrangian can be rewritten as

$$\mathcal{L}(\mu_N, \lambda, \alpha) = \sum_{t \in T} p(t) \sum_{d \in D} \mu_N(d | t) \sum_{i \in N} v_i(d, t, \lambda, \alpha) \quad (2.3)$$

<sup>9</sup>We have departed slightly from the formal definition of Holmström and Myerson (1983) in using strict inequalities rather than weak inequalities and one strict inequality.



Necessary and sufficient first order conditions (from duality theory of linear programming) imply the following result:

**Proposition 1.**

An incentive compatible mechanism  $\mu_N$  is incentive efficient if and only if there exist some vectors  $\lambda \geq 0$  ( $\lambda \neq 0$ ) and  $\alpha \geq 0$ , such that

$$\alpha_i(\tau_i | t_i) [U_i(\mu_N | t_i) - U_i(\mu_N, \tau_i | t_i)] = 0, \quad \forall i \in N, \forall t_i \in T_i, \forall \tau_i \in T_i \quad (2.4)$$

and  $\mu_N$  maximizes the Lagrangian in (2.3) over all mechanisms in  $\mathcal{M}_N$ , namely,

$$\sum_{d \in D} \mu_N(d | t) \sum_{i \in N} v_i(d, t, \lambda, \alpha) = \max_{d \in D} \sum_{i \in N} v_i(d, t, \lambda, \alpha), \quad \forall t \in T \quad (2.5)$$

Equation (2.4) is the usual *dual complementary slackness* condition. Condition (2.5) says that any incentive efficient mechanism  $\mu_N$  must put positive probability weight only on the decisions that maximize the sum of the players' virtual utilities, on each information state. This implies that if players are given the possibility to transfer virtual utility, conditionally on every state, then  $\mu_N$  would be ex-post efficient<sup>10</sup>. Incentive compatibility forces each player to act as if he was maximizing a distorted utility, which magnifies the differences between his true type and the types that would be tempted to imitate him. Myerson (1984b) refers to this idea as the *virtual utility hypothesis*. A more detailed discussion about the meaning and significance of the virtual utility can be found in Myerson (1991, ch. 10).

The natural vector  $\alpha$  in this Lagrangian analysis is the vector that solves the dual problem of (2.1). This *dual problem for  $\lambda$*  can be written as

$$\min_{\alpha \geq 0} \sum_{t \in T} p(t) \left( \max_{d \in D} \sum_{i \in N} v_i(d, t, \lambda, \alpha) \right) \quad (2.6)$$

### 2.3 The M-solution

Using the concept of virtual utility, Myerson (1984a,b) generalizes the Harsanyi-Shapley fictitious transfer procedure in order to extend the Shapley NTU value to an environment with incomplete information. Specifically, for any incentive efficient mechanism  $\mu_N$  one associates a vector  $(\lambda, \alpha)$  of virtual utility scales. These scales correspond to the utility weights  $\lambda$  for which  $\mu_N$  solves the primal problem and the associated Lagrange multipliers  $\alpha$ . Then, one considers the fictitious game in which players are allowed to transfer virtual utility, conditional on every state  $t \in T$ , w.r.t. the scales  $(\lambda, \alpha)$ . In the virtual game, each intermediate coalition  $S \subset N$  commits to a *rational threat* mechanism to be carried out in case the other players refuse to cooperate with the members of  $S$ . Rational threats are the basis for computing the (virtual) worth of each coalition, and thus they determine how much credit each type of a player can claim from the proceeds of cooperation in the grand coalition. Conditionally on every state, rational threats thus define a coalitional game with transferable virtual utility. A mechanism is *equitable* for the

<sup>10</sup>This property is specially useful for practical applications, in particular when computing value allocations.

grand coalition  $N$  if it gives each type of every player his (conditional) expected Shapley TU value of the fictitious game. A precise definition is given in Section 4 (see Remark 1).

Myerson (1984b) defines the *M-solution* to be an incentive efficient mechanism  $\mu_N$  for which there exist virtual scales  $(\lambda, \alpha)$  such that  $\mu_N$  is equitable for the grand coalition. The associated interim utility allocations are called an *M-value*. A formal definition of the M-solution can be deduced from our cooperative solution concept (cf. Definition 6) by removing the egalitarian restrictions from our optimal threat criterion (see Remark 3). Two variants of the value can be considered depending on whether coalitional threats are required to be incentive compatible or not. Myerson exclusively deals with the case in which only the mechanism of the grand coalition is constrained to be (equitable and) incentive compatible. The M-solution is justified only in situations where cooperative agreements are made before a coalition structure is determined, while expecting that only the grand coalition will be forming. A detailed discussion on this issue is given in Myerson (1984b, sec 6).

### 3 Motivating Examples

In this section we study two examples which motivate the introduction of our solution concept. In both examples, it is shown that the M-value exhibits some “difficulties”; specifically, there are compelling reasons leading to an outcome not consistent with the M-value.

#### 3.1 Example 1: A Collective Choice Problem

We consider the following cooperative game with incomplete information. The set of players is  $N = \{1, 2, 3\}$ . Only player 3 has private information represented by two possible types in  $T_3 = \{H, L\}$  with prior probabilities  $p(H) = 1 - p(L) = 9/10$ . Decision options for every coalition are  $D_i = \{d_i\}$  ( $i \in N$ ),  $D_{\{1,2\}} = \{D_1 \times D_2\} \cup \{d_{12}\} = \{[d_1, d_2], d_{12}\}$ ,  $D_{\{i,3\}} = \{D_i \times D_3\} \cup \{d_{i3}^i, d_{i3}^3\} = \{[d_i, d_3], d_{i3}^i, d_{i3}^3\}$  ( $i = 1, 2$ ) and  $D_N = \{D_{\{1,2\}} \times D_3\} \cup \{D_{\{1,3\}} \times D_2\} \cup \{D_{\{2,3\}} \times D_1\}$ . A detailed interpretation will be given below. Finally, utility functions are as follows:

$(u_1, u_2, u_3)$	$L$	$H$
$[d_1, d_2, d_3]$	$(0, 0, 0)$	$(0, 0, 0)$
$[d_{12}, d_3]$	$(5, 5, 0)$	$(5, 5, 0)$
$[d_{13}^1, d_2]$	$(0, 0, 5)$	$(0, 0, 10)$
$[d_{13}^3, d_2]$	$(10, 0, -5)$	$(10, 0, 0)$
$[d_{23}^2, d_1]$	$(0, 0, 5)$	$(0, 0, 10)$
$[d_{23}^3, d_1]$	$(0, 10, -5)$	$(0, 10, 0)$

This game can be interpreted as a collective choice problem in which three individuals have the option to cooperate by investing in a work project which would benefit them. The project would cost \$10. It is commonly known that the project is worth \$10 to player 1 as well as to player 2; but its value to player 3 depends on his type, which is unknown to the other players. If 3’s type is  $H$  (“high”) then the project is worth \$10 to him. However, if 3’s type is  $L$  (“low”) then the project is only worth \$5 to him.

Decision options for all coalitions are interpreted as follows. For each player  $i \in N$ ,  $d_i$  is the only available action for himself, which leaves him with his reservation utility normalized to \$0. If coalition  $\{1, 2\}$  forms, its members may decide not to undertake the project by choosing  $[d_1, d_2]$  or they can agree on the option  $d_{12}$  which carries out the project dividing the cost on equal parts. If players 1 and 3 form a coalition, decision  $d_{13}^j$  ( $j = 1, 3$ ) denotes the option to undertake the project at  $j$ 's expense. There is no need to consider intermediate financing options, because they can be represented by randomized decisions. They may also agree on  $[d_1, d_3]$  which does not implement the project. Decision options for coalition  $\{2, 3\}$  are similarly interpreted. If all three form a coalition, they may use a random device to pick a two-person coalition which must then make a decision as above.

To analyze this game, we first consider the situation in which players 1 and 3 must reach a cooperative agreement to be implemented in case player 2 refuses to cooperate with them. In such a situation, 1 and 3 face a threat-selection subgame described by a two-person cooperative game with incomplete information that can be analyzed applying the concepts of Section 2. Assume that threats are not required to be incentive compatible. Figure 1 illustrates the set of feasible (i.e., individually rational) interim utility allocations for this (sub)game.

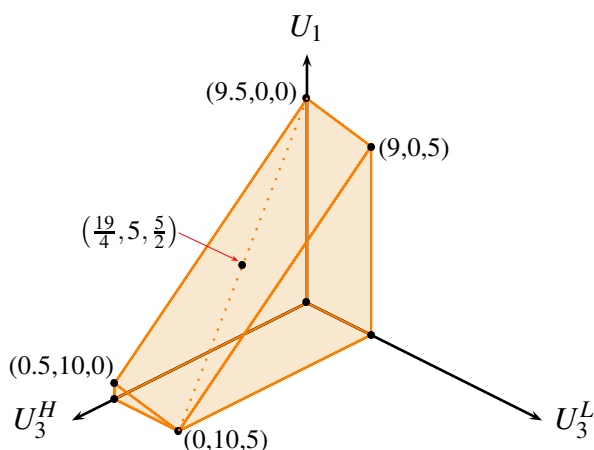


Figure 1: Feasible allocations for  $\{1, 3\}$

An equitable utility allocation in this game can be constructed as follows. Suppose that player 3 is given the right to act as a “dictator”, so that he may enforce any mechanism that is individually rational given the information that player 1 may infer from the selection of the mechanism. In this case, there is a clear decision that both types of player 3 would demand, namely,  $d_{13}^1$ . This decision implements the utility allocation  $(U_1, U_3^H, U_3^L) = (0, 10, 5)$  which gives *both* types of player 3 the highest expected utility they can get in the game. Moreover, it is efficient (see Figure 1) and *safe*, i.e., it remains individually rational no matter what player 1 can infer about 3's type from this proposal. In the terminology of Myerson (1983), it is a *strong solution*<sup>11</sup> for player 3. On the other hand, if player 1 were a dictator, then he would demand his strong solution which implements the allocation  $(19/2, 0, 0)$ . Now consider a *random-dictatorship* in which each player is given equal chance of enforcing his strong solution. Then, the interim

<sup>11</sup>A strong solution may not exist, but if so it is unique up to equivalence in utility.

efficient allocation  $(19/4, 5, 5/2) = \frac{1}{2}(0, 10, 5) + \frac{1}{2}(19/2, 0, 0)$  is equitable for  $\{1, 3\}$ . Indeed, random-dictatorship together with efficiency characterize Myerson's (1984a) generalization of the Nash bargaining solution. It is then the unique M-value for this subgame<sup>12</sup>.

The value of a player is an index based on his ability to guarantee high payoffs to all members of the coalitions to which he belongs (marginal contribution). From that perspective, player 3 should be considered as a weak player. By agreeing to cooperate with player 3, player 1 cannot expect to get more than  $19/4$  in an equitable allocation. Because players 1 and 2 are symmetric, the same reasoning is also true for a negotiation between players 2 and 3. Hence, both players 1 and 2 are better off in coalition  $\{1, 2\}$  in which case they both get 5 each, which is strictly preferred to  $19/4$ . When negotiating with player 3, 1 and 2 are adversely affected by the likely presence of 3's "bad" low type. However, by acting together players 1 and 2 face no uncertainty at all. Indeed, it is commonly known that the project is equally worth to each of them. A value allocation for our three player game should thus reward player 3 less than the other players in both states.

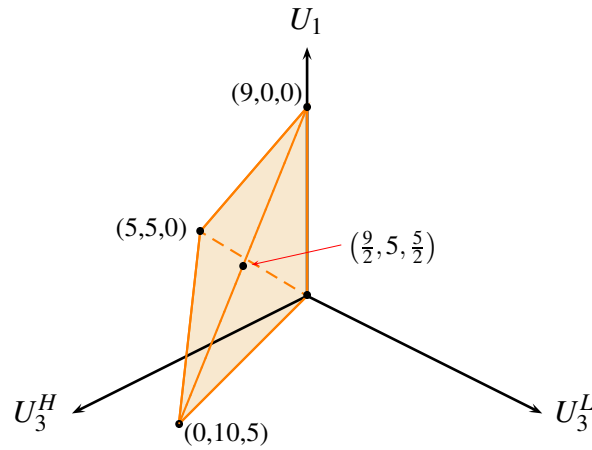


Figure 2: Incentive feasible allocations for  $\{1, 3\}$

Let us suppose now that threats are required to be incentive compatible. Figure 2 depicts the set of incentive feasible (i.e., incentive compatible and individually rational) interim utility allocations for the subgame faced by coalition  $\{1, 3\}$ . For this modified threat-selection game, the strong solution for player 3 implements again the utility allocation  $(0, 10, 5)$ .<sup>13</sup> However, the strong solution for player 1 now implements the allocation  $(9, 0, 0)$ . Proceeding as before, random-dictatorship prescribes the value allocation  $(9/2, 5, 5/2)$ .<sup>14</sup> We notice that both types of player 3 get the same expected utility in an equitable allocation regardless of whether incentive constraints are relevant or not. In contrast, 1's expected utility is reduced in the presence of

<sup>12</sup>This allocation is implemented by the mechanism  $\mu_{\{1,3\}}(d_{13}^1 | L) = 1 - \mu_{\{1,3\}}(d_{13}^3 | L) = 3/4$ ,  $\mu_{\{1,3\}}(d_{13}^1 | H) = \mu_{\{1,3\}}(d_{13}^3 | H) = 1/2$ .

<sup>13</sup>When incentive constraints matter, a safe mechanism for player 3 is one which would be incentive compatible and individually rational if player 1 knew 3's type.

<sup>14</sup>This allocation is implemented by the mechanism  $\mu_{\{1,3\}}(d_{13}^1 | L) = \mu_{\{1,3\}}([d_1, d_3] | L) = 1/2$ ,  $\mu_{\{1,3\}}(d_{13}^1 | H) = \mu_{\{1,3\}}(d_{13}^3 | H) = 1/2$ .

incentive constraints. Incentive compatibility leads to efficiency losses that are mainly beared by the uninformed party, hence increasing the incentives for 1 and 2 to form a coalition, and thus reducing the bargaining position of player 3. Therefore, 3's payoff from a value allocation in the whole game should be further reduced when coalitional incentive constraints matter.

The unique M-value of our three-player game is the utility allocation<sup>15</sup>

$$(U_1, U_2, U_3^H, U_3^L) = \left(\frac{10}{3}, \frac{10}{3}, \frac{10}{3}, \frac{5}{3}\right). \quad (3.1)$$

For instance, the incentive efficient mechanism  $\mu_N([d_{12}, d_3] | t) = \frac{2}{3}$ ,  $\mu_N([d_{23}^2, d_1] | t) = \mu_N([d_{13}^1, d_2] | t) = \frac{1}{6}$  for all  $t \in T_3$  is an M-solution. The value is supported by the utility weights<sup>16</sup>  $(\lambda_1, \lambda_2, \lambda_3^H, \lambda_3^L) = (1, 1, 9/10, 1/5)$  and the Lagrange multipliers  $(\alpha_1(L | H), \alpha_1(H | L)) = (0, 0)$ .<sup>17</sup>

An easy way to compute the M-solution in this game is simply to apply the random-dictatorship procedure to the grand coalition. The strong solution for player 3 in  $N$  implements the allocation  $(U_1, U_2, U_3^H, U_3^L) = (0, 0, 10, 5)$ . The strong solution for player 1 (resp. 2) in  $N$  implements the allocation  $(19/2, 1/2, 0, 0)$  (resp.  $(1/2, 19/2, 0, 0)$ ). Averaging these utility vectors we obtain (3.1). It is worth emphasizing that this procedure does not generally characterize the M-value. Yet for our example, it exhibits the reason why both types of player 3 extract a considerable amount of utility; namely, players are treated symmetrically. The random dictatorship procedure applied to  $N$  ignores the possibilities of cooperation among subsets of players, hence it is only acceptable when coalitions are symmetric. Indeed, Myerson's rational threats criterion cares only about the joint overall gains that can be allocated inside a coalition, but not about the way in which they are distributed. Since all coalitions can achieve the maximal gains from the project in both states of the transferable virtual utility game, the M-value treats all coalitions symmetrically. This is so even when threats are required to be incentive compatible. For instance, the mechanism that implements  $d_{j3}^j$  ( $j = 1, 2$ ) in both states is a rational threat for coalition  $\{j, 3\}$ . This mechanism however gives the whole surplus of cooperation to player 3, which is manifestly not equitable. Such a threat can be considered as being not "credible", in the sense that player  $i \notin \{j, 3\}$  could not believe that player  $j$  would agree to implement  $d_{j3}^j$  in case cooperation in  $N$  breaks down. In this example, the M-solution is insensitive to the negative externality that adverse selection exerts on 3's bargaining position.

Before proceeding with the construction of our solution concept, and for the sake of completeness, we shall briefly analyze an additional example introduced by de Clippel (2005).

### 3.2 Example 2: A Bilateral Trade Problem

Let us consider the following cooperative game with incomplete information.  $N = \{1, 2, 3\}$ ,  $T_1 = \{H, L\}$ ,  $p(H) = 1 - p(L) = 4/5$ ,  $D_i = \{d_i\}$  ( $i = 1, 2, 3$ ),  $D_{\{1,2\}} = \{[d_1, d_2], d_{12}^1, d_{12}^2\}$ ,

<sup>15</sup>We focus only on *non-degenerated* values, i.e., those which are supported by strictly positive utility weights  $\lambda$ .

<sup>16</sup>Utility weights are determined up to a positive scalar multiplication. We then normalize utility weights so that virtual utilities of the uninformed players coincide with their real utilities. This is possible since 1 and 2 are symmetric.

<sup>17</sup>Explicit computations are given in the Supplementary material.

$D_{\{1,3\}} = \{[d_1, d_3]\}$ ,  $D_{\{2,3\}} = \{[d_2, d_3]\}$ ,  $D_N = \{[d_1, d_2, d_3], [d_{12}^1, d_3], [d_{12}^2, d_3], d_{23}, d_{32}\}$  and

$(u_1, u_2, u_3)$	$[d_1, d_2, d_3]$	$[d_{12}^1, d_3]$	$[d_{12}^2, d_3]$	$d_{23}$	$d_{32}$
$H$	(0, 0, 0)	(90, 0, 0)	(0, 90, 0)	(0, 90, 0)	(0, 0, 90)
$L$	(0, 0, 0)	(30, 0, 0)	(-60, 90, 0)	(0, 30, 0)	(0, 0, 30)

The game can be interpreted as follows. Player 2 is the seller of a single good that has no value for himself. Player 1 is the only potential buyer and he has a valuation of the good that can be low (30\$), with probability 1/5, or high (90\$), with probability 4/5. Decision  $[d_1, d_2]$  represents the no-exchange alternative. Decision  $d_{12}^1$  (resp.  $d_{12}^2$ ) represents the situation where player 1 receives the good from player 2 for free (resp. in exchange of 90\$). Any other transfer of money from player 1 to player 2 (between 0\$ and 90\$) can be represented by a lottery defined on  $\{d_{12}^1, d_{12}^2\}$ . Because of the necessity to give player 1 an incentive to participate honestly, both players are limited in their abilities to share the gains from trade. Indeed, the mechanism that gives the entire surplus to player 2 in both states, is not incentive compatible. Player 3 does not generate any additional surplus from the trade. Yet, his participation partly releases players 1 and 2 from the incentive constraints they face when they cooperate. Indeed, when he joins coalition  $\{1, 2\}$  (so that the grand coalition forms), decisions  $d_{23}$  and  $d_{32}$  are added to  $D_{\{1,2\}} \times D_{\{3\}}$ . Decision  $d_{23}$  (resp.  $d_{32}$ ) gives the whole surplus to player 2 (resp. 3) in both states<sup>18</sup>.

As it is shown by de Clippel (2005), the unique M-value of this game is the interim utility allocation

$$(U_1^H, U_1^L, U_2, U_3) = (45, 15, 39, 0). \tag{3.2}$$

We observe that player 3 is considered a null player. Even though player 3 does not create any additional surplus, it would be fair to give him some positive payoff, as players 1 and 2 have to rely on him in order to weaken the incentive constraints they face. As in the previous example, requiring optimal threats to be incentive compatible does not change the M-value allocation. We conclude that the M-value is not sensitive to the informational contribution of player 3.

## 4 Equity Principles for Bayesian Cooperative Games

The Harsanyi NTU value can be characterized using two different fair allocation rules. The first of these two equity notions, introduced by Myerson (1980) under the name of *balanced contributions*, requires that, for any two members of a coalition, the amount that each player would gain by the other's participation should be equal when utility comparisons are made in some weighted utility scale. The second equity principle, denominated *subgame value equity* by Imai (1983), says that, for every coalition  $S \subseteq N$ , each player in  $S$  should obtain his Shapley TU value from the game restricted to the subcoalitions of  $S$  when utility has been made comparable in some weighted utility scale. These two equity notions are in dual relationship: for fixed utility

---

<sup>18</sup>It can be shown that when player 3 drops out of the game and coalition  $\{1, 2\}$  forms, the constraint asserting that type  $1_H$  has no incentive to report to be type  $1_L$  is binding in any incentive efficient mechanism for this coalition.

scales both allocation rules are equivalent (*equity equivalence*). In this Section, we extend this result to the case of incomplete information.

Given a vector of utility weights  $\lambda$  and a vector of Lagrange multipliers  $\alpha$ , let us consider the fictitious game in which players make interpersonal utility comparisons in the virtual utility scales  $(\lambda, \alpha)$ . In such a virtual game, each player's payoffs are represented in the virtual utility scales and virtual payoffs are transferable among the players (conditionally on every state). We assume that, as a threat during the bargaining process within the grand coalition  $N$ , each coalition  $S \subset N$  commits to some mechanism  $\mu_S : T_S \rightarrow \Delta(D_S)$ .<sup>19</sup> We denote by  $\mathcal{M}_S$  the set of mechanisms for  $S$ . Let  $\mathcal{M} = \prod_{S \subseteq N} \mathcal{M}_S$  denote the set of possible profiles of mechanisms that all various coalitions might select.

Let  $v_i(\mu_S, t, \lambda, \alpha)$  denote the linear extension of  $v_i(\cdot, t, \lambda, \alpha)$  (as defined in (2.2)) over  $\mu_S$ . We define  $W_S(\mu_S, t, \lambda, \alpha)$  as the sum of virtual utilities that the members of  $S \subseteq N$  would expect in state  $t$  when they select the mechanism  $\mu_S$ , that is

$$W_S(\mu_S, t, \lambda, \alpha) = \sum_{i \in S} v_i(\mu_S, t, \lambda, \alpha). \quad (4.1)$$

Let  $W(\eta, t, \lambda, \alpha) = (W_S(\mu_S, t, \lambda, \alpha))_{S \subseteq N}$  denote the characteristic function game when the vector of threats  $\eta = (\mu_S)_{S \subseteq N} \in \mathcal{M}$  is selected by the various coalitions<sup>20</sup> in the virtual game. For any vector  $\eta \in \mathcal{M}$ , let  $\eta_S = (\mu_R)_{R \subseteq S}$  denote its restriction to the subcoalitions of  $S$ . We define  $W|_S(\eta_S, t, \lambda, \alpha)$  as the subgame of  $W(\eta, t, \lambda, \alpha)$  obtained by restricting the domain of  $W(\eta, t, \lambda, \alpha)$  to the subsets of  $S$ . Let  $\phi$  be the Shapley TU value operator; for  $i \in S \subseteq N$ ,  $\phi_i(S, W|_S(\eta_S, t, \lambda, \alpha))$  will thus denote the Shapley TU value of player  $i$  in the subgame restricted to  $S$  when the vector of threats  $\eta_S$  is selected in the virtual game.

We denote  $V_i(\mu_S | t_i, \lambda, \alpha)$  the expected virtual utility of type  $t_i$  of player  $i \in S$  when the members of  $S$  agree on  $\mu_S$ , i.e.,

$$V_i(\mu_S | t_i, \lambda, \alpha) := \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) v_i(\mu_S, t, \lambda, \alpha). \quad (4.2)$$

**Definition 1 (Equitable mechanism).**

For any coalition  $S \subseteq N$ , the mechanism  $\mu_S$  is *equitable for  $S$*  w.r.t.  $\eta_S$ ,  $\lambda$  and  $\alpha$  if

$$V_i(\mu_S | t_i, \lambda, \alpha) = \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) \phi_i(S, W|_S(\eta_S, t, \lambda, \alpha)), \quad \forall t_i \in T_i, \forall i \in S. \quad (4.3)$$

If for all coalitions  $R \subseteq S$ ,  $\mu_R$  is equitable for  $R$  w.r.t.  $\eta_R$ ,  $\lambda$  and  $\alpha$ , then the vector of threats  $\eta_S = (\mu_R)_{R \subseteq S}$  is called *equitable* w.r.t.  $\lambda$  and  $\alpha$ .

<sup>19</sup>When a coalition  $S$  forms, it cannot rely on the information possessed by the players outside  $S$ . In other words, a communication mechanism for a coalition must be measurable with respect to the private information of its members. This is equivalent to define a mechanism as  $\mu_S : T \rightarrow \Delta(D_S)$  with  $\mu_S(t) = \mu_S(t')$  for every  $t, t' \in T$  such that  $t_S = t'_S$ .

<sup>20</sup>Strictly speaking, the component  $\mu_N \in \mathcal{M}_N$  of  $\eta$  is not a threat, since there is no coalition to threaten. However, we keep this terminology in order to simplify the exposition.

Then, a mechanism for coalition  $S$  is said to be equitable for  $S$  if it gives every type of a player in  $S$  his (conditionally) expected Shapley TU value from the virtual subgame obtained by restricting  $W(\eta, t, \lambda, \alpha)$  to the subcoalitions of  $S$ . This equity notion extends Imai's subgame value equity condition<sup>21</sup>.

**REMARK 1.** When  $S = N$ , the equality in (4.3) reduces to Myerson's (1984b) principle for equitable compromises.

**Definition 2 (Egalitarian mechanism).**

For any coalition  $S \subseteq N$ , the mechanism  $\mu_S$  is *egalitarian for  $S$*  w.r.t.  $(\mu_{S \setminus i})_{i \in S}$ ,  $\lambda$  and  $\alpha$  if

$$\sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) \sum_{j \in S \setminus i} [v_i(\mu_S, t, \lambda, \alpha) - v_i(\mu_{S \setminus j}, t, \lambda, \alpha)] = \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) \sum_{j \in S \setminus i} [v_j(\mu_S, t, \lambda, \alpha) - v_j(\mu_{S \setminus i}, t, \lambda, \alpha)], \quad \forall t_i \in T_i, \forall i \in S. \quad (4.4)$$

If for all coalitions  $R \subseteq S$ ,  $\mu_R$  is egalitarian for  $R$  w.r.t.  $(\mu_{R \setminus i})_{i \in R}$ ,  $\lambda$  and  $\alpha$ , then the vector of threats  $\eta_S = (\mu_R)_{R \subseteq S}$  is called *egalitarian* w.r.t.  $\lambda$  and  $\alpha$ .

Equation (4.4) says that the expected average virtual contribution of the different players in  $S$  to player  $i$  equals the expected average virtual contribution of player  $i$  to the different players in  $S$  as assessed by his type  $t_i$ . This egalitarian criterion generalizes Myerson's balanced contributions condition<sup>22</sup>. Indeed, when information is complete (i.e.,  $T_i$  is a singleton for every  $i \in N$ , so that we can set  $\alpha = 0$ ), condition (4.4) implies that the  $j$ -th terms on both sides are equal: the marginal contribution of  $j$  to  $i$ , measured by  $v_i(\mu_S, \lambda) - v_i(\mu_{S \setminus j}, \lambda)$ , equals the marginal contribution of  $i$  to  $j$ , symmetrically measured by  $v_j(\mu_S, \lambda) - v_j(\mu_{S \setminus i}, \lambda)$ . The same implication cannot be expected to generally hold in the case of asymmetric information. The reason is that, since negotiations take place at the interim stage, the individual probability assessments of the different types of the various players need not be the same. Then,  $i$ 's personal evaluation of  $j$ 's gains may not coincide with  $j$ 's evaluation of her own gains.

For given arbitrary vectors  $(\mu_R)_{R \subseteq S}$ ,  $\lambda$  and  $\alpha$ , equity and egalitarianism are in general two different notions of "fairness" for coalition  $S \subseteq N$ . In particular, notice that while an egalitarian mechanism  $\mu_S$  depends only on the mechanisms  $(\mu_{S \setminus i})_{i \in S}$ , an equitable mechanism depends on the whole profile of threats  $(\mu_R)_{R \subseteq S}$ . However, it turns out that if the *whole* profile  $\eta_S$  is egalitarian, then it is also equitable, and viceversa.

**Proposition 2 (Equity equivalence).**

For any coalition  $S \subseteq N$ , the vector of threats  $\eta_S = (\mu_R)_{R \subseteq S}$  is equitable (w.r.t.  $\lambda$  and  $\alpha$ ) if and only if it is egalitarian (w.r.t.  $\lambda$  and  $\alpha$ ).

This result is significant, first, in establishing a dual relationship between equity (as defined by the Shapley TU value) and the balanced contributions in environments with incomplete information. Second, and most important, Proposition 2 helps us to justify why our egalitarian criterion is (probably) the most appropriate generalization of the balanced contributions condition.

<sup>21</sup>When information is complete, so that  $T_i$  is a singleton for every  $i \in N$ , (4.3) reduces to the first condition in Proposition 6 of Imai (1983).

<sup>22</sup>It also extends the "preservation of average differences" principle introduced by Hart and Mas-Colell (1996).



When information is asymmetric, so that the probability assessments of the various types of distinct players are different, Proposition 2 cannot be deduced from the equity equivalence under complete information simply by taking (conditional) expectations. Instead we use a “consistency property” of the Shapley TU value: the value of a player is the average of his marginal contribution to the grand coalition  $W_N - W_{N \setminus i}$  and his values  $\phi_i(N \setminus j, W|_{N \setminus j})$  in the subgames with  $|N| - 1$  players (see Hart (2004, p. 39)). Apart from this clarification, the proof of Proposition 2 is straightforward .

**Proof.** We start proving the “only if” part. Let  $\eta \in \mathcal{M}$  be a vector of equitable threats (w.r.t.  $\lambda$  and  $\alpha$ ). Let  $S \subseteq N$  and  $i \in S$  be fixed. Then, for any  $j \in S \setminus i$ ,  $\mu_{S \setminus j}$  is equitable for  $S \setminus j$  (w.r.t.  $\eta_{S \setminus j}$ ,  $\lambda$  and  $\alpha$ ). Thus, for all  $t_i \in T_i$ ,

$$\sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) \sum_{j \in S \setminus i} \phi_i(S \setminus j, W|_{S \setminus j}(\eta_{S \setminus j}, t, \lambda, \alpha)) = \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) \sum_{j \in S \setminus i} v_i(\mu_{S \setminus j}, t, \lambda, \alpha). \quad (4.5)$$

Because  $\mu_S$  is equitable for  $S$  (w.r.t.  $\eta_S$ ,  $\lambda$  and  $\alpha$ ), we have that for any type  $t_i$  of a player  $i \in S$ ,

$$\begin{aligned} \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) v_i(\mu_S, t, \lambda, \alpha) &= \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) \phi_i(S, W|_S(\eta_S, t, \lambda, \alpha)) \\ &= \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) \frac{1}{|S|} \left[ W_S(\mu_S, t, \lambda, \alpha) - W_{S \setminus i}(\mu_{S \setminus i}, t, \lambda, \alpha) \right. \\ &\quad \left. + \sum_{j \in S \setminus i} \phi_i(S \setminus j, W|_{S \setminus j}(\eta_{S \setminus j}, t, \lambda, \alpha)) \right] \\ &= \frac{1}{|S|} \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) \left[ v_i(\mu_S, t, \lambda, \alpha) + \sum_{j \in S \setminus i} v_i(\mu_{S \setminus j}, t, \lambda, \alpha) \right. \\ &\quad \left. + \sum_{j \in S \setminus i} (v_j(\mu_S, t, \lambda, \alpha) - v_j(\mu_{S \setminus i}, t, \lambda, \alpha)) \right], \end{aligned} \quad (4.6)$$

where (4.5) has been used in the last. Finally, rearranging terms in (4.6) we get (4.4).

Consider now the “if” part. Let  $\eta \in \mathcal{M}$  be a vector of egalitarian threats (w.r.t.  $\lambda$  and  $\alpha$ ). For any coalition  $S \subseteq N$  and any player  $i \in S$  of type  $t_i$ , the (conditionally) expected marginal contribution of player  $i$  to coalition  $S$  is

$$\begin{aligned} &\sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) [W_S(\mu_S, t, \lambda, \alpha) - W_{S \setminus i}(\mu_{S \setminus i}, t, \lambda, \alpha)] \\ &= \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) v_i(\mu_S, t, \lambda, \alpha) + \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) \sum_{j \in S \setminus i} [v_j(\mu_S, t, \lambda, \alpha) - v_j(\mu_{S \setminus i}, t, \lambda, \alpha)] \\ &= \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) v_i(\mu_S, t, \lambda, \alpha) + \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) \sum_{j \in S \setminus i} [v_i(\mu_S, t, \lambda, \alpha) - v_i(\mu_{S \setminus j}, t, \lambda, \alpha)] \\ &= \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) \left[ |S| v_i(\mu_S, t, \lambda, \alpha) - \sum_{j \in S \setminus i} v_i(\mu_{S \setminus j}, t, \lambda, \alpha) \right], \end{aligned}$$

where the second equality is due to the fact that  $\mu_S$  is egalitarian for  $S$  w.r.t.  $(\mu_{S \setminus j})_{j \in S}$ ,  $\lambda$  and  $\alpha$ . Therefore,

$$\sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) \phi_i(S, W|_S(\eta_S, t, \lambda, \alpha)) = \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) v_i(\mu_S, t, \lambda, \alpha).$$

□

We conclude this section with a convenient characterization of an equitable mechanism for the grand coalition. It will allow us to identify the real interim utilities corresponding to an equitable allocation in the virtual game.

**Definition 3 (Warranted claims).**

Let  $(\lambda, \alpha)$  be a vector of virtual scales and  $\eta \in \mathcal{M}$  a vector of threats. The interim allocation  $\omega \in \prod_{i \in N} \mathbb{R}^{T_i}$  is *warranted* by  $\lambda$ ,  $\alpha$  and  $\eta$  if

$$\left( \lambda_i(t_i) + \sum_{\tau_i \in T_i} \alpha_i(\tau_i | t_i) \right) \omega_i(t_i) - \sum_{\tau_i \in T_i} \alpha_i(t_i | \tau_i) \omega_i(\tau_i) = \sum_{t_{-i} \in T_{-i}} p(t) \phi_i(N, W(\eta, t, \lambda, \alpha)), \quad \forall t_i \in T_i, \quad \forall i \in N. \quad (4.7)$$

The quantity  $\omega_i(t_i)$  is called the *warranted claim* of type  $t_i$  of player  $i$ .

**REMARK 2.** By Lemma 1 in Myerson (1983), the warrant equations have a unique solution in the vector of warranted claims of player  $i$ , provided that  $\lambda > 0$ . Furthermore, the solution (weakly) increases (in the vector sense) as the right-hand side is increased.

The following result follows from the equalities (3.10) and (3.11) in Myerson (1984b).

**Lemma 1.**

Let  $(\lambda, \alpha)$  be a vector of virtual scales such that  $\alpha$  is a solution of the dual for  $\lambda$ . Let  $\eta \in \mathcal{M}$  be a vector of threats such that  $\mu_N$  is a solution of the primal for  $\lambda$ . The mechanism  $\mu_N$  is equitable for  $N$  w.r.t.  $\eta$ ,  $\lambda$  and  $\alpha$  if and only if the vector of interim utilities  $U(\mu_N) := (U_i(\mu_N | t_i))_{i \in N, t_i \in T_i}$  is warranted by  $\lambda$ ,  $\alpha$  and  $\eta$ .

We can thus interpret the warrant equations: they implicitly define  $\omega$  to be the real utility allocation which would give every type of each player (in the grand coalition) his expected Shapley TU value in the virtual game.

## 5 Optimal Threats

In this section we use the equity principles previously developed in order to extend Harsanyi's (1963, sec. 9) optimal threat strategies. Specifically, we modify Myerson's (1984b) rational threats by requiring coalitional threats to meet our egalitarian criterion.

**Definition 4 (Optimal egalitarian threats).**

The mechanism  $\bar{\mu}_S \in \mathcal{M}_S$  is an *optimal egalitarian threat* for  $S \subseteq N$  w.r.t.  $(\mu_{S^c})_{i \in S}$ ,  $\lambda$  and  $\alpha$  if and only if  $\bar{\mu}_S$  is a solution to

$$\begin{aligned} \max_{\mu_S \in \mathcal{M}_S} \sum_{t \in T} p(t) W_S(\mu_S, t, \lambda, \alpha) \\ \text{s.t. (4.4)} \end{aligned} \quad (5.1)$$

The optimal threats criterion in (5.1) postulates that each coalition should maximize the ex-ante expected total virtual utility that its members would earn when coalitions commit to a vector

of egalitarian threats. In view of Proposition 2, we could also have defined an optimal threat replacing the egalitarian constraints (4.4) in (5.1) by the equity conditions in (4.3). However, this alternative definition is less tractable since threats of one coalition cannot be determined without knowledge of threats of *all* its subcoalitions<sup>23</sup>.

We notice that the maximization in (5.1) is carried out over all mechanisms in  $\mathcal{M}_S$ . However, we can alternatively require threats to be incentive compatible. A mechanism  $\mu_S$  is *incentive compatible for coalition*  $S \subseteq N$  if and only if<sup>24</sup>

$$\begin{aligned} \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) \sum_{d_S \in D_S} \mu_S(d_S | t_S) u_i(d_S, t) \\ \geq \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) \sum_{d_S \in D_S} \mu_S(d_S | \tau_i, t_{S \setminus i}) u_i(d_S, t), \quad \forall i \in S, \quad \forall t_i, \tau_i \in T_i. \end{aligned}$$

We denote as  $\mathcal{M}_S^*$  the set of incentive-compatible mechanisms for coalition  $S$ .

**Definition 5 (Incentive compatible optimal egalitarian threats).**

A mechanism  $\bar{\mu}_S \in \mathcal{M}_S$  is an *incentive compatible optimal egalitarian threat for*  $S \subseteq N$  w.r.t.  $(\mu_{S \setminus i})_{i \in S}$ ,  $\lambda$  and  $\alpha$  if and only if it solves (5.1) over all mechanisms in  $\mathcal{M}_S^*$ .

Given some virtual scales  $(\lambda, \alpha)$ , (incentive compatible) optimal egalitarian threats must be recursively constructed: for each  $S$ , given the threats  $(\bar{\mu}_{S \setminus i})_{i \in S}$ ,  $\bar{\mu}_S$  is determined solving (5.1). This recursion leads to a profile of threats  $\bar{\eta} = (\bar{\mu}_S)_{S \subseteq N}$  which we call an (*coalitionally incentive compatible*) *egalitarian solution* w.r.t.  $\lambda$  and  $\alpha$ .<sup>25</sup>

Myerson (1984b, sec. 6) argues that maximizing the ex-ante expected virtual worth of a coalition is appropriate in games where only the mechanism chosen by the grand coalition will be implemented. In such a situation, the final payoffs are granted by the grand coalition and therefore the mechanisms  $(\mu_S)_{S \subseteq N}$  need not be either equitable or incentive compatible. Thus, Myerson's (1984b) rational threats maximize the objective function in (5.1) constrained only by the feasibility of the mechanisms, i.e.,  $\mu_S \in \mathcal{M}_S$ . Even if we agree with this reasoning, the examples in Section 3 illustrate situations in which some relevant aspects of the intermediate coalitions are ignored by Myerson's rational threat criterion. In contrast, we think that for a mechanism  $\mu_S$  to constitute an appropriate measure of the strength of coalition  $S$ , it must be equitable regardless of whether it is expected to be implemented or not. It should be clear that a vector  $\eta = (\mu_S)_{S \subseteq N}$  of (inductively constructed) optimal egalitarian threats is egalitarian. Therefore, by Proposition 2,  $\eta$  is also equitable. This reasoning is summarized in the following proposition.

**Proposition 3.**

Let  $\eta = (\mu_S)_{S \subseteq N}$  be an (coalitionally incentive compatible) egalitarian solution w.r.t.  $\lambda$  and  $\alpha$ . Then, for each coalition  $S \subseteq N$ ,  $\eta_S$  is equitable w.r.t.  $\lambda$  and  $\alpha$ .

<sup>23</sup>A definition like that would be consistent with Imai's (1983) characterization of the Harsanyi NTU value.

<sup>24</sup>Clearly, when  $S = N$  this definition coincides with the one introduced in Section 2.

<sup>25</sup>Egalitarian solutions generalize the monotonic solutions introduced by Kalai and Samet (1985) to games with incomplete information.

## 6 The H-Solution

In this section we apply the ideas developed in the preceding sections to construct an egalitarian-based cooperative solution.

### Definition 6 (H-solution).

A mechanism  $\bar{\mu}_N \in \mathcal{M}_N$  is an *H-solution* if and only if there exist vectors  $\lambda > 0$ ,  $\alpha \geq 0$  and  $\eta = (\mu_S)_{S \subseteq N} \in \mathcal{M}$  with  $\mu_N = \bar{\mu}_N$  such that

- (i)  $\mu_N$  is a solution of the primal problem for  $\lambda$ .
- (ii)  $\alpha$  is a solution of the dual problem for  $\lambda$ .
- (iii) For each  $S \subset N$ ,  $\mu_S$  is an optimal egalitarian threat for  $S$  w.r.t.  $(\mu_{S \setminus i})_{i \in S}$ ,  $\lambda$  and  $\alpha$ .
- (iv)  $\mu_N$  is an egalitarian threat for  $N$  w.r.t.  $(\mu_{N \setminus i})_{i \in N}$ ,  $\lambda$  and  $\alpha$ .

The vector of interim utilities  $U(\bar{\mu}_N)$  is called an *H-value*.

Alternatively, a bargaining solution can be defined replacing condition (iii) by

- (iii') For each coalition  $S \subset N$ ,  $\mu_S$  is an incentive compatible optimal egalitarian threat for  $S$  w.r.t.  $(\mu_{S \setminus i})_{i \in S}$ ,  $\lambda$  and  $\alpha$ .

In that case an H-solution is called *coalitionally incentive compatible*.

Conditions (i) – (iv) in our definition of an H-value have natural interpretations: (i) generalizes the  $\lambda$ -weighted utilitarian criterion, (ii) says that  $\alpha$  is the vector of Lagrange multipliers associated with (i), and (iii) extends Harsanyi's (1963) optimal threats criterion to games with incomplete information. It follows from (i) that  $\mu_N$  maximizes the Lagrangian in (2.3). Hence, (i) and (iv) imply that  $\mu_N$  is also an optimal egalitarian threat for  $N$  w.r.t.  $(\mu_{N \setminus i})_{i \in N}$ ,  $\lambda$  and  $\alpha$  (i.e.,  $\mu_N$  solves (5.1) for  $N$ ). Therefore, the whole vector of threats  $\eta = (\mu_S)_{S \subseteq N}$  is an egalitarian solution.

**REMARK 3.** By Lemma 1, we can equivalently define an H-solution replacing condition (iv) by

- (iv')  $U(\mu_N)$  is warranted by  $\lambda$ ,  $\alpha$  and  $\eta$ .

This equivalent definition puts the H-solution in a form that makes it easily comparable with the M-solution. Indeed, with (iv'), the H-solution differentiates itself from the M-solution only in that the latter does not require optimal threats to meet our egalitarian criterion. Then, it follows that when  $n = 2$  both solution concepts coincide<sup>26</sup>.

### Theorem 1 (Two-player games).

Let  $\Gamma$  be a two-player cooperative game with incomplete information. Then, any (non-degenerated) M-solution is an H-solution and viceversa. Moreover, if  $\Gamma$  is a two-person bargaining problem<sup>27</sup>, both solution concepts coincide with Myerson's (1984a) generalization of the Nash bargaining solution.

<sup>26</sup>Singleton coalitions are not constrained by the egalitarian restrictions in (4.4)

<sup>27</sup>A two-person bargaining problem is a cooperative game satisfying:  $n = 2$ ,  $D_i = \{d_i\}$  for all  $i \in N$  and  $u_i(d^*, t) = 0$  for all  $i \in N$  and  $t \in T$ , where  $d^* := [d_i, d_j]$  is the disagreement outcome.

**Theorem 2 (Individual rationality).**

Both variants of the H-bargaining solution are interim individually rational.

**Proof.** Let  $\mu_N$  be an H-solution supported by  $\eta$ ,  $\lambda$  and  $\alpha$ . For each  $i \in N$ , let  $\hat{\mu}_i \in \mathcal{M}_i$  be defined by

$$\sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) \sum_{d_i \in D_i} \hat{\mu}_i(d_i | t_i) u_i(d_i, t) = \max_{d_i \in D_i} \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) u_i(d_i, t), \quad \forall t_i \in T_i, \quad (6.1)$$

For each  $t \in T$ , the TU game  $W(\eta, t, \lambda, \alpha)$  is weakly superadditive<sup>28</sup> (since decision sets are superadditive). Then,  $\phi_i(N, W(\eta, t, \lambda, \alpha)) \geq v_i(\mu_i, t, \lambda, \alpha)$  for every  $t \in T$ . Also, for all  $i \in N$ ,  $\sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) v_i(\mu_i, t, \lambda, \alpha) \geq \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) v_i(\hat{\mu}_i, t, \lambda, \alpha)$  for all  $t_i \in T_i$ , since  $\mu_i$  is an optimal egalitarian threat for  $i$ . Then, we have that for each  $i \in N$  and  $t_i \in T_i$ ,

$$\begin{aligned} & \left( \lambda_i(t_i) + \sum_{\tau_i \in T_i} \alpha_i(\tau_i | t_i) \right) U_i(\mu_N | t_i) - \sum_{\tau_i \in T_i} \alpha_i(t_i | \tau_i) U_i(\mu_N | \tau_i) \\ &= \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) \phi_i(N, W(\eta, t, \lambda, \alpha)) \\ &\geq \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) v_i(\hat{\mu}_i, t, \lambda, \alpha) \\ &\geq \left( \lambda_i(t_i) + \sum_{\tau_i \in T_i} \alpha_i(\tau_i | t_i) \right) \max_{d_i \in D_i} \sum_{t_{-i} \in T_{-i}} p(t_{-i} | t_i) u_i(d_i, t) \\ &\quad - \sum_{\tau_i \in T_i} \alpha_i(t_i | \tau_i) \max_{d_i \in D_i} \sum_{t_{-i} \in T_{-i}} p(t_{-i} | \tau_i) u_i(d_i, (\tau_i, t_{-i})), \end{aligned} \quad (6.2)$$

where the first line follows from the fact that  $U(\mu_N)$  is warranted by  $\eta$ ,  $\lambda$  and  $\alpha$  (cf. condition (iv')); the second line follows from the first part of the proof; and finally, the last inequality uses (6.1) applied to  $\tau_i$ . The desired conclusion is obtained from (6.2) together with Remark 2.  $\square$

The following result follows directly from the definitions (cf. Harsanyi (1963, sec. 10)).

**Theorem 3 (Generalization of the Harsanyi NTU value).**

Let  $\Gamma$  be a cooperative game with complete information, i.e.,  $T_i$  is a singleton for every  $i \in N$ . If  $\bar{\mu}_N$  is an H-bargaining solution, then the utility allocation  $U(\bar{\mu}_N)$  is a Harsanyi NTU value of  $\Gamma$ . Conversely, if the utility allocation  $\bar{U} = (\bar{U}_i)_{i \in N}$  is a (non-degenerated) Harsanyi NTU value of  $\Gamma$ , then there exists an H-bargaining solution of  $\Gamma$ ,  $\bar{\mu}_N$ , such that  $\bar{U} = U(\bar{\mu}_N)$ .

We are now ready to compute our bargaining solution for the examples introduced in Section 3.

**6.1 Example 1**

Let us consider the vector of utility weights  $\bar{\lambda} = (\bar{\lambda}_1, \bar{\lambda}_2, \bar{\lambda}_3^H, \bar{\lambda}_3^L) = (1, 1, 9/10, 1/5)$ . First, we notice that for any feasible mechanism  $\mu_N \in \mathcal{M}_N$  we have that

$$U(\mu_N, \bar{\lambda}) := U_1(\mu_N) + U_2(\mu_N) + \frac{9}{10} U_3(\mu_N | H) + \frac{1}{5} U_3(\mu_N | L) \leq 10, \quad (6.3)$$

<sup>28</sup>A TU game  $(N, W)$  is *weakly superadditive* if and only if for each player  $i \in N$ ,  $W(S \setminus i) + W(\{i\}) \leq W(S)$  for all coalitions  $S \subseteq N$  containing  $i$ . Clearly, by definition of the Shapley TU value, weak superadditivity implies that  $\phi_i(N, W) \geq W(\{i\})$  for every  $i \in N$ .

Consider now the problem of finding the best incentive compatible and individually rational utility allocation for each possible type of every player. Straightforward computations yield that the best allocation for player 1 is  $(U_1, U_2, U_3^H, U_3^L) = (19/2, 1/2, 0, 0)$ . By symmetry, the best allocation for player 2 is  $(1/2, 19/2, 0, 0)$ . Finally,  $(0, 0, 10, 5)$  is simultaneously the best allocation for *both* types of player 3. These three allocations are incentive efficient, and they lie on the hyperplane  $U(\mu_N, \bar{\lambda}) = 10$ . Then, by convexity of  $\mathcal{M}_N^*$ , any individually rational and incentive efficient mechanism  $\mu_N$  must satisfy  $U(\mu_N, \bar{\lambda}) \geq 10$ . Thus, (6.3) implies that the incentive efficient frontier coincides with the hyperplane  $U(\mu_N, \bar{\lambda}) = 10$  on the individually rational zone. Therefore, in view of Theorem 2, condition (i) implies that a value allocation can only be supported by the utility weights  $\bar{\lambda}$ .<sup>29</sup>

The utility weights  $\bar{\lambda}$  reflect the optimal inter-type compromise between both types of player 3. To conceal his type, player 3 must achieve a balance that puts extra weight on the payoff maximization goals of type  $L$  (*inscrutability principle*). This is what explains that  $\bar{\lambda}_3^L$  differs from the prior probability  $p(L)$  by scaling up the actual utility of type  $L$ . On the other hand, the optimal value of the dual variables in the dual problem for  $\bar{\lambda}$  is  $(\bar{\alpha}_3(L | H), \bar{\alpha}_3(H | L)) = (0, 0)$ .

Given these virtual scales, it can be easily verified that the only H-value of this game is<sup>30</sup>

$$(U_1, U_2, U_3^H, U_3^L) = \left(\frac{61}{18}, \frac{61}{18}, \frac{60}{18}, \frac{20}{18}\right). \quad (6.4)$$

The value allocation gives less to *both* types of player 3, compared to what they get players 1 and 2. This is due to the fact that, by requiring optimal threats to satisfy our egalitarian criterion, coalitions  $\{1, 3\}$  and  $\{2, 3\}$  cannot agree to fully distribute the total gains of cooperation in state  $L$ . Indeed, because players in coalition  $\{i, 3\}$  ( $i = 1, 2$ ) are constrained to choose a feasible allocation giving them equal gains (in the virtual utility scales), then they have to settle for a sum of payoffs of at most  $\$20/3 (< \$10)$  in state  $L$ . This implies that, in a two-person coalition with 3, players 1 and 2 cannot expect to get more than  $\$29/6 (< \$5)$  each. Hence, the expected “marginal contribution” of player 3 in a two-person coalition with him is strictly lower than what 1 and 2 can get by acting together. Consequently, 3 is perceived to have a weak bargaining position. It then appears that the H-value reflects the game situation better than the M-value.

The asymmetry reflected in the allocation (6.4) comes uniquely from the fact that players 1 and 2 are adversely affected by 3’s low type. None of the inefficiencies created by the incentive compatibility is taken into account: incentive constraints are not essential for the grand coalition (i.e., incentive constraints do not impose any restriction for achieving an ex-post efficient allocation) and optimal egalitarian threats are not required to be incentive compatible. The unique coalitionally incentive compatible H-value of this game is

$$(U_1, U_2, U_3^H, U_3^L) = \left(\frac{41}{12}, \frac{41}{12}, \frac{40}{12}, \frac{10}{12}\right). \quad (6.5)$$

When we take account of the incentive constraints that coalitions  $\{1, 3\}$  and  $\{2, 3\}$  face, our bargaining solution gives much less to player 3 in both states compared to the situation in which incentive constraints only matter for the grand coalition (compare (6.4) and (6.5)). In

<sup>29</sup>The same utility weights support the unique M-value (see Section 3).

<sup>30</sup>Detailed computations are provided in the Supplementary material.

fact, when coalition  $\{i, 3\}$  (with  $i = 1, 2$ ) is required to choose a mechanism that is incentive compatible, its members cannot agree on a virtual utility allocation giving them equal gains without an efficiency loss. Thus player 3's bargaining ability is further lowered by the necessity for players to trust each other.

All in all, it seems that, in this particular game, our solution concept provides much more agreement with what we expect the outcome to be.

## 6.2 Example 2

Proceeding as in Example 1, it can be shown that any incentive compatible and individually rational mechanism is incentive efficient if and only if it satisfies

$$\frac{4}{5}U_1(\mu_N | H) + \frac{1}{5}U_1(\mu_N | L) + U_2(\mu_N) + U_3(\mu_N) = 78, \quad (6.6)$$

The natural vector of utility weights is thus  $\bar{\lambda} = (\bar{\lambda}_1^H, \bar{\lambda}_1^L, \bar{\lambda}_2, \bar{\lambda}_3) = (4/5, 1/5, 1, 1)$ . For these utility weights, the corresponding dual variables are  $(\bar{\alpha}_1(L | H), \bar{\alpha}_1(H | L)) = (0, 0)$ . Then, we conclude that incentive constraints do not matter for the grand coalition. As it was previously discussed in Section 3, the participation of player 3 in the grand coalition releases players 1 and 2 from the incentive constraints they face in coalition  $\{1, 2\}$ . Unlike Example 1, here utility weights and prior probabilities coincide. This is so because player 3 allows 1 and 2 to fully distribute the gains from trade. Types are then essentially verifiable, as any transfer of utility can be implemented by a utility equivalent incentive compatible mechanism.

Given these virtual scales, it can be checked that the interim allocation in (3.2) is also the unique H-value of this game. Both the M-value and the H-value coincide because the virtual value of coalition  $\{1, 2\}$  is computed while using the vector  $(\lambda, \alpha)$  as specified for the grand coalition. By doing so, we act as if incentive constraints do not matter for coalition  $\{1, 2\}$ , although they do.

By imposing incentive constraints for all intermediate coalitions, we have that the unique coalitionally incentive compatible H-value of this game is the allocation

$$(U_1, U_2, U_3^H, U_3^L) = (45, 13, 38.6, 0.8). \quad (6.7)$$

The H-value generates an interesting alternative to the M-value in de Clippel's example. This game however also puts in evidence some "difficulties" with our bargaining solution. First, notice that while it is the case that the coalitionally incentive compatible value allocation rewards player 3, it is as if *both* players 1 and 2 pay \$0.8 to player 3 in exchange of his service. This may be considered as not reasonable since only player 2 needs the help of player 3 in order to extract the whole cooperative surplus in both states. Second, the virtual worths of all coalitions in our bargaining solution are computed using the vector  $(\lambda, \alpha)$  specified for the grand coalition. As a consequence, the efficiency losses due to the incentive compatibility at the level of all subcoalitions are not taken into account, unless incentive constraints are explicitly required.

It turns out that both examples presented in this paper are similar in nature, and that the H-solution prescribes intuitively appealing outcomes in each case.

## 7 Some Comments About the (Non-)Existence of the H-solution

The H-solution is characterized by strong equity conditions that may lead to its non-existence in some cases. In this section we shall exhibit an example of a 4-player cooperative game with *complete* information in which there is no H-solution. The following hinders the existence of the H-solution in this example: first, optimal egalitarian threats do not exist for some utility weights; second, optimal egalitarian threats vary discontinuously with the utility weights, which makes impossible the consistency of conditions (i) and (iv) in the definition of the H-solution. This example can be used to construct a game with incomplete information satisfying the same properties. The method is outlined in footnote 31 below. We study instead the game with complete information, this being however easier to analyze. Finally, we discuss the reasons why the methods and techniques used to obtain existence results of the Harsanyi NTU value cannot be well adapted to games with incomplete information.

### 7.1 Example 3: Non-existence of the H-solution

Let  $\Gamma_C$  be the following cooperative game (with complete information): the set of players is  $N = \{1, 2, 3, 4\}$ . Decision options for every coalition are  $D_i = \{d_i\}$  ( $i \in N$ ),  $D_{\{1,2\}} = \{D_1 \times D_2\} \cup \{d_{12}\} = \{[d_1, d_2], d_{12}\}$ ,  $D_{\{1,3\}} = \{D_1 \times D_3\} = \{[d_1, d_3]\}$ ,  $D_{\{2,3\}} = \{D_2 \times D_3\} \cup \{d_{23}\} = \{[d_2, d_3], d_{23}\}$ ,  $D_{\{1,2,3\}} = \{D_1 \times D_{\{2,3\}}\} \cup \{D_{\{1,2\}} \times D_3\}$ ,  $D_{S \cup \{4\}} = \{D_S \times D_4\}$  ( $S \subset N \setminus \{4\}$ ) and  $D_N = \{D_{\{1,2,3\}} \times D_4\} \cup \{d_N^1, d_N^2, d_N^3, d_N^4\}$ . Finally, utility functions are as follows:

$d_N$	$(u_1, u_2, u_3, u_4)$
$[d_1, d_2, d_3, d_4]$	$(0, 0, 0, 0)$
$[d_{12}, d_3, d_4]$	$(2, 2, 0, 0)$
$[d_1, d_{23}, d_4]$	$(0, 1, 1, 0)$
$d_N^1$	$(-1, 3, 3, 3)$
$d_N^2$	$(3, -1, 3, 3)$
$d_N^3$	$(3, 3, -1, 3)$
$d_N^4$	$(3, 3, 3, -1)$

We notice that decisions  $[d_1, d_2, d_3, d_4]$ ,  $[d_{12}, d_3, d_4]$  and  $[d_1, d_{23}, d_4]$  are strictly Pareto dominated. Hence, the Pareto frontier (of the grand coalition) is a bounded surface contained in the hyperplane

$$u_1 + u_2 + u_3 + u_4 = 8.$$

Natural utility weights supporting an H-solution are  $\bar{\lambda}_i = 1$  for all  $i \in N$ . However, no egalitarian solution can be constructed for  $\bar{\lambda}$ . Let us see this. For coalitions consisting of a single player  $i$  ( $i \in N$ ), the unique optimal egalitarian threat is the (deterministic) mechanism  $d_i$ . Similarly, coalition  $\{1, 2\}$  has a unique optimal egalitarian threat given by the (deterministic) mechanism  $d_{12}$ . For coalition  $\{2, 3\}$  (resp.  $\{1, 3\}$ ) we have that the unique optimal egalitarian threat is  $d_{23}$



(resp.  $[d_1, d_3]$ ). Then, the egalitarian restrictions in (4.4) for coalition  $S = \{1, 2, 3\}$  reduce to:

$$u_1(\mu_S) - u_3(\mu_S) = 1 \quad (7.1)$$

$$u_2(\mu_S) - u_1(\mu_S) = 1 \quad (7.2)$$

Condition (7.2) implies that  $\mu_S([d_{12}, d_3]) = 1 - \mu_S([d_1, d_{23}]) = 0$ . However, (7.1) requires that  $\mu_S([d_{12}, d_3]) > 0$ , which is a contradiction. We conclude that no egalitarian solution exists w.r.t.  $\bar{\lambda}$ .

**Proposition 4.**

The game  $\Gamma_C$  has no H-solution. This holds even if we allow some (but not all) utility weights to vanish.<sup>31</sup>

A detailed proof is given in the Appendix. A further difficulty prevents the existence of an H-solution in this game, namely, the optimal solution correspondence of (5.1) may not be upper-hemicontinuous in the utility weights. Consider, for instance, coalition  $\{1, 2\}$ . The set of feasible expected utility allocations for  $\{1, 2\}$  is given by the line segment  $\vec{ow}$  illustrated in Figure 3.

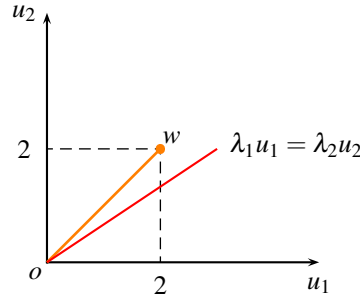


Figure 3: Feasible allocations for  $\{1, 2\}$

For any vector  $\lambda > 0$  such that  $\lambda_1/\lambda_2 \neq 1$ , the unique optimal egalitarian threat is  $\bar{\mu}_{\{1,2\}}(d_{12}) = 1 - \bar{\mu}_{\{1,2\}}([d_1, d_2]) = 0$ , this being the unique feasible mechanism satisfying the egalitarian constraint  $\lambda_1 u_1 = \lambda_2 u_2$ . The corresponding utility allocation is  $o$ . However, when  $\lambda_1/\lambda_2 = 1$ , the unique optimal egalitarian threat is  $\tilde{\mu}_{\{1,2\}}(d_{12}) = 1 - \tilde{\mu}_{\{1,2\}}([d_1, d_2]) = 1$ , achieving the utility allocation  $w$ . We conclude that the optimal solutions correspondence of (5.1) for  $S = \{1, 2\}$ , viewed as a function of  $\lambda_1/\lambda_2$ , is discontinuous. As stated earlier, this lack of continuity impedes conditions (i) and (iv) to be simultaneously satisfied as the utility weights  $\lambda$  accommodate. This issue can only be appreciated while exhaustively analyzing conditions (i) and (iv) for all values of  $\lambda$  (see proof of Proposition 4).

<sup>31</sup> The game  $\Gamma_C$  can be used to construct a game with incomplete information for which there is no H-solution. Let  $N$  and  $(D_S)_{S \subseteq N}$  be defined as in  $\Gamma_C$ . For each  $i = 1, 2, 3$ , let  $T_i$  be a singleton. Player 4 has private information in the form of two possible types  $T_4 = \{A, B\}$  with prior probabilities  $q(A) = 1 - q(B) > 0$ . Utility functions are defined as follows:  $w_i(d_N, A) = u_i(d_N)$  and  $w_i(d_N, B) = \beta u_i(d_N)$  (with  $\beta > 0$ ), where  $(u_i)_{i \in N}$  is defined as in  $\Gamma_C$ . Then, the game  $\Gamma_I = \{N, (D_S)_{S \subseteq N}, (w_i, T_i)_{i \in N}, q\}$  has no H-solution. Indeed, because the incentives of player 4 are fully aligned in both states, incentive constraints are not essential. Thus, we can set the Lagrange multipliers to be  $\alpha_4(A | B) = \alpha_4(B | A) = 0$ . Virtual utilities then reduce to  $\lambda$ -weighted utilities. The rest of the analysis follows, *mutatis mutandis*, the same reasoning as for  $\Gamma_C$ .

### 7.2 Free Disposal and the Structure of Incentives

When information is complete, the above difficulties are ruled out by considering games whose characteristic function is comprehensive (“free disposal” assumption). Then, one is tempted to accommodate free disposal activities by introducing decisions in each  $D_S$  specifying how much utility a player may discard. This has no significant consequence when information is complete, however under asymmetric information, adding new decisions may change the incentive structure of the game: free disposal can be used for signaling purposes, i.e., for weakening incentive compatibility. As a result, for any interim utility allocation on the interim incentive efficient frontier (of the grand coalition), we cannot *generally* extend the original game by introducing additional decisions allowing players to discard utility (conditional on every state), while leaving the original utility allocation efficient in the expanded problem<sup>32</sup>. In order to illustrate this issue, consider again the (sub)game faced by players 1 and 3 in Example 1. Assume now that player 3 is allowed to dispose of his utility in state  $H$ . Specifically, let  $\tilde{d}$  be such that  $u_3(\tilde{d}, H) = 0$ ,  $u_3(\tilde{d}, L) = 5$  and  $u_1(\tilde{d}, H) = u_1(\tilde{d}, L) = 0$ . Decision  $\tilde{d}$  is equivalent to implement decision  $d_{13}^1$  but then player 3 agrees to discard 10 units of his utility in state  $H$ . Now consider the expanded problem with decision set  $\tilde{D}_{\{1,3\}} = D_{\{1,3\}} \cup \{\tilde{d}\}$ . The new set of incentive feasible interim utility allocations is depicted in Figure 4.

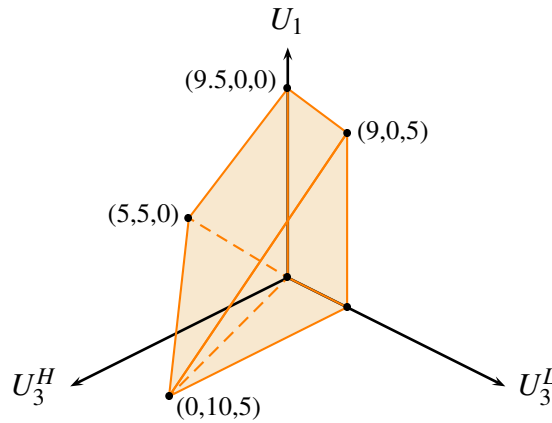


Figure 4: Incentive feasible allocations for  $\{1, 3\}$  in the expanded problem

When comparing Figures 2 and 4, we observe that the game has substantially changed after  $\tilde{d}$  was introduced. As required, type  $H$  of player 1 can now achieve all the allocations in which he discards any nonnegative amount of utility. However, permitting free disposal facilitated also the fulfillment of incentive constraints, thus allowing both players to achieve higher interim utilities with respect to the original problem<sup>33</sup>. In particular, any incentive efficient allocation in the expanded game is ex-post efficient, which is not the case in the original game (cf. Figure

<sup>32</sup>Clearly, this issue is not present in  $\Gamma_C$ , this being a game with complete information. Neither is it in  $\Gamma_I$ , as incentive constraints are not essential in this game. Nevertheless, in more general games in which incentives constraints are binding, the same difficulties are also encountered. In that cases, in addition to the utility weights  $\lambda$ , also the dual variables  $\alpha$  have to be taken into consideration. Exemplify such situations is, however, more difficult due to the endogenous nature of the dual variables.

<sup>33</sup>In this game, only type  $H$  has incentives to impersonate type  $L$ . In an effort to distinguish himself from type  $H$ , type  $L$  may agree on a mechanism that discards an appropriate amount of utility in state  $H$ . Clearly, this

1). This implies that incentive constraints are not essential in the enlarged problem. In addition, the supporting utility weights to the Pareto frontier of the expanded bargaining problem are no longer the same as in the original problem<sup>34</sup>.

## 8 Appendix: Proof of Proposition 4

Let  $\mu_N$  be an H-solution of  $\Gamma_C$  supported by  $\lambda$  and  $\eta = (\mu_S)_{S \subseteq N}$ . We verify recursively conditions (i) – (iv). Because  $\Gamma_C$  has complete information, there are no incentive constraints, which is equivalent to set  $\alpha = 0$ , so that virtual utility reduces to  $\lambda$ -weighted utility and the egalitarian criterion in (4.4) becomes

$$\lambda_i (u_i(\mu_S) - u_i(\mu_{S \setminus j})) = \lambda_j (u_j(\mu_S) - u_j(\mu_{S \setminus i})), \quad \forall i, j \in S. \quad (8.1)$$

For coalitions consisting of a single player  $i$ , it is clear that  $u_i(\mu_i) = 0$ . For all two-person coalitions containing player 4,  $D_{\{i,4\}} = \{[d_i, d_4]\}$ . Then, it follows immediately that  $u_i(\mu_{\{i,4\}}) = u_4(\mu_{\{i,4\}}) = 0$  for all  $i \in N \setminus 4$ . Similarly,  $D_{\{1,3\}} = \{[d_1, d_3]\}$ , thus  $u_1(\mu_{\{1,3\}}) = u_3(\mu_{\{1,3\}}) = 0$ . Consider now coalition  $\{1,2\}$ . It can be easily verified that an optimal egalitarian threat for this coalition satisfies:

$$(u_1(\mu_{\{1,2\}}), u_2(\mu_{\{1,2\}})) = \begin{cases} (2, 2), & \text{if } \lambda_1 = \lambda_2 > 0 \\ (y, y), y \in [0, 2], & \text{if } \lambda_1 = \lambda_2 = 0 \\ (0, 0), & \text{if } \lambda_1 \neq \lambda_2 \end{cases}$$

Similarly,

$$(u_2(\mu_{\{2,3\}}), u_3(\mu_{\{2,3\}})) = \begin{cases} (1, 1), & \text{if } \lambda_2 = \lambda_3 > 0 \\ (y, y), y \in [0, 1], & \text{if } \lambda_2 = \lambda_3 = 0 \\ (0, 0), & \text{if } \lambda_2 \neq \lambda_3 \end{cases}$$

We proceed now by cases.

**Case 1:**  $\lambda_1 = \lambda_2 = \lambda_3 > 0$  Condition (8.1) applied to  $S = \{1, 2, 3\}$  lead to the equations (7.1) and (7.2). We have already shown in Section 7 that these two equations are inconsistent.

**Case 2:**  $\lambda_1 = \lambda_2 > 0, \lambda_2 \neq \lambda_3$  Without loss of generality, we can set  $\lambda_1 = \lambda_2 = 1$ . It can be easily verified that, (8.1) implies:  $u_1(\mu_{\{1,2,3\}}) = u_2(\mu_{\{1,2,3\}}) = u_1(\mu_{\{1,2,4\}}) = u_2(\mu_{\{1,2,4\}}) = 2$

---

“commitment strategy” does not affect type  $L$ , but is harmful for type  $H$ . Therefore, type  $H$  will never accept any such mechanism. In this way, types become essentially verifiable. This example shares some features in common with an exchange economy with differential information proposed by Forges, Mertens and Vohra (2002, sec. 2.5).

<sup>34</sup>It is worth noticing that in two-person games the H-solution always exists. This follows from the existence of the M-solution, since by Theorem 1 both solutions coincide whenever  $n = 2$ . The issue illustrated in the previous example does not bring any difficulty for the existence of the H-solution in two-person games. The reason is that, equity imposes no restrictions for singleton coalitions. Hence, in this case, allowing for free disposal is no longer necessary for guaranteeing existence and (upper-hemi)continuity of the optimal egalitarian threats.

and  $u_3(\mu_{\{1,2,3\}}) = u_4(\mu_{\{1,2,4\}}) = u_2(\mu_{\{2,3,4\}}) = u_3(\mu_{\{2,3,4\}}) = u_4(\mu_{\{2,3,4\}}) = 0$ . Then, condition (8.1) applied to  $N$  reduces to

$$u_1(\mu_N) = u_2(\mu_N) \quad (8.2a)$$

$$u_1(\mu_N) = \lambda_3 u_3(\mu_N) + 2 \quad (8.2b)$$

$$u_1(\mu_N) = \lambda_4 u_4(\mu_N) + 2 \quad (8.2c)$$

On the other hand, we have that

$$u_1(d_N) + u_2(d_N) + \lambda_3 u_3(d_N) + \lambda_4 u_4(d_N) = \begin{cases} 2 + 3(\lambda_3 + \lambda_4), & \text{if } d_N = d_N^1, d_N^2 \\ 6 - \lambda_3 + 3\lambda_4, & \text{if } d_N = d_N^3 \\ 6 + 3\lambda_3 - \lambda_4, & \text{if } d_N = d_N^4 \end{cases} \quad (8.2d)$$

**Subcase 2.1:**  $\lambda_3 > 1, \lambda_4 > 1$  Condition (i) implies that  $\mu_N(d_N^3) = \mu_N(d_N^4) = 0$ . But then, (8.2a) requires that  $\mu_N(d_N^1) = \mu_N(d_N^2) = 1/2$ . Hence,  $u_1(\mu_N) = u_2(\mu_N) = 1$  and  $u_3(\mu_N) = u_4(\mu_N) = 3$ . With this, (8.2b) reduces to  $\lambda_3 = -1/3$ , which is a contradiction.

**Subcase 2.2:**  $\lambda_3 > 1, \lambda_4 = 1$  Condition (i) implies that  $\mu_N(d_N^3) = 0$ . But then,  $u_1(\mu_N) < \lambda_3 u_3(\mu_N) + 2$ , which contradicts (8.2b).

**Subcase 2.3:**  $\lambda_3 > 1, \lambda_4 < 1$  Condition (i) implies that  $\mu_N(d_N^4) = 1$ . But then, (8.2c) reduces to  $\lambda_4 = -1$ .

**Subcase 2.4:**  $\lambda_3 < 1, \lambda_4 \geq 1$  Condition (i) implies that  $\mu_N(d_N^3) = 1$ . But then,  $u_1(\mu_N) > \lambda_3 u_3(\mu_N) + 2$ , which contradicts (8.2b).

**Subcase 2.5:**  $\lambda_4 < \lambda_3 < 1$  Condition (i) implies that  $\mu_N(d_N^4) = 1$ . The same conclusion as in the case 2.3 is obtained.

**Subcase 2.6:**  $\lambda_3 < \lambda_4 < 1$  Condition (i) implies that  $\mu_N(d_N^3) = 1$ . The same conclusion as in the case 2.4 is obtained.

**Subcase 2.7:**  $0 < \lambda_3 = \lambda_4 < 1$  Condition (i) implies that  $\mu_N(d_N^3) = 1 - \mu_N(d_N^4) = \beta$  with  $\beta \in [0, 1]$ . Condition (8.2b) implies

$$\beta[1 + \lambda_3] + (1 - \beta)[1 - 3\lambda_3] = 0 \Rightarrow \beta = \frac{3\lambda_3 - 1}{4\lambda}$$

Similarly, (8.2c) implies

$$\beta[1 - 3\lambda_3] + (1 - \beta)[1 + \lambda_3] = 0 \Rightarrow \beta = \frac{1 + \lambda_3}{4\lambda}$$

Therefore,  $3\lambda_3 - 1 = 1 + \lambda_3$ , or equivalently,  $\lambda_3 = 1$ , which is a contradiction.

**Subcase 2.8:**  $\lambda_3 = \lambda_4 = 0$  As in the previous case, condition (i) implies that  $\mu_N(d_N^3) = 1 - \mu_N(d_N^4) = \beta$  with  $\beta \in [0, 1]$ . However, (8.2a) and (8.2b) imply that  $u_1(\mu_N) = u_2(\mu_N) = 2$ , which cannot be satisfied by any such mechanism.

**Case 3:**  $\lambda_1 = \lambda_2 = \lambda_3 = 0$  From (8.1) with  $S = N$ ,  $i = 2$  and  $j = 4$ , we get that  $\lambda_4 u_4(\mu_N) = 0$  (since  $u_4(\mu_{\{1,3,4\}}) = 0$ ). Hence,  $\sum_{i \in N} \lambda_i u_i(\mu_N) = 0$ . However, condition (i) implies that  $\sum_{i \in N} \lambda_i u_i(\mu_N) = 3$ , which is a contradiction.

**Case 4:**  $\lambda_1 = \lambda_2 = 0, \lambda_2 \neq \lambda_3$  Condition (8.1) applied to  $S = N$  with  $i = 2$  and  $j = 3, 4$  gives  $\lambda_3 u_3(\mu_N) = \lambda_4 u_4(\mu_N) = 0$  (since  $u_3(\mu_{\{1,3,4\}}) = u_4(\mu_{\{1,3,4\}}) = 0$ ). Hence,  $\sum_{i \in N} \lambda_i u_i(\mu_N) = 0$ . However, condition (i) implies that  $\sum_{i \in N} \lambda_i u_i(\mu_N) > 0$ , which is a contradiction.

**Case 5:**  $\lambda_1 \neq \lambda_2, \lambda_2 = \lambda_3 > 0$  Without loss of generality, we can set  $\lambda_2 = \lambda_3 = 1$ . It can be easily verified that, (8.1) implies:  $u_2(\mu_{\{1,2,3\}}) = u_3(\mu_{\{1,2,3\}}) = u_2(\mu_{\{2,3,4\}}) = u_3(\mu_{\{2,3,4\}}) = 1$  and  $u_1(\mu_{\{1,2,3\}}) = u_1(\mu_{\{1,2,4\}}) = u_2(\mu_{\{1,2,4\}}) = u_4(\mu_{\{1,2,4\}}) = u_4(\mu_{\{2,3,4\}}) = 0$ . Then, condition (8.1) applied to  $N$  reduces to

$$\lambda_1 u_1(\mu_N) = u_2(\mu_N) - 1 \quad (8.3a)$$

$$\lambda_1 u_1(\mu_N) = u_3(\mu_N) - 1 \quad (8.3b)$$

$$\lambda_1 u_1(\mu_N) = \lambda_4 u_4(\mu_N) \quad (8.3c)$$

On the other hand, we have that

$$\lambda_1 u_1(d_N) + u_2(d_N) + u_3(d_N) + \lambda_4 u_4(d_N) = \begin{cases} 2 + 3(\lambda_1 + \lambda_4), & \text{if } d_N = d_N^2, d_N^3 \\ 6 - \lambda_1 + 3\lambda_4, & \text{if } d_N = d_N^1 \\ 6 + 3\lambda_1 - \lambda_4, & \text{if } d_N = d_N^4 \end{cases} \quad (8.3d)$$

**Subcase 5.1:**  $\lambda_1 > 1, \lambda_4 > 1$  Condition (i) implies that  $\mu_N(d_N^2) = 1 - \mu_N(d_N^3) = \beta$  with  $\beta \in [0, 1]$ . But then, since  $u_2(\mu_N) = u_3(\mu_N)$  (by (8.3a) and (8.3b)), we must necessarily have that  $\beta = 1/2$ . Therefore,  $u_2(\mu_N) = u_3(\mu_N) = 1$  and  $u_1(\mu_N) = u_4(\mu_N) = 3$ . However, this together with (8.3a) imply that  $\lambda_1 = 0$ , which is a contradiction.

**Subcase 5.2:**  $\lambda_1 > 1, \lambda_4 = 1$  Condition (i) implies that  $\mu_N(d_N^1) = 0$ . Hence, (since  $\lambda_1 > 2/3$ )  $\lambda_1 u_1(\mu_N) > u_2(\mu_N) - 1$ , which contradicts (8.3a).

**Subcase 5.3:**  $\lambda_1 > 1, \lambda_4 < 1$  Condition (i) implies that  $\mu_N(d_N^4) = 1$ . With this, (8.3a) implies that  $\lambda_1 = 2/3 < 1$ , which is a contradiction.

**Subcase 5.4:**  $\lambda_1 < 1, \lambda_4 \geq 1$  Condition (i) implies that  $\mu_N(d_N^1) = 1$ . This together with (8.3a) imply that  $\lambda_1 = -2$ .

**Subcase 5.5:**  $\lambda_1 < \lambda_4 < 1$  Condition (i) implies that  $\mu_N(d_N^1) = 1$ . The same conclusion as in case 5.4 is obtained.

**Subcase 5.6:**  $\lambda_4 < \lambda_1 < 1$  Condition (i) implies that  $\mu_N(d_N^4) = 1$ . Hence, (8.3c) implies that  $3\lambda_1 = -\lambda_4$ . Therefore, (since  $\lambda_1, \lambda_2 \geq 0$ )  $\lambda_1 = \lambda_2 = 0$ , which is a contradiction.

**Subcase 5.7:**  $0 < \lambda_4 = \lambda_1 < 1$  Condition (i) implies that  $\mu_N(d_N^1) = 1 - \mu_N(d_N^4) = \beta$ , with  $\beta \in [0, 1]$ . On the other hand, (8.3c) implies that  $u_1(\mu_N) = u_4(\mu_N)$ . Hence, we must have that  $\beta = 1/2$ . But then, (8.3a) implies that  $\lambda_1 = 2$ , which is a contradiction.

**Subcase 5.8:**  $0 = \lambda_4 = \lambda_1$  As in the previous case, condition (i) implies that  $\mu_N(d_N^1) = 1 - \mu_N(d_N^4) = \beta$ , with  $\beta \in [0, 1]$ . Hence,  $u_2(\mu_N) = u_3(\mu_N) = 3$ . However, by (8.3a) and (8.3b),  $u_2(\mu_N) = u_3(\mu_N) = 1$ , which contradicts the previous fact.

**Case 6:**  $\lambda_1 \neq \lambda_2, \lambda_2 = \lambda_3 = 0$  Condition (8.1) applied to  $S = N$  with  $i = 2$  and  $j = 1, 4$  gives  $\lambda_1 u_1(\mu_N) = \lambda_4 u_4(\mu_N) = 0$  (since  $u_1(\mu_{\{1,3,4\}}) = u_4(\mu_{\{1,3,4\}}) = 0$ ). Hence,  $\sum_{i \in N} \lambda_i u_i(\mu_N) = 0$ . However, condition (i) implies that  $\sum_{i \in N} \lambda_i u_i(\mu_N) = 3(1 + \lambda_4) > 0$ , which is a contradiction.

**Case 7:**  $0 = \lambda_2 \neq \lambda_1, \lambda_2 \neq \lambda_3$  Condition (8.1) applied to  $S = N$  with  $i = 2$  and  $j = 1, 3, 4$  gives  $\lambda_1 u_1(\mu_N) = \lambda_3 u_3(\mu_N) = \lambda_4 u_4(\mu_N) = 0$  (since  $u_1(\mu_{\{1,3,4\}}) = u_3(\mu_{\{1,3,4\}}) = u_4(\mu_{\{1,3,4\}}) = 0$ ). Hence,  $\sum_{i \in N} \lambda_i u_i(\mu_N) = 0$ . However, condition (i) implies that  $\sum_{i \in N} \lambda_i u_i(\mu_N) = 3(\lambda_1 + \lambda_3 + \lambda_4) > 0$  (since we must have  $\lambda_1 > 0$  and  $\lambda_3 > 0$ ). But this is a contradiction.

**Case 8:**  $\lambda_2 > \lambda_1, \lambda_2 \neq \lambda_3$  Condition (8.1) applied to  $S = N$  with  $i = 1$  and  $j = 2$  gives  $\lambda_1 u_1(\mu_N) = \lambda_2 u_2(\mu_N)$  (since  $u_1(\mu_{\{1,3,4\}}) = u_2(\mu_{\{2,3,4\}}) = 0$ ). On the other hand, condition (i) implies that  $\mu_N(d_N^2) = 0$ . But then, this implies that  $\lambda_1 u_1(\mu_N) < \lambda_2 u_2(\mu_N)$ , which is a contradiction.

**Case 9:**  $\lambda_2 < \lambda_1, \lambda_2 \neq \lambda_3$  As in the previous case, we have that  $\lambda_1 u_1(\mu_N) = \lambda_2 u_2(\mu_N)$ . We also have that condition (i) implies  $\mu_N(d_N^1) = 0$ . But then, this implies that  $\lambda_1 u_1(\mu_N) > \lambda_2 u_2(\mu_N)$ , which is a contradiction.

We conclude that  $\Gamma_C$  has no H-solution. This completes the proof.



# On the Values for Bayesian Cooperative Games with Sidepayments<sup>1</sup>

## 1 Introduction

In this paper we explore the relationship between the following value like solution concepts for cooperative games with incomplete information<sup>1</sup>: Myerson's (1984b) M-solution, Salamanca's (2016) H-solution and A. Kalai and E. Kalai's (2013) cooperative-competitive solution. We consider a model in which utility transfers in the form of sidepayments are allowed. Transferable utility is a common assumption in Cooperative Game Theory. It states that utilities are quasi-linear in money and *unrestricted* monetary transfers can be performed (see Aumann (1960)). Our model, however, may exhibit *restricted* monetary transfers. The reason is that, in principle, private information is unverifiable. As a consequence, allowable state-contingent contracts are required to be incentive compatible and, thus, not all state-contingent plans of sidepayments might be feasible.

It is well known that utility transfers serve as a linear activity that can be used for signaling purposes, i.e., for helping to satisfy incentive compatibility. Yet, in our model, a transfers scheme will typically affect the interim utilities, which makes impossible to transfer interim utility across types without affecting the incentive constraints. In particular, one cannot generally construct, corresponding to a first best interim utility allocation, a transfer scheme satisfying incentive compatibility (see, e.g., example 1 in Myerson (2007))<sup>2</sup>.

We separate the analysis of our model in two sections depending on the number of players and the amount of externalities involved in the game situation. We start by considering general  $n$ -person Bayesian cooperative games in which the actions available to any particular coalition do not have an impact on the utilities of the players in the complementary coalition (lack of strategic externalities). Coalitions are then said to be *orthogonal*. We allow, however, that a

---

<sup>1</sup>This chapter corresponds to the paper: *On the Values for Bayesian Cooperative Games with Sidepayments*, HAL Working paper 01468867, 2017.

<sup>2</sup>Similar difficulties are also encountered by Forges, Mertens and Vohra (2002) to analyze the incentive compatible interim (coarse) core of an exchange economy with differential information and quasi-linear utilities.



player's utility can depend on other players' information (informational externalities). When considering the particular case in which information is complete, a cooperative analysis can be done by applying any standard value-like solution concept for non-transferable utility (NTU) games. In addition, incentive constraints disappear, thus utility becomes fully transferable. Therefore, as it is widely known, both Harsanyi's (1963) NTU value and Shapley's (1969) NTU value coincide. Moreover, their formula is given by Shapley's (1953) transferable utility (TU) value. We extend this result to games with incomplete information.

Myerson (1984a,b) developed an approach in which incentive constraints are used to define the virtual utility of players. Virtual utilities play the same role as that of weighted utility in the Harsanyi-Shapley method of fictitious transfers<sup>3</sup>. This approach has been used in Myerson (1984b) to extend the Shapley NTU value to an environment with incomplete information. It has also been recently used in Salamanca (2016) to generalize the Harsanyi NTU value. Our first main result (Theorem 1) establishes that these two cooperative solutions are interim utility equivalent in our model with sidepayments and orthogonal coalitions. Their common definition, however, cannot be described by a simple closed form expression as the Shapley TU value in games with complete information. The reason is that, due to the restricted nature of the utility transfers, the interim incentive efficient frontier is not generally an hyperplane as it would be in a game with complete information.

The subsequent analysis is devoted to cooperation in two-person Bayesian cooperative games in which the players face both informational and strategic externalities. In its more general form, Myerson's (1984b) solution allows for both kinds of externalities. Thus it can be applied to this particular class of games. Indeed, his cooperative solution extends Nash's (1953) bargaining solution with variable threats. More recently, A. Kalai and E. Kalai (2013) proposed a value for two-person Bayesian games (in strategic form) with transferable utility. Their semi-cooperative solution, called the *cooperative-competitive (or "coco") value*, is based on a decomposition of the game into cooperative and competitive component games. The coco value conceptually differs from Myerson's solution in that the former ignores potential incentive compatibility issues. Also, it is defined only at the ex-ante stage. Kalai and Kalai (2013) however conjectured the existence of a close relationship between these two solution concepts in the case private information is verifiable ex-post<sup>4</sup>. Our second main result (Theorem 2) provides a positive answer to this open problem. We show that under de Clippel and Minelli's (2004) verifiable types assumption, Myerson's solution and the coco value are ex-ante utility equivalent. Myerson's solution is thus an appropriate generalization of the coco value to games in which information is not verifiable ex-post. Our result also exhibits why extending the coco value to general environments in which no information becomes observable ex-post cannot be done without referring to an appropriate interim framework as in Myerson's (1984a,b) approach.

The paper is organized as follows. Section 2 is devoted to specifying formally the model of a Bayesian cooperative game with sidepayments. In Section 3 we introduce the concept of incentive efficiency and its relation to the virtual utility approach. Sections 4 and 5 contain the main body of results. Section 4 analyzes  $n$ -player games with orthogonal coalitions. Finally, Section 5 focus on the two-player case.

<sup>3</sup>See Myerson (1992) for a detailed explanation of the fictitious transfers method.

<sup>4</sup>This open problem was also pointed out by Forges and Serrano (2013).

## 2 Bayesian Cooperative Game

A *cooperative game with incomplete information* is a tuple  $\Gamma = \{N, (D_S)_{S \subseteq N}, (T_i, u_i)_{i \in N}, p\}$ , where  $N = \{1, 2, \dots, n\}$  denotes the set of players and for any (nonempty) coalition<sup>5</sup>  $S \subseteq N$ ,  $D_S$  is the set of feasible decisions for  $S$ ; for any player  $i \in N$ ,  $T_i$  denotes the (finite) set of possible types of player  $i$ ,  $p$  is the prior probability distribution over  $T_N = \prod_{i \in N} T_i$ , and  $u_i : D_N \times T_N \rightarrow \mathbb{R}$  is the utility function of player  $i$ . The sets of feasible decisions are finite and *superadditive*, namely, for any two disjoint coalitions  $S$  and  $R$ ,  $D_R \times D_S \subseteq D_{R \cup S}$ . This definition allows for both informational and strategic externalities, since the payoffs of the members of a coalition  $S$  may depend on the types and decisions of the players in  $N \setminus S$ .

We assume types are stochastically independent<sup>6</sup> and  $p(t_i) > 0$  for all  $t_i \in T_i$  and all  $i \in N$ . We use the notations  $t_{-i} = (t_j)_{j \in N \setminus i} \in T_{-i} = \prod_{j \in N \setminus i} T_j$  and  $t_S = (t_i)_{i \in S} \in T_S = \prod_{i \in S} T_i$ . For simplicity, we drop the subscript  $N$  in the case of the grand coalition, so we define  $D = D_N$  and  $T = T_N$ .

A *mechanism* for coalition  $S \subseteq N$  is a pair of functions  $(\mu_S, x_S)$  defined by<sup>7</sup>

$$\begin{aligned} \mu_S : T_S &\rightarrow \Delta(D_S) & x_S : T_S &\rightarrow \mathbb{R}^S \\ t_S &\mapsto \mu_S(\cdot | t_S) & t_S &\mapsto (x_S^i(t_S))_{i \in S} \end{aligned}$$

The component  $\mu_S$  is a type-contingent lottery on the set of feasible decisions for  $S$ , while  $x_S$  is a vector of type-contingent monetary transfers. Monetary transfers must satisfy the following budget feasibility condition<sup>8</sup>:

$$\sum_{i \in S} x_S^i(t_S) \leq 0, \quad \forall t_S \in T_S \quad (2.1)$$

In the case  $S \neq N$ , the mechanism  $(\mu_S, x_S)$  stands as a *threat* to be carried out only if  $N \setminus S$  refuses to cooperate with  $S$ . We let the set of *budget-feasible* mechanisms satisfying (2.1) be denoted  $\mathcal{F}_S$ .

The (interim) expected utility of player  $i \in N$  of type  $t_i$  under the mechanism  $(\mu_N, x_N)$  when he pretends to be of type  $\tau_i$  (while all other players are truthful) is

$$U_i(\mu_N, x_N, \tau_i | t_i) = \sum_{t_{-i} \in T_{-i}} p(t_{-i}) \left[ x_N^i(\tau_i, t_{-i}) + \sum_{d \in D} \mu_N(d | \tau_i, t_{-i}) u_i(d, (t_i, t_{-i})) \right]$$

Monetary transfers are added linearly to the expected utilities. As is standard, we denote  $U_i(\mu_N, x_N | t_i) = U_i(\mu_N, x_N, t_i | t_i)$ .

<sup>5</sup>For any two sets  $A$  and  $B$ ,  $A \subseteq B$  denotes *weak* inclusion (i.e., possibly  $A = B$ ), and  $A \subset B$  denotes strict inclusion.

<sup>6</sup>This assumption is made without loss of generality, since the solution concepts studied in this paper satisfy the *probability-invariance axiom* described by Myerson (1984a), and so for any game with dependent types, prior probabilities and utilities can be jointly modified in a way that the new game has independent types and both games impute probability and utility functions that are decision-theoretically equivalent.

<sup>7</sup>For any finite set  $A$ ,  $\Delta(A)$  denotes the set of probability distributions over  $A$ .

<sup>8</sup>Other forms of budget feasibility can be defined. For instance, Prescott and Townsend (1984) and Myerson (2007) consider *average* budget feasibility, i.e.,  $\sum_{t_S \in T_S} p(t_S) \sum_{i \in S} x_S^i(t_S) \leq 0$ . In Myerson (2007), the amount  $\sum_{t_S \in T_S} p(t_S) \sum_{i \in S} x_S^i(t_S)$  is interpreted as the expected payoff to the mediator.

A mechanism  $(\mu_N, x_N)$  is *incentive compatible* if and only if

$$U_i(\mu_N, x_N | t_i) \geq U_i(\mu_N, x_N, \tau_i | t_i), \quad \forall i \in N, \quad \forall t_i, \tau_i \in T_i$$

By the revelation principle, any Nash equilibrium of any noncooperative game that the players could design in order to exchange information and make decisions can be equivalently represented by an incentive compatible mechanism. Hence, there is no loss of generality in restricting our attention to such incentive compatible mechanisms. Also, we notice that the incentive constraints only depend on the conditional expected monetary transfers. Hence we can restrict ourselves without loss of generality to deterministic money transfers. A budget-feasible and incentive-compatible mechanism for the grand coalition is said to be *feasible* for  $N$ . We denote  $\mathcal{F}_N^*$  the set of feasible mechanisms for  $N$ .

### 3 Incentive Efficiency and Virtual Utility

A mechanism is (*interim*) *incentive efficient* for the grand coalition if and only if it is feasible and no other feasible mechanism yields strictly higher expected utilities to all types of all players. Incentive efficient mechanisms can be characterized using the concept of virtual utility.

Given vectors  $\lambda \in \prod_{i \in N} \mathbb{R}_+^{T_i}$  and  $\alpha \in \prod_{i \in N} \mathbb{R}_+^{T_i \times T_i}$ , the *virtual utility* of a decision  $d \in D$  in state  $t \in T$  for player  $i \in N$  is defined as follows:

$$v_i(d, t, \lambda, \alpha) = \frac{1}{p(t_i)} \left[ \left( \lambda_i(t_i) + \sum_{\tau_i \in T_i} \alpha_i(\tau_i | t_i) \right) u_i(d, t) - \sum_{\tau_i \in T_i} \alpha_i(t_i | \tau_i) u_i(d, (\tau_i, t_{-i})) \right]$$

The vectors  $\lambda$  and  $\alpha$  are called the virtual utility scales. The virtual utility of player  $i$  is a distorted utility scale that exaggerates the difference between his actual utility and the utilities of  $i$ 's other types that would be tempted to imitate him. The following characterization follows from duality theory of linear programming (a detailed reasoning is given in Myerson (2007)).

**Proposition 1 (Incentive efficiency).**

A feasible mechanism  $(\mu_N, x_N) \in \mathcal{F}_N^*$  is incentive efficient if and only if there exist  $\lambda \in \prod_{i \in N} \mathbb{R}_+^{T_i}$  and  $\alpha \in \prod_{i \in N} \mathbb{R}_+^{T_i \times T_i}$  such that

$$\alpha_i(\tau_i | t_i) [U_i(\mu_N, x_N | t_i) - U_i(\mu_N, x_N, \tau_i | t_i)] = 0, \quad \forall i \in N, \forall t_i \in T_i, \forall \tau_i \in T_i \quad (3.1)$$

$$\lambda_i(t_i) + \sum_{\tau_i \in T_i} \alpha_i(\tau_i | t_i) - \sum_{\tau_i \in T_i} \alpha_i(t_i | \tau_i) = p(t_i), \quad \forall i \in N, \forall t_i \in T_i \quad (3.2)$$

$$\sum_{d \in D} \mu_N(d | t) \sum_{i \in N} v_i(d, t, \lambda, \alpha) = \max_{d \in D} \sum_{i \in N} v_i(d, t, \lambda, \alpha), \quad \forall t \in T \quad (3.3)$$

$$\sum_{i \in N} x_N^i(t) = 0, \quad \forall t \in T \quad (\text{exact budget feasibility}) \quad (3.4)$$

Equation (3.1) is the usual *complementary slackness* condition. Equations in (3.2) are called by Myerson (2007) *hydraulic equations*: consider a network in which at each node a type  $t_i$  is located. If we interpret  $p(t_i)$  as the flow into the network at  $t_i$ ,  $\lambda_i(t_i)$  as the flow out of the network at  $t_i$ , and  $\alpha_i(\tau_i | t_i)$  as the flow from  $\tau_i$  to  $t_i$ , then (3.4) says that these flows are balanced at each node. Finally, (3.3)-(3.4) implies that any incentive efficient mechanism determines an allocation that is ex-post efficient in terms of the virtual utility scales. Hence, one could say that incentive compatibility compels the players to behave according to their virtual utilities. Myerson (1984b) refers to this idea as the *virtual utility hypothesis* (see Myerson (1991, ch. 10) for a detailed discussion).

## 4 Values for Bayesian Cooperative Games with Orthogonal Coalitions

As in most of the literature in cooperative game theory, in this section we shall assume that coalitions are *orthogonal*, namely, when coalition  $S \subseteq N$  chooses an action which is feasible for it, the payoffs to the members of  $S$  do not depend on the actions of the complementary coalition  $N \setminus S$ . Formally,

$$u_i((d_S, d_{N \setminus S}), t) = u_i((d_S, d'_{N \setminus S}), t)$$

for every  $S$ ,  $i \in S$ ,  $d_S \in D_S$ ,  $d_{N \setminus S}, d'_{N \setminus S} \in D_{N \setminus S}$  and  $t \in T$ . Then we can let  $u_i(d_S, t)$  denote the utility for player  $i \in S$  if  $d_S \in D_S$  is carried out. That is,  $u_i(d_S, t) := u_i((d_S, d_{N \setminus S}), t)$  for some  $d_{N \setminus S} \in D_{N \setminus S}$  (recall that  $D_S \times D_{N \setminus S} \subseteq D$ ). This assumption excludes strategic externalities. However, the payoffs of the members of a coalition  $S$  might still depend on the types of the players in  $N \setminus S$ . Hence, informational externalities are allowed.

When information is complete, the orthogonal coalitions hypothesis allows to describe a NTU game as a collection of feasible utility sets. This *characteristic function form* suppresses any explicit mention of the decisions generating the utilities. Although implicitly, we assume that an utility allocation  $u_S$  is feasible for  $S$  if the players in  $S$  together have a joint strategy<sup>9</sup> that enables them to allocate  $u_S$ . If utilities are linear in money and players can make unrestricted sidepayments of money, we obtain a TU game<sup>10</sup>. It is well known that both the Shapley NTU value and the Harsanyi NTU value of a TU game coincide and its formula is given by the Shapley TU value<sup>11</sup>. Although, in our model, utility may not be fully transferable due to the presence of incentive constraints, we provide an analogous result in the class of Bayesian cooperative games.

<sup>9</sup>This may include a correlated strategy or a joint decision, discarding utility or even transferring utility.

<sup>10</sup>Indeed, let  $V = (V(S))_{S \subseteq N}$  be a NTU game. For each  $S \subseteq N$ , let  $P(S) := \{u \in \mathbb{R}^S \mid \sum_{i \in S} u_i \leq 0\}$  denote the set of (unrestricted) sidepayments for the members of  $S$ . Then, the game  $W = (V(S) + P(S))_{S \subseteq N}$  is a TU game for which the worth of coalition  $S$  is  $w(S) = \max_{v \in V(S)} \sum_{i \in S} v_i$  and  $W(S) = \{w \in \mathbb{R}^S \mid \sum_{i \in S} w_i \leq w(S)\}$ .

<sup>11</sup>The fact that the Shapley NTU value coincides with the Shapley TU value for TU games can be consulted in Myerson (1991, pp. 470). On the other hand, Proposition 4.10 in Hart (1985b) establishes the equivalence between the Harsanyi NTU value and the Shapley TU value in TU games.

## 4.1 The M-Solution

We consider the fictitious game in which the players make interpersonal utility comparisons in terms of some fixed virtual scales  $(\lambda, \alpha)$ . For any coalition  $S$ , we let  $W_S(\mu_S, x_S, t, \lambda, \alpha)$  be the sum of virtual utilities that the members of  $S$  would expect in state  $t$  when they select the mechanism  $(\mu_S, x_S)$  as a threat, that is<sup>12</sup>

$$W_S(\mu_S, x_S, t, \lambda, \alpha) = \sum_{i \in S} \left[ v_i(\mu_S, t, \lambda, \alpha) + x_S^i(t_S) \right] \quad (4.1)$$

where  $v_i(\mu_S, t, \lambda, \alpha)$  is the linear extension of  $v_i(\cdot, t, \lambda, \alpha)$  over  $\mu_S$ . For any vector of threats  $\eta = (\mu_S, x_S)_{S \subseteq N}$ , we define  $W(\eta, t, \lambda, \alpha) = (W_S(\mu_S, x_S, t, \lambda, \alpha))_{S \subseteq N}$  as the characteristic function game when virtual utility is fully transferable in state  $t$ . Let  $\phi$  be the Shapley TU value operator; for  $i \in N$ ,  $\phi_i(N, W(\eta, t, \lambda, \alpha))$  will thus denote the Shapley TU value of player  $i$  in the virtual game when the vector of threats  $\eta$  is selected.

The interim allocation  $\omega \in \prod_{i \in N} \mathbb{R}^{T_i}$  is warranted by  $\lambda, \alpha$  and  $\eta$  if and only if

$$\left( \lambda_i(t_i) + \sum_{\tau_i \in T_i} \alpha_i(\tau_i | t_i) \right) \omega_i(t_i) - \sum_{\tau_i \in T_i} \alpha_i(t_i | \tau_i) \omega_i(\tau_i) = \sum_{t_{-i} \in T_{-i}} p(t) \phi_i(N, W(\eta, t, \lambda, \alpha)), \quad \forall i \in N, \quad \forall t_i \in T_i \quad (4.2)$$

In other words,  $\omega$  corresponds to the real utility allocation which would give every type of each player his expected Shapley TU value in the virtual game.

We say that  $\eta = (\mu_S, x_S)_{S \subseteq N}$  is a vector of *rational threats* (w.r.t.  $\lambda$  and  $\alpha$ ) if, for each  $S \subseteq N$ , the mechanism  $(\mu_S, x_S)$  is an optimal solution to

$$\max_{(v_S, y_S) \in \mathcal{F}_S} \sum_{t \in T} p(t) W_S(v_S, y_S, t, \lambda, \alpha) \quad (4.3)$$

A mechanism  $(\mu_S, x_S)$  is an optimal solution to (4.3) if and only if, for every  $t_S \in T_S$ ,

$$\sum_{i \in S} v_i(\mu_S, t_S, \lambda, \alpha) = \max_{d_S \in D_S} \sum_{i \in S} v_i(d_S, t_S, \lambda, \alpha) \quad \text{and} \quad \sum_{i \in S} x_S^i(t_S) = 0,$$

where

$$v_i(d_S, t_S, \lambda, \alpha) := \sum_{t_{N \setminus S} \in T_{N \setminus S}} p(t_{N \setminus S}) v_i(d_S, t, \lambda, \alpha)$$

and  $v_i(\mu_S, t_S, \lambda, \alpha)$  is the linear extension of  $v_i(\cdot, t_S, \lambda, \alpha)$  over  $\mu_S$  (recall that  $\mu_S$  is measurable w.r.t.  $T_S$ ).

<sup>12</sup>In the virtual game, sidepayments are meant to be done in terms of the virtual scales  $(\lambda, \alpha)$ . Hence an appropriate definition for  $W_S$  should be  $W_S(\mu_S, x_S, t, \lambda, \alpha) = \sum_{i \in S} v_i(\mu_S, t, \lambda, \alpha) + \beta_i(t_i, \lambda, \alpha) x_S^i(t_S)$ , where  $\beta_i(t_i, \lambda, \alpha) := [\lambda_i(t_i) + \sum_{\tau_i \in T_i} \alpha_i(\tau_i | t_i) - \sum_{\tau_i \in T_i} \alpha_i(t_i | \tau_i)] / p(t_i)$ . However, the scales  $(\lambda, \alpha)$  are selected endogenously in a way that the mechanism  $(\mu_N, x_N)$  satisfies (3.2). Then, we can set  $\beta_i(t_i, \lambda, \alpha) = 1$  for all  $i \in N$  and all  $t_i \in T_i$ .

**Definition 1 (M-solution (Myerson, 1984b)).**

A feasible mechanism for the grand coalition  $(\bar{\mu}_N, \bar{x}_N) \in \mathcal{F}_N^*$  is a *M-solution*, if there exist vectors  $\lambda > 0$ ,  $\alpha \geq 0$ , and  $\eta = (\mu_S, x_S)_{S \subseteq N}$  with  $(\mu_N, x_N) = (\bar{\mu}_N, \bar{x}_N)$  such that:

- (i)  $(\mu_N, x_N)$  satisfies (3.1)-(3.4) for  $\lambda$  and  $\alpha$ .
- (ii)  $\eta$  is a vector of rational threats w.r.t.  $\lambda$  and  $\alpha$ .
- (iii)  $U(\mu_N, x_N) = (U_i(\mu_N, x_N | t_i))_{t_i \in T_i, i \in N}$  is warranted by  $\lambda$ ,  $\alpha$  and  $\eta$ .

The vector  $U(\mu_N, x_N)$  of warranted claims is called a *M-value*. We denote by  $\mathcal{V}_M(\Gamma)$  the set of M-values of  $\Gamma$ .<sup>13</sup>

## 4.2 The H-Solution

A first component in the definition of the H-solution is the construction of a vector of threats called an *egalitarian solution*. An egalitarian solution requires threats to satisfy a condition of average balanced contributions (principle of equal gains), and differs from the H-solution in that the latter endogenously determines the virtual scales  $(\lambda, \alpha)$  by additionally imposing a utilitarian criterion.

Given a vector of virtual scales  $(\lambda, \alpha)$ , a vector of threats  $\eta = (\mu_S, x_S)_{S \subseteq N}$  is an *egalitarian solution* (w.r.t.  $\lambda$  and  $\alpha$ ) if and only if, for all  $S \subseteq N$ , the mechanism  $(\mu_S, x_S)$  is an optimal solution to

$$\begin{aligned}
 & \max_{(v_S, y_S) \in \mathcal{F}_S} \sum_{t \in T} p(t) W_S(v_S, y_S, t, \lambda, \alpha) \\
 & \text{s.t. } \sum_{t_{-i} \in T_{-i}} p(t_{-i}) \sum_{j \in S \setminus i} \left[ v_i(v_S, t, \lambda, \alpha) + y_S^i(t_S) - v_i(\mu_{S \setminus j}, t, \lambda, \alpha) - x_{S \setminus j}^i(t_{S \setminus j}) \right] \\
 & \quad = \sum_{t_{-i} \in T_{-i}} p(t_{-i}) \sum_{j \in S \setminus i} \left[ v_j(v_S, t, \lambda, \alpha) + y_S^j(t_S) - v_j(\mu_{S \setminus i}, t, \lambda, \alpha) - x_{S \setminus i}^j(t_{S \setminus i}) \right],
 \end{aligned} \tag{4.4}$$

$$\forall t_i \in T_i, \forall i \in S$$

We notice that an egalitarian solution must be constructed recursively: given the vector of threats  $(\mu_{S \setminus j}, x_{S \setminus j})_{j \in S}$ ,  $(\mu_S, x_S)$  is determined solving (4.4). The possibility to make unrestricted sidepayments in terms of the virtual utility scales guarantees that this construction is always possible<sup>14</sup>.

In the problem (4.4), the objective function is the same as in (4.3). In an egalitarian solution, however, optimal threats are required to be “equitable”. Here, equitable means that the expected average virtual contribution of the different players in  $S$  to player  $i$  of type  $t_i$  (in coalition

<sup>13</sup>Definition 1 involves strictly positive utility weights  $\lambda$ . This complicates matters for obtaining existence results of the M-solution. Myerson (1984b) solves this dilemma by slightly enlarging the solution set to include utility allocations that are reasonable as emerging from limit points.

<sup>14</sup>In the absence of sidepayments, the optimization problem in (4.4) may not be feasible. The difficulty is due to a lack of comprehensiveness in the set of attainable virtual utility allocations (see Section 7 in Salamanca (2016)).

$S$ ) equals the expected average virtual contribution of player  $i$  to the different players in  $S$  as assessed by type  $t_i$  (see Section 4 in Salamanca (2016) for a justification of this equity criterion).

**Definition 2 (H-Solution (Salamanca, 2016)).**

A feasible mechanism for the grand coalition  $(\bar{\mu}_N, \bar{x}_N) \in \mathcal{F}_N^*$  is an *H-solution*, if there exist vectors  $\lambda > 0$ ,  $\alpha \geq 0$ , and  $\eta = (\mu_S, x_S)_{S \subseteq N}$  with  $(\mu_N, x_N) = (\bar{\mu}_N, \bar{x}_N)$  such that:

- (i)  $(\mu_N, x_N)$  satisfies (3.1)-(3.4) for  $\lambda$  and  $\alpha$ .
- (ii)  $\eta$  is an egalitarian solution w.r.t.  $\lambda$  and  $\alpha$ .

The vector  $U(\mu_N, x_N)$  of interim utilities is called an *H-value*. We denote by  $\mathcal{V}_H(\Gamma)$  the set of H-values of  $\Gamma$ .

### 4.3 Reconciling the Differences

When comparing the previous solutions, both satisfy the utilitarian conditions (3.1)-(3.4). In addition, for any egalitarian solution  $\eta = (\mu_S, x_S)_{S \subseteq N}$ ,  $U(\mu_N, x_N)$  is warranted by  $\lambda$ ,  $\alpha$  and  $\eta$ .<sup>15</sup> Hence, the M-solution and the H-solution differ only in the way both determine the threats for intermediate coalitions. We will show, however, that this difference disappears in the present model with sidepayments. Formally, we are going to prove the following:

**Theorem 1.**

Let  $\Gamma$  be a Bayesian cooperative game with orthogonal coalitions and sidepayments. Then,  $\mathcal{V}_M(\Gamma) = \mathcal{V}_H(\Gamma)$ .

In order to establish this result, we shall construct a particular class of threat mechanisms which will help us to establish a certain connection between the rational threats and the egalitarian solution. This relationship is stated in Lemma 1. The final conclusion of Theorem 1 follows from the double inclusion established in Propositions 2 and 3.

Fix the virtual scales  $(\lambda, \alpha)$  and let  $S \subseteq N$  be a coalition. Given a vector of threats  $(\mu_S, x_S)_{R \subseteq S}$ , we define

$$r_S^i(t_S, \lambda, \alpha) := \sum_{\substack{R \subseteq S \\ i \in R}} (-1)^{|S \setminus R|+1} [v_i(\mu_R, t_S, \lambda, \alpha) + x_R^i(t_R)], \quad \forall t_S \in T_S. \quad (4.5)$$

The quantity  $r_S^i(t_S, \lambda, \alpha)$  can be thought as the cumulated “virtual dividends” that player  $i \in S$  expects in state  $t_S$  from his participation in all coalitions  $R \subset S$  to which he belongs.

Given the vector  $r_S(\lambda, \alpha) = (r_S^i(t_S, \lambda, \alpha))_{i \in S, t_S \in T_S}$ , consider a threat mechanism  $(\bar{\mu}_S, \bar{x}_S)$  for coalition  $S$  defined by<sup>16</sup>

$$\sum_{i \in S} v_i(\bar{\mu}_S, t_S, \lambda, \alpha) = \sum_{i \in S} v_i^*(t_S, \lambda, \alpha), \quad \forall t_S \in T_S, \quad (4.6a)$$

<sup>15</sup>See Remark 3 in Salamanca (2016).

<sup>16</sup>It is worth noting that  $(\bar{\mu}_S, \bar{x}_S)$  is not *uniquely* determined by (4.6a)-(4.6b). Indeed, there may be several random joint decisions  $\mu_S$  satisfying (4.6a). Yet, once  $\mu_S$  is determined, there exists a unique  $\bar{x}_S$  satisfying (4.6b) and (4.7a)-(4.7b).

$$\bar{x}_S^i(t_S) = v_i^*(t_S, \lambda, \alpha) - v_i(\bar{\mu}_S, t_S, \lambda, \alpha), \quad \forall i \in S, \forall t_S \in T_S \quad (4.6b)$$

where  $v^*(\lambda, \alpha) = (v_i^*(t_S, \lambda, \alpha))_{i \in S, t_S \in T_S}$  is the solution to

$$v_i^*(t_S, \lambda, \alpha) - r_S^i(t_S, \lambda, \alpha) = v_j^*(t_S, \lambda, \alpha) - r_S^j(t_S, \lambda, \alpha), \quad \forall i, j \in S, \forall t_S \in T_S. \quad (4.7a)$$

$$\sum_{i \in S} v_i^*(t_S, \lambda, \alpha) = \max_{d_S \in D_S} \sum_{i \in S} v_i(d_S, t_S, \lambda, \alpha), \quad \forall t_S \in T_S. \quad (4.7b)$$

The system of linear equations in (4.7a)-(4.7b) is always solvable, and its solution is unique<sup>17</sup>, hence  $v_i^*(\lambda, \alpha)$  is well defined. Notice also that for each  $t_S \in T_S$ , the transfers  $\bar{x}_S$  are exactly balanced

$$\sum_{i \in S} \bar{x}_S^i(t_S) = \sum_{i \in S} v_i^*(t_S, \lambda, \alpha) - \sum_{i \in S} v_i(\bar{\mu}_S, t_S, \lambda, \alpha) = 0$$

**Lemma 1.**

Let  $(\lambda, \alpha)$  be fixed. For a given coalition  $S \subseteq N$ , let  $(\bar{\mu}_S, \bar{x}_S)$  be defined by (4.6a)-(4.6b) with  $(\mu_R, x_R)_{R \subset S}$ . Then,  $(\bar{\mu}_S, \bar{x}_S)$  is an optimal solution to (4.3). If, in addition, for every  $R \subset S$ ,  $(\mu_R, x_R)$  is feasible in (4.4) w.r.t  $(\mu_{R \setminus j}, x_{R \setminus j})_{j \in S}$ . Then,  $(\bar{\mu}_S, \bar{x}_S)$  is also an optimal solution to (4.4) w.r.t.  $(\mu_{S \setminus j}, x_{S \setminus j})_{j \in S}$ .

**Proof.** The fact that  $(\bar{\mu}_S, \bar{x}_S)$  is an optimal solution to (4.3) is straightforward. Let  $(\mu_S, x_S)$  be an optimal solutions of (4.4) w.r.t.  $(\mu_{S \setminus j}, x_{S \setminus j})_{j \in S}$ . We notice that

$$\begin{aligned} \sum_{t_S \in T_S} p(t_S) \left[ \sum_{i \in S} v_i(\bar{\mu}_S, t_S, \lambda, \alpha) + \bar{x}_S^i(t_S) \right] &= \sum_{t_S \in T_S} p(t_S) \max_{d_S \in D_S} \sum_{i \in S} v_i(d_S, t_S, \lambda, \alpha) \\ &\geq \sum_{t_S \in T_S} p(t_S) \left[ \sum_{i \in S} v_i(\mu_S, t_S, \lambda, \alpha) + x_S^i(t_S) \right] \end{aligned}$$

Then, it suffices to show that  $(\bar{\mu}_S, \bar{x}_S)$  is feasible in (4.4) (w.r.t.  $(\mu_{S \setminus j}, x_{S \setminus j})_{j \in S}$ ). By construction, for any  $t_S \in T_S$ ,  $(\bar{\mu}_S, \bar{x}_S)$  satisfies

$$v_i(\bar{\mu}_S, t_S, \lambda, \alpha) + \bar{x}_S^i(t_S) - r_S^i(t_S, \lambda, \alpha) = v_j(\bar{\mu}_S, t_S, \lambda, \alpha) + \bar{x}_S^j(t_S) - r_S^j(t_S, \lambda, \alpha), \quad \forall i, j \in S. \quad (4.8)$$

Let  $t_i \in T_i$  be a fixed type of a player  $i \in S$ . Multiplying both sides of (4.8) by  $p(t_{S \setminus i})$ , summing over all

<sup>17</sup>Consider the homogeneous system. For each  $t_S \in T_S$ , the system has exactly  $|S|$  linearly independent equations, namely  $v_1(t_S) - v_j(t_S) = 0$  for each  $j \in S \setminus 1$  and  $\sum_{i \in S} v_i(t_S) = 0$ . The unique solution is  $v_i(t_S) = 0$  for all  $i \in S$ . Hence, the non-homogenous system is solvable and its solutions is unique.



$t_{S \setminus i} \in T_{S \setminus i}$  and all  $j \in S \setminus i$ , and rearranging terms yields

$$\begin{aligned}
& \sum_{t_{-i} \in T_{-i}} p(t_{-i}) \sum_{j \in S \setminus i} [v_i(\bar{\mu}_S, t, \lambda, \alpha) + \bar{x}_S^i(t_S) - v_i(\mu_{S \setminus j}, t, \lambda, \alpha) - x_{S \setminus j}^i(t_{S \setminus j})] \\
& - \sum_{t_{-i} \in T_{-i}} p(t_{-i}) \sum_{j \in S \setminus i} [v_j(\bar{\mu}_S, t, \lambda, \alpha) + \bar{x}_S^j(t_S) - v_j(\mu_{S \setminus i}, t, \lambda, \alpha) - x_{S \setminus i}^j(t_{S \setminus i})] \\
& = \sum_{\substack{R \subset S \\ i \in R \\ |R| \geq 2}} (-1)^{|S \setminus R|} \left[ \sum_{t_{-i} \in T_{-i}} p(t_{-i}) \sum_{j \in R \setminus i} [v_j(\mu_R, t, \lambda, \alpha) + x_R^j(t_R) - v_j(\mu_{R \setminus i}, t, \lambda, \alpha) - x_{R \setminus i}^j(t_{R \setminus i})] \right. \\
& \quad \left. - \sum_{t_{-i} \in T_{-i}} p(t_{-i}) \sum_{j \in R \setminus i} [v_i(\mu_R, t, \lambda, \alpha) + x_R^i(t_R) - v_i(\mu_{R \setminus j}, t, \lambda, \alpha) - x_{R \setminus j}^i(t_{R \setminus j})] \right]. \tag{4.9}
\end{aligned}$$

If, for every  $R \subset S$ ,  $(\mu_R, x_R)$  is feasible in (4.4) w.r.t  $(\mu_{R \setminus j}, x_{R \setminus j})_{j \in S}$ , then the right hand side (and therefore also the left hand side) of (4.9) is zero. This concludes the proof.  $\square$

The following inclusion readily follows from Lemma 1.

**Proposition 2.**

$$\mathcal{V}_M(\Gamma) \supseteq \mathcal{V}_H(\Gamma).$$

**Proof.** Let  $(\mu_N, x_N)$  be an H-solution supported by  $\eta = (\mu_S, x_S)_{S \subseteq N}$ ,  $\lambda$  and  $\alpha$ . For each  $S \subseteq N$ , Lemma 1 implies that  $(\mu_S, x_S)$  attains the optimal value of (4.3). Hence,  $\eta$  is a vector of rational threats<sup>18</sup>.  $\square$

We shall now prove the reverse inclusion. The basic idea will be to show that, given a M-solution with corresponding rational threats  $\eta = (\mu_S, x_S)_{S \subseteq N}$ , one can construct a transfers scheme  $(\tilde{x}_S)_{S \subseteq N}$  as in (4.6a)-(4.6b), such that the vector of threats  $(\mu_S, \tilde{x}_S)_{S \subseteq N}$  is an egalitarian solution. This is possible thanks to Lemma 1.

**Proposition 3.**

$$\mathcal{V}_M(\Gamma) \subseteq \mathcal{V}_H(\Gamma).$$

**Proof.** Let  $(\mu_N, x_N)$  be a M-solution supported by  $\eta = (\mu_S, x_S)_{S \subseteq N}$ ,  $\lambda$  and  $\alpha$ . Recursively define for each  $S \subset N$

$$\tilde{x}_S^i(t_S) := \tilde{v}_i^*(t_S, \lambda, \alpha) - v_i(\mu_S, t_S, \lambda, \alpha), \quad \forall i \in S, \forall t_S \in T_S$$

where  $\tilde{v}^*(\lambda, \alpha) = (\tilde{v}_i^*(t_S, \lambda, \alpha))_{i \in S, t_S \in T_S}$  is the solution to (4.7a)-(4.7b) with  $\tilde{r}_S(\lambda, \alpha)$  computed using  $(\mu_R, \tilde{x}_R)_{R \subset S}$  (already defined in the recursion). Define  $\tilde{x}_N = x_N$  and  $\tilde{\eta} = (\mu_S, \tilde{x}_S)_{S \subseteq N}$ . Obviously,  $(\mu_N, \tilde{x}_N)$  is a feasible mechanism satisfying (3.1)-(3.4) w.r.t.  $\lambda$  and  $\alpha$  and  $U(\mu_N, \tilde{x}_N) = U(\mu_N, x_N)$ . Hence, we only need to show that  $\tilde{\eta}$  is an egalitarian solution w.r.t.  $\lambda$  and  $\alpha$ . Notice that, for each  $S \subset N$ ,  $(\mu_S, \tilde{x}_S)$  satisfies (4.6a)-(4.6b) and (4.7a)-(4.7b) with  $(\mu_R, \tilde{x}_R)_{R \subset S}$ . Then, Lemma 1 (applied inductively) implies that, for each  $S \subset N$ ,  $(\mu_S, \tilde{x}_S)$  is an optimal solutions of (4.4) w.r.t.  $(\mu_{S \setminus j}, \tilde{x}_{S \setminus j})_{j \in S}$ . On the other hand, since  $\tilde{x}_S$  is exactly balanced for all subcoalitions,  $W_S(\mu_S, x_S, t, \lambda, \alpha) = W_S(\mu_S, \tilde{x}_S, t, \lambda, \alpha)$  for every  $S \subseteq N$ .

<sup>18</sup>Indeed, we must have that  $\sum_{i \in S} x_S^i(t_S) = 0$  for every  $t_S \in T_S$ , since  $\sum_{t_S \in T_S} p(t_S) \sum_{i \in S} x_S^i(t_S) = 0$  and  $\sum_{i \in S} x_S^i(t_S) \leq 0$  for all  $t_S \in T_S$ .

Then,  $U(\mu_N, \tilde{x}_N)$  is warranted by  $\lambda$ ,  $\alpha$  and  $\tilde{\eta}$ .<sup>19</sup> Therefore, by Lemma 1 and Proposition 2 in Salamanca (2016),  $(\mu_N, \tilde{x}_N)$  is feasible in (4.4) w.r.t.  $(\mu_{S \setminus j}, \tilde{x}_{S \setminus j})_{j \in S}$ .<sup>20</sup> But,

$$\sum_{i \in N} v_i(\mu_N, t, \lambda, \alpha) + \tilde{x}_N^i(t) = \max_{d \in D} \sum_{i \in N} v_i(d, t, \lambda, \alpha), \quad \forall t \in T.$$

Hence,  $(\mu_N, \tilde{x}_N)$  is an optimal solution of (4.4) w.r.t.  $(\mu_{S \setminus j}, \tilde{x}_{S \setminus j})_{j \in S}$ . We conclude that  $\tilde{\eta}$  is an egalitarian solution w.r.t.  $\lambda$  and  $\alpha$ .  $\square$

Theorem 1 says that, in our model with sidepayments, the M-solution and the H-solution are interim utility equivalent. Moreover, as it is also deduced from the proofs of Propositions 2 and 3, any M-solution is an H-solution and viceversa<sup>21</sup>. Henceforth, a cooperative solution will be simply called a *MH-solution*. Notice that if  $\Gamma$  is a game with complete information (i.e.,  $T_i$  is a singleton for every  $i \in N$ ), there are no incentive constraints (or equivalently  $\alpha = 0$ ) and consequently (3.2) implies that  $\lambda_i = 1$  for every  $i \in N$ . Hence, all efficient mechanisms are supported by the same utility weights  $\lambda_i = 1$ , which means that the Pareto efficient frontier is thus characterized by an hyperplane. From Definition 1, it then follows that a MH-solution coincides with the Shapley TU value<sup>22</sup>. A MH-solution is thus a valid generalization of the Shapley TU value to Bayesian cooperative games with sidepayments.

Under incomplete information, however, (3.2) only implies that

$$\sum_{t_i \in T_i} \lambda_i(t_i) = \sum_{t_i \in T_i} p(t_i) = 1, \quad \forall i \in N$$

therefore  $\lambda_i(t_i) > 0$  for some  $t_i$  for each player  $i \in N$ , but we still have  $\sum_{i \in N} (|T_i| - 1)$  degrees of freedom for choosing  $\lambda$ . With incomplete information, the restricted nature of the utility transfers implies that the interim incentive efficient frontier is not generally an hyperplane. This indeterminacy makes utility transfers less useful for games with incomplete information than it is for games with complete information. In particular, it makes imposible to get a general and simple closed form expression for the value as in the case of the Shapley value for TU games with complete information.

**Example 1** Consider the following Bayesian cooperative game introduced by Salamanca (2017). Let  $r$  be a parameter with  $0 < r < 1/2$ . The set of players is  $N = \{1, 2, 3\}$ . Only player 1 has private information about one of two possible states  $T = \{H, L\}$  with prior probabilities

<sup>19</sup>In particular,  $(\mu_N, \tilde{x}_N)$  is also a M-solution supported by  $\tilde{\eta}$ ,  $\lambda$  and  $\alpha$ .

<sup>20</sup>Using the terminology and results developed by Salamanca (2016), for any  $S \subset N$  we have that  $\tilde{\eta}_S = (\mu_R, \tilde{x}_R)_{R \subset S}$  is egalitarian w.r.t.  $\lambda$  and  $\alpha$ . Then, applying Proposition 2 in his paper, for any  $S \subset N$ ,  $\tilde{\eta}_S$  is equitable w.r.t.  $\lambda$  and  $\alpha$ . But  $U(\mu_N, \tilde{x}_N)$  is warranted by  $\lambda$ ,  $\alpha$  and  $\tilde{\eta}$ , therefore by Lemma 1 in the same paper,  $(\mu_N, \tilde{x}_N)$  is equitable for  $N$  w.r.t.  $\lambda$ ,  $\alpha$  and  $\tilde{\eta}$ . We conclude that  $\tilde{\eta}$  is equitable w.r.t.  $\lambda$  and  $\alpha$ , thus by Proposition 2 again  $\tilde{\eta}$  is egalitarian w.r.t.  $\lambda$  and  $\alpha$ , which means, in particular, that  $(\mu_N, \tilde{x}_N)$  is feasible in (4.4) w.r.t.  $(\mu_{N \setminus j}, \tilde{x}_{N \setminus j})_{j \in N}$ ,  $\lambda$  and  $\alpha$ .

<sup>21</sup>Notice however that Definitions 1 and 2 are not equivalent: an optimal solution to (4.3) is not necessarily also an optimal solution to (4.4), unless an appropriate transfers scheme is used (see proof of Proposition 3).

<sup>22</sup>Since  $\alpha = 0$  and  $\lambda_i = 1$  for every  $i \in N$ , the TU characteristic function can be simply described by  $v(S) = \max\{\sum_{i \in S} u_i(\mu_S) \mid \mu_S \in \Delta(D_S)\}$ . Thus, condition (ii) implies that, for every  $S \subseteq N$ , a rational threat achieves  $v(S)$ . Hence, condition (iii) reduces to  $u_i(\mu_N) = \phi_i(N, v)$ .

$p(H) = 1 - p(L) = 4/5$ . Feasible decisions for each coalition are:  $D_{\{i\}} = \{d_i\}$  ( $i = 1, 2, 3$ ),  $D_{\{i,j\}} = \{[d_i, d_j], d_{ij}\}$  ( $i \neq j$ ) and  $D_N = D_{\{1,2\}} \times D_{\{3\}} \cup D_{\{1,3\}} \times D_{\{2\}} \cup D_{\{2,3\}} \times D_{\{1\}}$ . For each value of  $r$ , utility functions are given by

$(u_1, u_2, u_3)$	$[d_1, d_2, d_3]$	$[d_{12}, d_3]$	$[d_{13}, d_2]$	$[d_{23}, d_1]$
$H$	$(0, 0, 0)$	$(50, 50, 0)$	$(100r, 0, 100(1-r))$	$(0, 100r, 100(1-r))$
$L$	$(0, 0, 0)$	$(40, 40, 0)$	$(40r, 0, 40(2-r))$	$(0, 40r, 40(2-r))$

Decision  $[d_1, d_2, d_3]$  denotes the noncooperative outcome which leaves each player with his reservation utility (normalized to zero), whereas  $[d_{ij}, d_k]$  denotes the cooperative outcome in which players  $i$  and  $j$  form a coalition (leaving player  $k$  alone) and they share the proceeds from cooperation as specified above.

Let us consider the game when sidepayments are *not* allowed, i.e., we set  $x_S^i(t_S) = 0$  for all  $i \in S$ ,  $t_S \in T_S$  and  $S \subseteq N$ . A mechanism for a coalition  $S$  is thus only composed by a state-dependent lottery  $\mu_S : T_S \rightarrow \Delta(D_S)$ .<sup>23</sup> In this case, no matter the state, player 3 can only offer players 1 and 2 a payoff which is strictly lower than what they both can get by acting together (lack of transferability). Hence, a reasonable outcome for this game should reward player 3 strictly less than players 1 and 2. On the other hand, incentive constraints do not reduce interim efficiency, that is, incentives play no role in this game. This implies that player 1 has no information rents ( $\alpha = 0$ ) and therefore, 1 and 2 must be treated symmetrically. In addition, as  $r$  goes to zero, 3's ability to transfer payoffs decreases, thus his reward should also reduce.

For any value of  $r$ , the unique M-value of this game is the interim allocation<sup>24</sup>

$$(U_1^H, U_1^L, U_2, U_3) = \left(\frac{100}{3}, \frac{80}{3}, 32, 32\right) \quad (4.10a)$$

The M-value treats all players symmetrically, and this irrespectively of the value of  $r$ . On the other hand, the unique H-value is the allocation<sup>25</sup>

$$(U_1^H, U_1^L, U_2, U_3) = \left(50 - \frac{100}{3}r \left(\frac{88-88r}{96-88r}\right), 40 - \frac{80}{3}r \left(\frac{88-44r}{96-88r}\right), 48 - \frac{88}{3}r, \frac{176}{3}r\right) \quad (4.10b)$$

The arguments presented above are better reflected by this allocation. In particular, as  $r$  decreases to 0, the H-value converges to  $(50, 40, 48, 0)$ .

Let us now turn to the situation in which players are allowed to transfer payoffs using state-contingent sidepayments. In such a situation, utility is fully transferable<sup>26</sup>, so that in any two-person coalition with player 3, its members can agree on an arbitrary distribution of the proceeds on every state. In particular, if coalition  $\{1, 3\}$  or  $\{2, 3\}$  forms, its members can get half of the gains each. There is no puzzle in the present case; the three players are symmetric, thus the allocation (4.10a) is a reasonable outcome. In fact, (4.10a) is the unique MH-value of the game with sidepayments.

<sup>23</sup>When sidepayments are not allowed, the hydraulic equations in (3.2) are removed from Proposition 1. Moreover, condition (3.4) is automatically satisfied. Therefore, conditions (3.2) and (3.4) disappear from Definitions 1 and 2.

<sup>24</sup>The value is supported by the vectors  $\lambda = (4/5, 1/5, 1, 1)$  and  $\alpha = (0, 0)$ .

<sup>25</sup>The value is supported by the same vectors  $\lambda$  and  $\alpha$  as the M-value.

<sup>26</sup>Since  $\alpha = 0$ , condition (3.2) then implies that  $\lambda_i(t_i) = p(t_i)$  for all  $t_i \in T_i$  of every player  $i \in N$ . Real and virtual utilities coincide, hence the interim Pareto frontier is an hyperplane.

## 5 Values for Two-person Bayesian Games

In this section, we shall study in detail both the M-solution and the H-solution in two-player games ( $n = 2$ ). In order to explore all the particularities of this case, we first consider games with orthogonal coalitions in which sidepayments are not allowed (as in Example 1). We then extend the analysis to allow for both strategic and informational externalities. In this latter situation, we investigate an additional semi-cooperative solution introduced by Kalai and Kalai (2013) under the further assumptions of sidepayments and ex-post verifiable information.

In the previous section we have shown that in our model with orthogonal coalitions and sidepayments, the M-solution and the H-solution are interim utility equivalent. This holds, in particular, for the two-player case. The question now is whether the same result also extends to a situation in which sidepayments are not allowed. We notice that when there are just two players, the only intermediate coalitions are the singletons  $\{i\}$  ( $i \in \{1, 2\}$ ). For any of these coalitions, the optimization problem in (4.4) is the same as in (4.3).<sup>27</sup> Therefore, it follows immediately that Definitions 1 and 2 are equivalent<sup>28</sup>. This reasoning is summarized in the following proposition (which already appeared in Salamanca (2016)).

### Proposition 4.

Let  $\Gamma$  be a two-player Bayesian cooperative game with orthogonal coalitions (with or without sidepayments). Then, the M-solution and the H-solution for  $\Gamma$  coincide.

Thus, we may continue to call a M(or H)-solution simply a MH-solution.

Let  $\Gamma$  be a two-person cooperative games with orthogonal coalitions. Suppose that, for every  $i \in \{1, 2\}$ ,  $D_i = \{d_i\}$  and  $u_i(d^*, t) = 0$  for all  $t \in T$ , where  $d^* := [d_1, d_2](= D_1 \times D_2 \subseteq D_N)$  is called the *disagreement decision*. To make this game interesting, we further assume that there exists at least one joint decision in  $D_N$  which is beneficial for both individuals. Such a game is called a *two-person bargaining problem with incomplete information*. Clearly, the issue of coalitional threats does not arise in this model. For this kind of problems, Myerson (1984a) defined a *neutral bargaining solution* which generalizes Nash's (1950) bargaining solution with fixed threats. The next result follows from the characterization theorem 5 in Myerson (1984a)<sup>29</sup>.

### Proposition 5.

Let  $\Gamma$  be a two-person bargaining problem with incomplete information. Then, the MH-solution and the neutral solution for  $\Gamma$  coincide.

In the subsequent analysis we allow for strategic externalities, that is, we let the utility of every player to depend on the choices of the other individual. We will represent the underlying game situation by a *non-cooperative* Bayesian game in strategic form  $\Gamma = \{N, (T_i, D_i, u_i)_{i \in N}, p\}$ , where  $N = \{1, 2\}$ . All components of  $\Gamma$  have the same interpretation as in the model of the previous section, but the decision set  $D$  is now defined to be  $D = D_1 \times D_2$ . For this specific model, we extend the MH-solution. We follow the method developed by Nash (1953) in the proposal to

<sup>27</sup>This conclusion is independent of whether sidepayments are allowed or not.

<sup>28</sup>Recall from the previous section that the M-solution and the H-solution differ only in the way they determine the threats for intermediate coalitions.

<sup>29</sup>see also Myerson (1991, sec. 10)

modify his bargaining solution (Nash (1950)) by making the threat point endogenous. The generalization here presented is not new. Indeed, Myerson (1984b) offers a more general approach to the case  $n \geq 2$  without sidepayments.

Let  $(\lambda, \alpha)$  be some fixed virtual scales and consider the virtual game in which players make interpersonal utility comparison in the scales  $(\lambda, \alpha)$ . As in the previous section, before entering into negotiations within the grand coalition, each player  $i$  commits to a threat strategy (mechanism)  $\mu_i : T_i \rightarrow \Delta(D_i)$  to be used in case the players fail to reach a cooperative agreement<sup>30</sup>. Then,  $v_i((\mu_i, \mu_j), t, \lambda, \alpha)$  is the disagreement virtual payoff to player  $i \in \{1, 2\}$  in state  $t \in T$ . The total transferable virtual utility in state  $t$  available to both players if they cooperate is

$$w_{12}(t, \lambda, \alpha) := \max_{d \in D} (v_1(d, t, \lambda, \alpha) + v_2(d, t, \lambda, \alpha))$$

Therefore, the Shapley TU value (or equivalently, the Nash bargaining solution) of player  $i$  in the virtual game conditional on state  $t$  is

$$\begin{aligned} \phi_i(W((\mu_1, \mu_2), t, \lambda, \alpha)) &= v_i((\mu_i, \mu_j), t, \lambda, \alpha) \\ &+ \frac{1}{2} [w_{12}(t, \lambda, \alpha) - v_i((\mu_i, \mu_j), t, \lambda, \alpha) - v_j((\mu_i, \mu_j), t, \lambda, \alpha)] \end{aligned} \quad (5.1)$$

That is, the total virtual surplus “ $w_{12} - v_i - v_j$ ” is equally divided among the two players. Suppose now that players expect to reach a cooperative agreement  $(\mu_N, x_N)$  giving every type of each player his conditionally expected Shapley TU value from the virtual game. Then, the real interim utility allocation  $U(\mu_N, x_N) = (U_i(\mu_N, x_N | t_i))_{t_i \in T_i, i \in N}$  must be warranted by  $\lambda, \alpha$  and  $(\mu_i, \mu_j)$ , that is<sup>31</sup>,

$$\begin{aligned} \left( \lambda_i(t_i) + \sum_{\tau_i \in T_i} \alpha_i(\tau_i | t_i) \right) U_i(\mu_N, x_N | t_i) - \sum_{\tau_i \in T_i} \alpha_i(t_i | \tau_i) U_i(\mu_N, x_N | \tau_i) = \\ \sum_{t_j \in T_j} p(t) \phi_i(W((\mu_1, \mu_2), t, \lambda, \alpha)), \quad \forall t_i \in T_i. \end{aligned} \quad (5.2)$$

It must be the case that players should evaluate their threats only in terms of their impact on the payoffs granted by the final agreement  $(\mu_N, x_N)$ . By Lemma 1 in Myerson (1983), the final warranted payoffs are (weakly) increasing in the right hand side of (5.2). Then each player  $i$  should want to choose his threat strategy  $\mu_i$  so as to maximize his expected Shapley TU value. Notice that the virtual payoff to player  $i$  in the Shapley TU value increases as the disagreement virtual payoff to player  $j$  decreases. Therefore, an optimal threat strategy for player  $i$  should solve

$$\max_{\mu_i \in \mathcal{F}_i} \sum_{t \in T} p(t) (v_i((\mu_i, \mu_j), t, \lambda, \alpha) - v_j((\mu_i, \mu_j), t, \lambda, \alpha)),$$

given  $j$ 's threat strategy  $\mu_j$ . That is, optimal threats form an equilibrium of a two-person Bayesian zero-sum game in which the players' utility functions are given by

$$w_i((\mu_i, \mu_j), t, \lambda, \alpha) := v_i((\mu_i, \mu_j), t, \lambda, \alpha) - v_j((\mu_i, \mu_j), t, \lambda, \alpha). \quad (5.3)$$

<sup>30</sup>The use of transfers inside a single person coalition is unnecessary for the coalition member.

<sup>31</sup>Myerson (1984a, sec. 9) provides a rationale for equation (5.2).

We let  $\Gamma^{\text{ad}}(\lambda, \alpha)$  denote this zero-sum game. The superscript “ad” stands for *advantage*, since the utility scale  $w_i(\cdot, \lambda, \alpha)$  is a measure of the payoff advantage of player  $i$  in the virtual game<sup>32</sup>.

**Definition 3 (MH-Solution).**

Let  $\Gamma$  be a two-player noncooperative Bayesian game with sidepayments. A feasible mechanism for the grand coalition  $(\mu_N, x_N)$  is a *MH-solution* of  $\Gamma$ , if there exist vectors  $\lambda > 0$ ,  $\alpha \geq 0$ , and threat strategies  $(\mu_1, \mu_2)$  such that:

- (i)  $(\mu_N, x_N)$  satisfies (3.1)-(3.4) for  $\lambda$  and  $\alpha$ .
- (ii)  $(\mu_1, \mu_2)$  is a profile of minmax strategies of  $\Gamma^{\text{ad}}(\lambda, \alpha)$ .
- (iii) The vector  $U(\mu_N, x_N)$  of interim utilities satisfies (5.2) w.r.t.  $\lambda$ ,  $\alpha$  and  $(\mu_1, \mu_2)$ .

The vector  $U(\mu_N, x_N)$  is called a *MH-value*. In the case that sidepayments are not allowed, so that  $x_N^i(t) = 0$  for all  $i \in \{1, 2\}$  and  $t \in T$ , conditions (3.2) and (3.4) are removed from (i).

It is worth noticing that Definition 3 generalizes Nash’s (1953) bargaining solution with variable threats (cf. Myerson (1991, sec. 8)).

Recently, A. Kalai and E. Kalai (2013) proposed a semi-cooperative solution for the class of two-player noncooperative Bayesian games with sidepayments. Their solution, called the *cooperative-competitive (or “coco”) value*, is based on a decomposition of the game  $\Gamma$  into cooperative and competitive components. Let  $\Gamma^{\text{ad}}$  be the Bayesian zero-sum game defined as  $\Gamma$  except that player  $i$ ’s utility function is now  $u_i - u_j$ . In a similar fashion, we define the game  $\Gamma^{\text{eq}}$ , which differs from  $\Gamma$  in that both players share the same utility function given by  $u_i + u_j$ . The superscript “eq” stands for equal payoffs. Clearly,  $u_i = \frac{1}{2}(u_i + u_j) + \frac{1}{2}(u_i - u_j)$ . Thus, the game  $\Gamma^{\text{eq}}$  describes the *cooperative* component of  $\Gamma$  in which the interests of both players are aligned. On the other hand, the game  $\Gamma^{\text{ad}}$  reflects each player’s *competitive* advantage in  $\Gamma$ . Incentives to reveal private information are opposed in both components games. While in  $\Gamma^{\text{eq}}$  the obvious incentive is to truthfully disclose any private information, in  $\Gamma^{\text{ad}}$  each player would like to prevent any information disclosure increasing the opponent’s payoff advantage.

The *team optimum* of  $\Gamma^{\text{eq}}$  is defined by

$$\text{Opt}(\Gamma^{\text{eq}}) = \sum_{t \in T} p(t) \max_{d \in D} \sum_{i \in N} u_i(d, t)$$

In words, the team optimum is the maximum expected utility that each player can get in  $\Gamma^{\text{eq}}$  when they all share truthfully their information.

The game  $\Gamma^{\text{ad}}$  is a zero-sum game in which each player is trying to maximize the difference of his payoff and that of his opponent. This game has a unique minmax (ex-ante) expected value denoted  $\text{Val}_i(\Gamma^{\text{ad}})$ . The *coco value* of  $\Gamma$ , denoted  $\kappa_i(\Gamma)$ , is defined by

$$\kappa_i(\Gamma) = \frac{1}{2} \left( \text{Opt}(\Gamma^{\text{eq}}) + \text{Val}_i(\Gamma^{\text{ad}}) \right)$$

---

<sup>32</sup>Myerson (1991, sec. 8) refers to  $\Gamma^{\text{ad}}(\lambda, \alpha)$  as the *difference game*. We keep the designation “advantage” in order to be consistent with Kalai and Kalai’s (2013) terminology.

The MH-solution deeply differs from the coco value. On one hand, the coco value assumes that the incentives in playing each of the component games are independent of each other, thus it ignores possible incentive compatibility issues. As a consequence, the coco value is ex-post (first best) efficient. In contrast, the MH-solution takes into account the tension that may exist between ex-post efficiency and incentive compatibility (see Myerson and Satterthwaite (1983)). On the other hand, the coco value is defined only at the ex-ante stage so that its interim evaluation cannot be determined without specifying a protocol for its interim implementation. Finally, in the MH-solution players bargain over mechanisms and not directly over payoffs as in the coco value. Despite these differences, there is a close relationship between these two solution concepts when incentive constraints are unnecessary, in the sense that incentive compatibility is not an issue for the implementation of the final agreement.

The simplest bargaining situation in which incentive constraints are not required is when all private information becomes publicly verifiable at the implementation stage. In principle, therefore, any budget-feasible mechanism in  $\mathcal{F}_N$  can be enforced once it is agreed upon. We summarize this situation in the following assumption introduced by de Clippel and Minelli (2004):

**Definition 4 (Verifiable types).**

A game  $\Gamma$  satisfies the *verifiable types assumption* if, when agreements are implemented, each individual can costlessly verify the true information state<sup>33</sup>.

The verifiable types assumption is satisfied in games where private information relates to signals about states that eventually become public. For instance, differential forecasts about weather conditions, as in Kalai and Kalai's (2013) vendors example. This assumption is particularly strong in the sense that it rules out game situations such as auctions or trading problems where payoffs depend on valuations which are not observable ex-post. Kalai and Kalai's (2013) non-cooperative interim implementation of the coco value heavily relies on this observability assumption, which severely limits its practical applicability.

Let  $(\mu_N, x_N)$  be a MH-solution supported by the virtual scales  $(\lambda, \alpha)$  and the minmax strategies  $(\mu_1, \mu_2)$ . The verifiable types assumption is equivalent to setting  $\alpha = 0$ , since incentive constraints are dispensable. Under such circumstance, equation (3.1) is straightforwardly satisfied. Condition (3.2) becomes  $\lambda_i(t_i) = p(t_i)$  for all  $t_i \in T_i$  and  $i \in N$ . Hence, virtual utilities and real utilities coincide. Conditions (3.3) and (3.4) become:

$$\max_{d \in D} \sum_{i \in N} u_i(d, t) \quad \text{and} \quad \sum_{i \in N} x_N^i(t) = 0, \quad \forall t \in T.$$

On the other hand, the zero-sum game  $\Gamma^{\text{ad}}(\lambda, \alpha)$  coincides with the competitive component

---

<sup>33</sup>In our model, the entire game structure  $\Gamma$  is commonly known to all players, so that each player knows how much each player would have received in every information state given any combination of strategies. Therefore, the states of nature are completely determined by the information states. As a consequence, the verifiable types assumption is equivalent to Kalai and Kalai's (2013) (*unrestricted*) *revealed-payoff assumption*.

game  $\Gamma^{\text{ad}}$  of  $\Gamma$ . Finally, the warrant equations (5.2) simplify to

$$U_i(\mu_N, x_N | t_i) = \frac{1}{2} \sum_{t_j \in T_j} p(t_j) \max_{d \in D} \sum_{k \in N} u_k(d, t) + \frac{1}{2} \sum_{t_j \in T_j} p(t_j) (u_i((\mu_i, \mu_j), t) - u_j((\mu_i, \mu_j), t)), \quad \forall t_i \in T_i, \forall i \in N.$$

Hence, the ex-ante evaluation of any MH-solution of  $\Gamma$  is

$$\sum_{t_i \in T_i} p(t_i) U_i(\mu_N, x_N | t_i) = \frac{1}{2} \sum_{t \in T} p(t) \max_{d \in D} \sum_{i \in N} u_i(d, t) + \frac{1}{2} \text{Val}_i(\Gamma^{\text{ad}}) = \kappa_i(\Gamma)$$

The following result is deduced:

**Theorem 2.**

Let  $\Gamma$  be a two-player Bayesian noncooperative game with sidepayments satisfying the verifiable types assumption. Then, the MH-solution and the coco value of  $\Gamma$  are ex-ante utility equivalent.

Notice that each player might have multiple minmax strategies in the game  $\Gamma^{\text{ad}} (= \Gamma^{\text{ad}}(\lambda, \alpha))$ . The multiplicity of equilibria in the competitive zero-sum component game may lead to different type-conditional vectors payoffs<sup>34</sup>. Of course, all combinations of minmax strategies are equilibria with the same ex-ante expected payoff (minmax value). Therefore, there might exist various MH-values, all of them guaranteeing the same ex-ante expected payoffs, in accordance with Theorem 2. This indeterminacy of the conditional payoffs makes difficult to get an appropriate definition of the “interim” coco value.

The following bilateral trading problem, adapted from Myerson (1985, 2007), analyzes how the HM-solution relates to Myerson and Satterthwaite’s (1983) result about the impossibility to achieve an ex-post efficient trade. In particular, it illustrates a situation in which the applicability of the coco value is limited by its assumptions.

**Example 2** There are two players  $N = \{1, 2\}$ . Player 1 is the seller of a single good whose quality may be high ( $H$ ) with probability  $p = 1/2$  or low ( $L$ ) with probability  $1 - p = 1/2$ . Player 2 is the only potential buyer. The quality can only be observed by the seller, that is,  $T_1 = \{H, L\}$  is the set of types of player 1. If the quality of the good is  $L$  then the value of the good is \$1 to player 1 and is \$2 to player 2. If the quality is  $H$  then the value of the good is \$5 to player 1 and is \$6 to player 2. Players must decide whether to trade ( $T$ ) or not ( $NT$ ) and at which price. Utilities from no trade are normalized to be 0. The utility functions are:

		$L$		$H$	
$u_1, u_2$		$T$	$NT$	$u_1, u_2$	
$T$		-1, 2	0, 0	$T$	-5, 6    0, 0
$NT$		0, 0	0, 0	$NT$	0, 0    0, 0

<sup>34</sup>See the payoff matrices proposed in Section IV.C. of Kalai and Kalai (2013).



For the game under consideration, the verifiable types assumption says that the quality of the good becomes publicly observable at the execution of the contract. In particular, it implies that if trade does not occur, the buyer would still observe the quality of the good. This might be considered not reasonable in this particular game situation. Therefore, the first best interim implementation of the coco value is limited in this game. For illustrative purposes only, we compute the MH-solution under the verifiable types assumption (and a fortiori also the coco value).

The cooperative component game  $\Gamma^{\text{eq}}$  has a team optimum equal to 1, whereas the competitive component has a minmax value 0. Therefore, the coco value of this games is  $\kappa_1(\Gamma) = \kappa_2(\Gamma) = \frac{1}{2}$ . It can be easily checked that the conditional minmax value is also 0 in both states. Thus, the unique MH-value is the (ex-post efficient) interim allocation

$$(U_1^H, U_1^L, U_2) = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$$

This allocation is implemented by the mechanism

$$\mu_N((T, T) | H) = \mu_N((T, T) | L) = 1, \quad x_N^1(H) = \frac{11}{2} = -x_N^2(H), \quad x_N^1(L) = \frac{3}{2} = -x_N^2(L).$$

That is, trade occurs in both states at a price that gives each player an equal gain from the trade. This is an efficient and fair solution, yet it cannot be implemented whenever information is not verifiable. Indeed, this mechanism is not incentive compatible: type  $L$  will always gain by reporting that the good is of high quality.

Let us consider now the more reasonable situation in which the quality of the good is not verifiable before consumption takes place. It can be shown that all incentive efficient mechanisms in this example satisfy Proposition 1 for the virtual scales (see Myerson (1985)):

$$\lambda_1^H = \frac{5}{8}, \quad \lambda_1^L = \frac{3}{8}, \quad \lambda_2 = 1, \\ \alpha_1(L | H) = 0, \quad \alpha_1(H | L) = \frac{1}{8}.$$

The virtual utility game is thus described by the following payoff matrices:

		$L$		$H$	
		$T$	$NT$	$T$	$NT$
$v_1, v_2$	$T$	-1, 2	0, 0	-6, 6	0, 0
$v_1, v_2$	$NT$	0, 0	0, 0	0, 0	0, 0

The only difference between virtual utilities and actual utilities is for 1's type  $H$ , for which the virtual value of the good is 6. The advantage game  $\Gamma^{\text{ad}}(\lambda, \alpha)$  has a conditional minmax value that equals 0 in both states. Hence, the unique solution to the warrant equations in (5.2) is:

$$(U_1^H, U_1^L, U_2) = \left(\frac{1}{10}, \frac{1}{2}, \frac{1}{4}\right) \tag{5.4}$$

The virtually equitable and incentive efficient allocation in (5.4) is the unique MH-value of this game. It is implemented by the mechanism

$$\mu_N((T, T) | H) = \frac{1}{10}, \quad x_N^1(H) = 6 = -x_N^2(H), \\ \mu_N((T, T) | L) = 1, \quad x_N^1(L) = \frac{3}{2} = -x_N^2(L). \tag{5.5}$$

According to this mechanism, if player 1 announces that the quality is  $L$  then with probability 1 they trade at a price  $\$3/2$ , yet if player 1 reports a high quality then with probability  $1/10$  they trade at a price  $\$6$ . In state  $H$  no-trade has probability  $9/10$ , but the seller never trades at a price lower than his virtual valuation of the good, which guarantees that he extracts the whole surplus from trade.

At this point, it should be noted that the unresolvable tension between incentives and ex-post efficiency in Myerson and Satterthwaite's (1983) impossibility theorem holds only if we insist on (interim) individually rational outcomes. In the current model, individual rationality does not appear as an explicit constraint in the set of feasible mechanisms. However, it does appear as an indirect restriction imposed by the definition of the MH-solution. Indeed, the MH-solutions are individually rational (see theorem 3 in Myerson (1984b))<sup>35</sup>. To illustrate this issue, consider again the incentive compatible mechanism in (5.5). Notice that if the actual quality of the good is high, and of course type  $H$  is telling the truth, there is a chance of not trading. This efficiency loss reduces type  $H$ 's expected utility from cooperation, which tightens individual rationality. Therefore, the whole surplus from cooperation must be given to type  $H$  whenever trade occurs.

The coco value has been constructed to be a general semi-cooperative solution with a tractable definition. Its description by a simple closed form expression comes at the cost of restrictive conditions such as two-person games and verifiable information. Theorem 2 exhibits why relaxing these assumptions cannot be easily done while preserving the tractable nature of the coco value. Firstly, keeping track of the incentive constraints requires referring to an appropriate interim framework. In particular, bargaining directly over payoffs, as in the coco value, does not allow to determine what the utility of a player would be when he reports a different type. In contrast, incentive compatible mechanisms include all equilibria that can be achieved by any communication system, and so this set can be viewed as the feasible set. Also, interim preferences matter when defining a suitable definition of an equitable cooperative agreement under incomplete information. As Myerson (1983, 1984a,b) argues, identifying a cooperative solution among the set of incentive efficient mechanisms requires to define some principles for equitable compromises not only among the different players (as in the case of complete information), but also among the different possible types that a player pretends to be. Secondly, even under the verifiable types assumption, there is no a unique way of extending the coco value to  $n > 2$ . Myerson's (1984b) approach provides a reference in this direction. However, other cooperative solutions can be constructed to achieve this goal. For instance, proceeding as in Harsanyi's (1963) solution for  $n$ -person strategic games with complete information, the H-solution can be straightforwardly generalized to allow for both strategic and informational externalities.

---

<sup>35</sup>Even though we could restrict attention to individually rational mechanisms, we cannot apply Myerson and Satterthwaite's (1983) result to our example simply because they require a continuum of types.



# The Value of Mediated Communication

## 1 Introduction

This paper provides an analytical framework for studying *Bayesian persuasion* problems in which the sender cannot commit himself to truthfully communicate his information to the receiver, so that incentive compatibility becomes one of the major issues for communication to be meaningful<sup>1</sup>. By allowing the two players to communicate with a neutral third party, we are able to solve two analytical problems that could possibly prevent a tractable analysis of incentive compatibility: first, truthful revelation of information when communication is direct may considerably limit the ability of the sender to credibly signal his information by himself. In particular, Forges (1985) and Farrell (1993) propose some examples in which no substantive communication can occur between the players. However, it is well known that the set of implementable outcomes may be strictly larger when players use mediated rather than direct communication (see for instance, Forges (1985, 1990)). Second, revelation of influential information with direct communication requires the sender to be indifferent between all signals he sends with positive probability. This is a strong form of incentive compatibility that reflects the fact that the sender sends a random signal by himself. In contrast, when the players use a mediator to perform the randomization of the signals on behalf of the sender, incentive compatibility will only demand each type of the sender to prefer the expected allocation designated to him.

In a recent pioneering work, Kamenica and Gentzkow (2011) offer a general approach to Bayesian persuasion under full commitment on the part of the sender. They consider a sender-receiver game in which before learning his type (*ex-ante* stage), the sender publicly chooses a signaling strategy, i.e., a conditional distribution of signals for each of his types, that he will use for transmitting his information to an uninformed receiver<sup>2</sup>. The sender produces a signal according to his true type and the corresponding distribution of signals. He cannot distort

---

<sup>1</sup>This chapter corresponds to the paper: *The Value of Mediated Communication*, HAL Working paper 01289379, 2016.

<sup>2</sup>Kamenica and Gentzkow (2011) refer to a signaling strategy simply as a signal. In order to distinguish the conditional distribution from its realizations, they call the latter *signal realization*. This paper follows the terminology developed in the literature of communication games.

the signal realization, nor can he misrepresent his information (full commitment assumption). The receiver observes the signal realization and then takes an action that affects the welfare of both individuals. Drawing on a geometric approach developed by Aumann and Maschler (1995), Kamenica and Gentzkow characterize the sender's optimal value of persuasion. They first construct a *non-revealing payoff function* over prior beliefs,  $\hat{a}(\cdot)$ , describing the (ex-ante) expected equilibrium payoffs the sender can achieve in the absence of communication. Then, they compute the *concavification* of  $\hat{a}(\cdot)$ , denoted  $\text{cav } \hat{a}(\cdot)$ , i.e., the smallest concave function that is larger or equal to  $\hat{a}(\cdot)$ . Their main result establishes that, for given prior beliefs  $p$ , the sender's optimal expected payoff (value of persuasion) is  $\text{cav } \hat{a}(p)$ .

Under full commitment, restricting attention to either direct or indirect communication systems does not change the value of persuasion. Yet, in the absence of commitment, the communication system determines the strategic opportunities the sender has for manipulating his private information and, thus, it might have an impact on the sender's payoffs. We consider the Bayesian persuasion setup studied by Kamenica and Gentzkow (2011), but with limited commitment and mediated communication. Under mediation, the sender reports a type to a neutral trustworthy mediator who then recommends an action to the receiver. The sender's report is not verifiable either by the mediator or the receiver, which allows the sender to strategically manipulate his private information. The mediator's recommendation is not binding, that is, the receiver is free to choose any action different from the recommended one. The mediator can only create value by controlling the flow of information between both players. He introduces noise in the communication, which may relax the incentive constraints faced by the sender. In our setting, a very broad form of the *revelation principle* applies and, thus, without loss of generality, we can restrict attention to mediation protocols in which the sender reports his type truthfully and the receiver obeys the prescribed recommendation. A mediation protocol in which the sender always reports the truth and the receiver always follows the recommendation is called a *communication equilibrium* (see Myerson (1986) and Forges (1986)). The sender's problem is thus to select a communication equilibrium maximizing his ex-ante expected payoff.

The Lagrange multipliers associated to the truth-telling incentive constraints yield "shadow prices",  $\gamma$ , that can be used to define the sender's *virtual utility* (see Myerson (1991, ch. 10)). These virtual utility scales incorporate into the sender's utility function the signaling costs associated with incentive compatibility<sup>3</sup>. Considering virtual utilities rather than real utilities, we construct a *fictitious persuasion problem* in which there are no truth-telling incentive constraints. For this game, the non-revealing payoff function,  $\hat{\alpha}(\cdot; p, \gamma)$ , depends on the prior probability  $p$  and the signaling costs  $\gamma$ . Our main result (Theorem 1) says that the sender's optimal expected payoff at the prior belief  $p$ , denoted  $a^*(p)$ , equals the value of persuasion in the fictitious game with virtual scales defined by the optimal signaling costs, namely,

$$a^*(p) = \min_{\gamma} \text{cav } \hat{\alpha}(p; p, \gamma).$$

Ex-post inefficiencies derived from the signaling costs are geometrically expressed in the form of non convexities in the subgraph (or hypograph) of  $a^*$ . That is, the optimal value of persuasion

---

<sup>3</sup>The approach is similar to the one used in Auction theory to define the virtual surplus that takes into account the bidders' information rents (see Myerson (1981)).

may not be a concave function of the prior beliefs, as it is in the full commitment model. In fact, it may contain convex segments lying strictly below  $\text{cav } \hat{a}(\cdot)$ . Moreover, it may also exhibit discontinuities.

We characterize the optimal mediation protocol through a constrained splitting of the prior belief into a distribution over posterior beliefs. Given the optimal signaling costs  $\gamma^*$ , the posterior beliefs induced by any optimal mediation protocol correspond to the points on the domain of  $\hat{\alpha}(\cdot; p, \gamma^*)$  for which the convex combination of their images yields  $a^*(p)$ . The corresponding distribution of posteriors is constrained by the Bayes plausibility (martingale property) together with *complementary slackness* conditions from Duality Theory (see Proposition 1). Complementary slackness says that if a shadow price is positive, then the associated informational incentive constraint must be binding.

We also provide an upper-bound on the number of messages that the sender requires to transmit in order to attain the value of persuasion. Under full commitment, the greatest number of messages the sender needs to convey does not exceed his number of types. However, when misrepresentation is problematic, the sender might need to transmit one extra message with every binding incentive constraint. Hence the number of messages is bounded by the total number of types plus the total number of incentive constraints (see Proposition 2). To understand the idea behind this result, consider for instance a situation in which the sender could be a “good” type or a “bad” type. Assume that he would prefer to be perceived as the good type, so that the incentive constraint asserting that the bad type should not gain by imitating the good type is binding. The good type of the sender would like to communicate his true type to the receiver (which requires every type to send a different message). However, since the receiver will face difficulties preventing the bad type from claiming to be the good type, the latter will need to find a way to separate himself from the bad type. To do this, he may commit to recommend some additional action that might be unpleasant for him but highly hurtful for the bad type.

We conclude the paper with some discussions about the cheap-talk implementation of the optimal mediation protocols and the extension of our approach to general information design problems.

**Related literature** Our analytical framework is the same as in Kamenica and Gentzkow (2011), except that we consider a more general interaction situation in which communication is mediated and the sender may strategically manipulate his private information. In that respect, our paper relates to the recent literature on information design known as Bayesian persuasion. To our best knowledge, this literature so far has been rather unsuccessful in developing a tractable approach for an explicit analysis of the sender’s informational incentive compatibility. It is worth mentioning that Kolotilin, Li, Mylovanov and Zapechelnyuk (2016) study Bayesian persuasion with a privately informed receiver. In their framework, the sender designs a communication device that gathers information from the receiver and then sends a recommendation to the receiver conditional on her report and the sender’s *true* type. In addition to the strategic incentive constraints ensuring that the receiver will follow the recommendation, the sender is also led to consider informational incentive constraints guaranteeing that the receiver finds it optimal to report truthfully her information. In their setup, the communication device is

a mediation rule unable to verify the receiver's private information, but capable of identifying the sender's type.

This paper also relates to the literature on contracting with limited commitment. This literature considers a principal-agent setup in which the principal (receiver) is imperfectly informed about the agent's (sender's) type. The principal cannot contractually commit herself to chose any action, however, she may extract information from the agent by using a communication system. Bester and Strausz (2001) study direct communication in which the agent simply sends a single message to the principal. In contrast, Bester and Strausz (2007) allow the principal to use general communication devices which may enlarge the set of implementable contracts. Contrary to Bayesian persuasion, here the communication device (mechanism) is designed by the uninformed party, i.e., the principal. However, because the agent cannot commit to truthfully transmit his information, informational incentive compatibility is a relevant matter.

Mitusch and Strausz (2005) and, Golstman, Hörner, Pavlov and Squintani (2009) compare different communication protocols in the framework of Crawford and Sobel (1982). In this regard, our paper is also connected with this literature. They study the conditions under which mediation improves upon direct communication. As with contracting problems with adverse selection and limited commitment, it is also assumed that the communication procedures are designed to maximize the ex-ante welfare of the receiver. However, due to the particular structure of payoffs, it turns out that this is also equivalent to maximize the ex-ante welfare of the sender.

Finally, it is worth mentioning that our model is mathematically analogous to a problem of mechanism design by an informed principal (as in Myerson (1983)) in which contracting takes place at the ex-ante stage<sup>4</sup>. Indeed, by using the concept of virtual utility, we borrow some analytical tools that were developed by Myerson (1983) in order to characterize his neutral optimum.

This paper is organized as follows. In the next section we present a motivating example. Section 3 is devoted to formally describing the basic interaction scenario. The concept of communication equilibrium is also defined. Section 4 introduces the mediated persuasion problem and the virtual utility approach. The main results are presented. It also contains some examples illustrating our findings. Finally, Section 5 presents some concluding discussions.

## 2 Motivating Example

In this section we study an example which motivates several aspects of our analytical framework. First, it illustrates how, in the absence of commitment, the necessity of players to trust each other may lead to inefficiencies derived from the signaling costs associated with incentive compatibility. Second, the example shows that mediated communication may help the sender to reach equilibrium outcomes that cannot be achieved under direct communication. Finally, the example provides an instance in which, by allowing for more signaling opportunities, me-

---

<sup>4</sup>Myerson (1983) considers contracting at the *interim* stage, whereby the principal chooses the mechanism after she has received her private information. This is a more involved problem, as the choice of the mechanism may signal information about the principal's type.

diated communication compels the sender to disclose more information than when he is fully committed.

Consider the following sender-receiver game. Player 1 (the sender) has a privately known type that may be  $H$  with probability  $p = 1/10$  or  $L$  with probability  $1 - p = 9/10$ , and player 2 (the receiver) must choose an action from the set  $J = \{j_1, j_2, j_3\}$ . Payoffs for both players depend on the sender's type and the receiver's action as follows:

	$j_1$	$j_2$	$j_3$
$H$	1,3	3,1	-5,-3
$L$	-1,-3	2,-1	0,0

We can set this example in an economic situation described as follows. The informed player is a financial analyst knowing whether the general state of the financial markets is more favorable for investments in portfolios  $j_1$  and  $j_2$  (type  $H$ ) rather than in  $j_3$  (type  $L$ ). The uninformed player is an investor who must select among these three different portfolios offered by the analyst<sup>5</sup>. Thus, each portfolio generates an expected return for the investor that depends on the state. On the other hand, the analyst's preferences are explained by fact that he gets profits with investments in the portfolio  $j_2$  but he wants also to give good advice to the investor.

The expected payoff of the receiver, as a function of her belief  $q \in [0, 1]$  about the type  $H$ , is depicted in Figure 1. Thick lines denote her best-reply payoff (optimal actions appear above the corresponding best-reply payoff). At  $q = 1/5$  (resp.  $q = 1/2$ ) any randomization between  $j_3$  and  $j_2$  (resp.  $j_2$  and  $j_1$ ) is a best reply for the receiver.

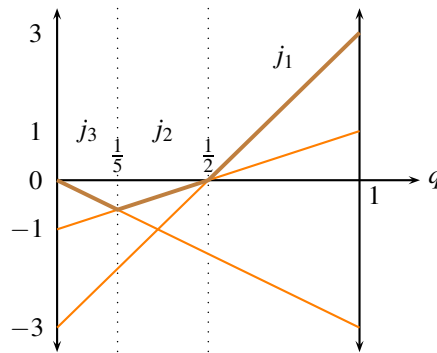


Figure 1: Receiver's expected payoffs (thin lines) and best-reply expected payoffs (thick lines)

Given the receiver's best-replies, the sender's (ex-ante) expected payoffs, as a function of the belief  $q \in [0, 1]$ , are represented in Figure 2. In particular, according to the prior belief  $p = 1/10$ , in the absence of communication, the receiver will choose action  $j_3$ . Thus, leaving the sender with an expected payoff equal to  $-1/2 (= -5 \times p + 0 \times (1 - p))$ .

Assume now that, in order to persuade the receiver to change her action, the sender publicly *commits* to disclose his information according to a signaling strategy (i.e., a conditional distribution of signals for each of his types). This commitment assumption means that the sender

<sup>5</sup>To keep the example as simple as possible, we suppose that the investor has no outside option.



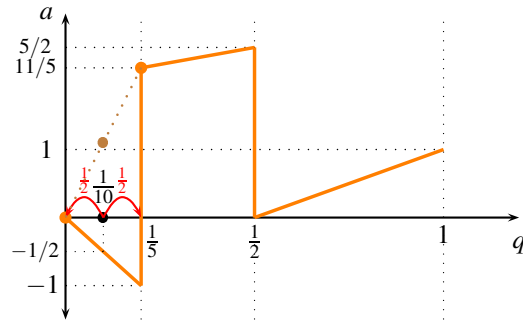
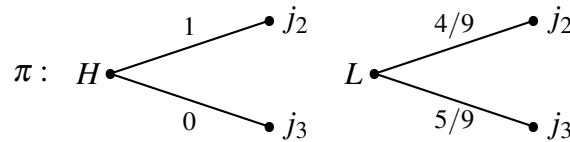


Figure 2: Non-revealing payoffs and optimal split under full commitment

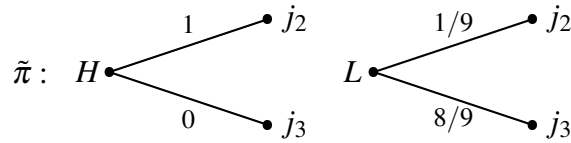
cannot distort the signal realization, nor can he misrepresent his private information. The problem for the sender is then to induce posterior beliefs leading the receiver to choose actions maximizing his expected payoff. Without loss of generality, the sender may choose to send “recommendations” in  $J$ . Then, the unique optimal signaling strategy can be described as follows:



According to  $\pi$ , the sender recommends  $j_2$  with probability 1 when he is type  $H$  and randomizes between  $j_2$  and  $j_3$  with probabilities  $4/9$  and  $5/9$  respectively, when he is type  $L$ . After receiving the recommendation to play  $j_2$  (resp.  $j_3$ ), the receiver forms a posterior belief  $p_2 = \frac{1}{5}$  (resp.  $p_3 = 0$ ). Since  $j_2$  (resp.  $j_3$ ) is optimal for the receiver at  $p_2$  (resp.  $p_3$ ), she will follow the recommendation. Both actions are recommended with an expected probability equal to  $1/2 (= 1 \times p + 4/9 \times (1 - p))$ , and thus the sender’s expected payoff is  $11/10 (= 1/2 \times 0 + 1/2 \times 11/5)$ . Figure 2 illustrates this situation.

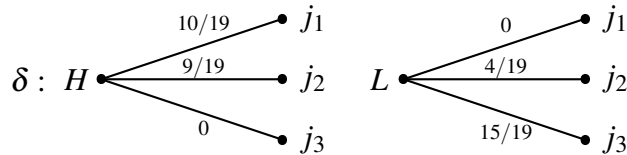
Let us now consider the situation in which commitment cannot be (legally) enforced, for instance because the sender’s private information is not verifiable by any authority. The question is then: why would the sender like to maintain his commitment? Knowing that the receiver will follow the recommendation, type  $L$  of the sender will send message  $j_2$  with probability 1. But then a message that is sent with probability 1 by both types will convey no information to the receiver. Therefore, in anticipation of this behavior, the receiver will rationally choose her optimal action at  $p = 1/2$ . As a consequence, the information transmitted by the sender is not credible and thus  $\pi$  has no persuasive effect. In the absence of commitment, the signaling strategy must be part of a Nash equilibrium of the underlying cheap-talk game. Thus, truthful revelation of information requires the sender to be indifferent between the distinct outcomes that his messages lead the receiver to choose. This is a strong form of incentive compatibility that may considerably limit the ability of the sender for credibly signaling his information. In this game, the most the sender can get in the absence of commitment is  $1/6$ . To achieve this payoff, he can use the signaling strategy<sup>6</sup>

<sup>6</sup>The signaling strategy  $\tilde{\pi}$  is part of the unique cheap-talk equilibrium of this game that maximizes the sender’s



After receiving the recommendation to play  $j_2$  and  $j_3$ , the receiver forms posterior beliefs  $p_2 = 1/2$  and  $p_3 = 0$ , respectively. At  $p_3$  action  $j_3$  is her unique best-reply, so she follows the recommendation. At  $p_2$  action  $j_2$  is optimal for the receiver, yet following this recommendation (with probability 1) will lead type  $L$  of the sender to recommend  $j_2$  with probability 1, thus making communication not credible. Instead, the receiver may randomize between actions  $j_1$  and  $j_2$  with probabilities  $2/3$  and  $1/3$ , respectively, which is also optimal for her at  $p_2$ . In this way, the receiver makes type  $L$  indifferent between the recommendations  $j_1$  and  $j_2$  and, thereby communicating his information according to  $\tilde{\pi}$  remains optimal for the sender.

Assume now that, instead of communicating his information directly to the receiver, the sender sends a confidential report about his type to an impartial mediator commissioned to produce a recommendation for the receiver. Because information is nonverifiable, even with the help of a mediator, the sender may strategically manipulate his information. For instance, consider a mediator that recommends actions according to  $\pi$ . This mediation rule is not incentive compatible for the sender. Indeed, type  $L$  would have incentives to report that he is type  $H$ . As a consequence, the sender cannot get  $11/10$ , yet he can do better than  $1/6$ . Consider a mediator that recommends actions as follows:



According to this mediation protocol, the receiver will infer posterior beliefs  $p_1 = 1$ ,  $p_2 = 1/5$  and  $p_3 = 0$ , where as before  $p_j$  denotes the probability of type  $H$  conditional on the recommendation to play  $j$ . For any  $j \in J$ , action  $j$  is a best-reply at  $p_j$ , then following the recommendation is optimal for the receiver. On the other hand, given that the receiver is obedient, no type of the sender has incentives to misrepresent his information. Indeed, by reporting to the mediator that he is type  $H$ , type  $L$  gets an expected payoff equal to  $8/19 (= -1 \times 10/19 + 2 \times 9/19 + 0 \times 0)$ , while by telling the truth he gets  $8/19 (= -1 \times 0 + 2 \times 4/19 + 0 \times 15/19)$ . That is, type  $L$  is indifferent, and thus he has no incentives to lie. A similar analysis reveals that type  $H$  has no incentives to misrepresent his information either.

The mediation protocol  $\delta$  gives the sender an (ex-ante) expected payoff equal to  $109/190 (\approx 0.57 > 1/6)$ . No other mediation rule can guarantee a higher expected payoff to the sender. In this particular game, mediation facilitates incentive compatibility, thus allowing the sender to achieve an outcome that cannot be attained under direct communication. Mediation alleviates the conflict between the incentives of both players. However, its potential benefits are reduced by the degree of such a conflict. In particular, the sender's expected payoff in the absence of

---

(ex-ante) expected payoff.

commitment is larger under mediation than under direct communication, but lower than what he gets under full commitment.

Finally, notice that  $\delta$  requires the sender to transmit 3 ( $> 2$ ) different messages with positive probability. That is, the sender communicates more information than just his true type (which has two possible values). Unlike the full commitment case, in which the number of messages need not exceed the number of types, when misrepresentation is problematic, the sender might need to disclose more information. The idea is that the sender requires to signal as much information as when he is fully committed (i.e., he needs to induce posterior beliefs  $p_3 = 0$ ,  $p_2 = \frac{1}{5}$ ), but also he needs to make such revelation credible to the receiver. Because type  $L$  has incentives to imitate type  $H$ , the latter type will need to find a way to separate himself from type  $L$ . To do this, type  $H$  can commit himself to recommend action  $j_1$  (with probability  $10/19$ ), something that is unpleasant for type  $L$ . In this manner, the receiver can discriminate between both sender's types, so that whenever  $j_1$  is recommended, she deduces that this message can only come from type  $H$ , i.e.,  $p_1 = 1$ , and thus, she follows the recommendation.

### 3 Basic Game

Our basic framework is a two-person finite Bayesian game in which player 1 has no decision to make, but is the only player to have private (nonverifiable) information. Let  $K$  be the (finite) set of types of player 1. A type  $k \in K$  is chosen according to<sup>7</sup>  $p \in \Delta(K)$ , and only player 1 is informed about  $k$ . We assume that  $p^k > 0$  for every  $k \in K$ . Player 2 chooses an action in a (finite) set  $J$ . When action  $j$  is chosen by player 2 and player 1 is of type  $k$ , then player 1 and player 2 get respective payoffs  $a_j^k$  and  $b_j^k$ . We refer to this *basic game* as  $\Gamma(p)$ .

A (*mediated*) *communication device*  $\delta$  is a mapping  $\delta : K \rightarrow \Delta(J)$ , namely a vector of probability distributions  $(\delta^k)_{k \in K}$  over  $J$  for every  $k \in K$ . By adding a communication device  $\delta$  to the game  $\Gamma(p)$ , one generates an extended game  $\Gamma_\delta(p)$ , which is played as follows:

1. A type  $k \in K$  is randomly chosen according to  $p$ .
2. Player 1 learns his type  $k \in K$ .
3. Player 1 sends a confidential report  $k' \in K$  to a mediator.
4. The mediator chooses an action  $j \in J$  with probability  $\delta_j^{k'}$ .
5. The mediator recommends the action  $j$  to player 2.
6. Player 2 chooses an action and both players receive payoffs as in  $\Gamma(p)$ .

For obvious reasons, we refer to player 1 in  $\Gamma_\delta(p)$  as the sender, and player 2 as the receiver. In this game, a strategy for the sender is a transition probability  $\tau : K \rightarrow \Delta(K)$  where  $\tau(k' | k)$  is the probability to report  $k'$  if his type is  $k$ . A strategy  $\tau$  is called *sincere* if  $\tau(k | k) = 1$  for every  $k \in K$ , namely, if the sender always reveals honestly his type to the mediator. A

---

<sup>7</sup>For any finite set  $A$ ,  $|A|$  denotes its cardinality and  $\Delta(A)$  denotes the set of probability distributions over  $A$ .

strategy for the receiver in  $\Gamma_\delta(p)$  is a transition probability  $\zeta : J \rightarrow \Delta(J)$  where  $\zeta(i | j)$  is the probability to choose  $i$  when  $j$  is recommended by the mediator. A strategy  $\zeta$  is called *obedient* if  $\zeta(j | j) = 1$  for every  $j \in J$ , i.e., if the receiver always follows the recommendation made by the mediator. When both players are sincere and obedient, respectively, in  $\Gamma_\delta(p)$ , the (ex-ante) expected payoff of the sender is

$$a(\delta; p) := \sum_{k \in K} p^k \sum_{j \in J} \delta_j^k a_j^k. \quad (3.1)$$

The communication device  $\delta$  is *incentive-compatible for the sender* if and only if the sincere strategy is a best response for the sender in  $\Gamma_\delta(p)$  whenever the receiver is obedient, that is,

$$\sum_{j \in J} \delta_j^k a_j^k \geq \sum_{j \in J} \delta_j^{k'} a_j^k, \quad \forall k, k' \in K. \quad (3.2)$$

The *informational incentive constraints* in (3.2) reflect the fact that neither the receiver nor the mediator can verify the sender's private information (adverse selection problem).

Suppose action  $j$  is recommended to the receiver according to the communication device  $\delta$ , provided that the sender is sincere in  $\Gamma_\delta(p)$ . Then, the receiver computes posterior probabilities  $p_j(\delta) = \left( p_j^k(\delta) \right)_{k \in K}$  given by

$$p_j^k(\delta) = \frac{\delta_j^k p^k}{\sum_{k' \in K} \delta_j^{k'} p^{k'}}. \quad (3.3)$$

The communication device  $\delta$  is *incentive-compatible for the receiver* if and only if the obedient strategy is a best response for the receiver in  $\Gamma_\delta(p)$  whenever the sender is sincere, namely,

$$\sum_{k \in K} p_j^k(\delta) b_j^k \geq \sum_{k \in K} p_j^k(\delta) b_i^k, \quad \forall i, j \in J. \quad (3.4)$$

The *strategic incentive constraints* in (3.4) characterize the receiver's inalienable right to control her action in  $J$  (moral hazard problem). By definition of the posterior probabilities in (3.3), both sides of (3.4) are divided by the total probability of receiving the recommendation to play  $j$ . Then, the strategic incentive constraints can be equivalently written as

$$\sum_{k \in K} \delta_j^k p^k b_j^k \geq \sum_{k \in K} \delta_j^k p^k b_i^k, \quad \forall i, j \in J. \quad (3.5)$$

We define  $Y(q)$  as the set of receiver's optimal actions at belief  $q \in \Delta(K)$ , i.e.,

$$Y(q) = \left\{ y \in \Delta(J) \mid \sum_{k \in K} q^k \sum_{j \in J} y_j b_j^k = \max_{j \in J} \sum_{k \in K} q^k b_j^k \right\}.$$

Let  $\pi_j(\delta) := \sum_{k \in K} p^k \delta_j^k$  be the probability of sending the recommendation  $j$  when  $\delta$  is implemented. Then,  $\delta$  is incentive compatible for the receiver if and only if for each  $j \in J$ ,  $\pi_j(\delta) > 0$  implies that  $j$  is optimal for the receiver given the posterior probabilities  $p_j(\delta)$ , i.e.,  $j \in Y(p_j(\delta))$ .

**Definition 1 (Communication equilibrium).**

A communication device  $\delta$  is a *communication equilibrium* of  $\Gamma(p)$  if and only if the sincere and obedient strategies form a Nash equilibrium of  $\Gamma_\delta(p)$ , that is,  $\delta$  satisfies the incentive constraints in (3.2) and (3.5). We let  $\mathcal{D}^*(p)$  denote the set of communication equilibria of  $\Gamma(p)$ .

REMARK 1. Communication equilibria are defined by a set of linear inequalities, hence the set  $\mathcal{D}^*(p)$  is a convex polyhedron. Furthermore, this set is closed, bounded and non-empty.

A communication equilibrium  $\delta$  is *fully revealing* (FR) if it recommends different actions for every type of the sender, so that the receiver can infer the true state by looking at the prescribed recommendation. It is *non-revealing* (NR) if  $\delta^k = \delta^{k'}$  for every  $k, k' \in K$ , so that no useful information is revealed to the receiver. It is *partially revealing* (PR) if it is neither NR nor FR.

Thanks to a very general form of the *revelation principle* for Bayesian games (see Forges (1985, 1986)), there is no loss of generality in restricting attention to communication equilibria, in the following sense: assume that the game  $\Gamma(p)$  is extended by allowing the players to communicate for a possibly infinite number of stages through a *general* communication device, sending signals to every player at every stage but also receiving messages from them. Such devices may involve preplay communication, before player 1 learns his type, but also interplay communication, after player 1 has learnt his type but before player 2 chooses his action. The set of all Nash equilibrium payoffs of all extensions of  $\Gamma(p)$  by general communication devices coincides with the set of all communication equilibrium payoffs.

## 4 Mediated Persuasion

In the basic game  $\Gamma(p)$ , player 1 has the option to remain silent and let player 2 to choose an action given her prior belief  $p$ . He can also design a communication system to signal his private information, and try to persuade player 2 to change her action. We assume that player 1 publicly chooses a mediated communication device  $\delta$  (i.e., a mediator) *before* learning his type. Then both players interact as in  $\Gamma_\delta(p)$ . Because the selection of the communication device is done at the ex-ante stage, this choice is by itself uninformative. The problem of player 1 is then to choose a communication equilibrium maximizing his ex-ante payoff, namely, to select a communication device solving

$$\max_{\delta \in \mathcal{D}^*(p)} a(\delta; p). \quad (4.1)$$

We shall refer to this optimization problem as the *primal problem for  $p$* .

REMARK 2. The optimization problem in (4.1) is a linear programming problem: the objective function is linear in  $\delta$  (see (3.1)) and the feasible set is defined by a system of linear inequalities in  $\delta$  (see Remark 1).

**Definition 2 (Value of persuasion).**

The optimal value of the primal problem for  $p$  will be called the *value of persuasion at  $p$*  and is denoted  $a^*(p)$ .

REMARK 3. Viewed as a correspondence defined on  $\Delta(K)$ , the set of communication equilibria is upper-hemicontinuous. Then,  $a^*$  is an upper-semicontinuous function. It may however fail to

be continuous.

#### 4.1 Mediated Persuasion Under Verifiable Information

Before proceeding with the analysis of the primal problem, let us consider the more simplified persuasion game in which the type of the sender is verifiable by the mediator but not by the receiver. In such a situation, the informational incentive constraints are not relevant, so that our framework reduces to Kamenica and Gentzkow's (2011).

##### Definition 3.

For a given prior  $p$ , we denote as  $\mathcal{D}(p)$  the set of communication devices satisfying the strategic incentive constraints in (4.5).

Under verifiable information, the problem for the sender can thus be expressed as

$$\max_{\delta \in \mathcal{D}(p)} a(\delta; p). \quad (4.2)$$

In this case, the sender has nothing to communicate to the mediator. The only thing he has to do is to choose a communication device that will recommend an action to the receiver depending on his *true* type. The verifiability assumption is thus equivalent to the full commitment assumption.

Given any prior belief  $q$ , the maximal ex-ante utility that the sender can expect in the absence of communication is

$$\hat{a}(q) := \max_{y \in Y(q)} \sum_{k \in K} q^k \sum_{j \in J} y_j a_j^k.$$

We refer to the function  $\hat{a}$  as the *non-revealing payoff function*. Let  $\text{cav } \hat{a}$  be the *concavification* of  $\hat{a}$ , i.e., the smallest concave function that is larger or equal to  $\hat{a}$ . As observed by Aumann and Maschler (1995) and Kamenica and Gentzkow (2011), the optimal value of the relaxed primal problem (4.2) is  $\text{cav } \hat{a}(p)$ . Also there exists a subset  $I \subseteq J$  of actions with  $|I| \leq |K|$  and posterior probabilities  $\{p_i\}_{i \in I}$  with  $p_i \in \Delta(K)$  for every  $i \in I$ , such that there exists a unique probability vector  $\rho \in \Delta(I)$  satisfying

$$\sum_{i \in I} \rho_i p_i = p \quad (4.3a)$$

and

$$\sum_{i \in I} \rho_i \hat{a}(p_i) = \text{cav } \hat{a}(p). \quad (4.3b)$$

Then, it is possible to “split” the total prior probability  $p$  into a set of conditional distributions  $\{p_i\}_{i \in I}$ , such that (i), for every  $i \in I$ , the posterior probabilities that the receiver computes after receiving the recommendation to play  $i$  are  $p_i \in \Delta(K)$ ; and (ii) the sender guarantees an expected payoff equal to  $\text{cav } \hat{a}(p)$ . Condition (4.3a) is called *Bayes plausibility*, while condition (4.3b) is an optimality requirement. Since the distribution  $\rho$  is unique, the receiver will update her prior beliefs from  $p$  to  $p_i$  with probability  $\rho_i$ .

Notice that in the previous result the number of signals required for achieving the optimal value of (4.2) is bounded by the number of types of the sender<sup>8</sup>. Namely, the most the sender needs to transmit to the receiver is just  $k$ , which has  $|K|$  possible values.

<sup>8</sup>This result follows from Carathéodory's theorem.

Given the set  $\{p_i\}_{i \in I}$  and the corresponding distribution of posteriors  $\rho$ , an optimal communication device can be computed from the following formula:

$$\delta_j^k = \begin{cases} \frac{p_j^k \rho_j}{p^k}, & \text{if } j \in I \\ 0, & \text{otherwise} \end{cases}, \quad \forall k \in K. \quad (4.4)$$

Finally, we observe that for any  $p \in \Delta(K)$ , we have that

$$\hat{a}(p) \leq a^*(p) \leq \text{cav } \hat{a}(p). \quad (4.5)$$

The first inequality follows from the fact that  $\hat{a}(p)$  can always be achieved by a NR communication equilibrium. The second inequality is due to the fact that  $\mathcal{D}^*(p) \subseteq \mathcal{D}(p)$ .

We start the analysis of the primal problem in Section 4.2 by dealing only with its optimal value. Then, in Section 4.3 we shall characterize its optimal solutions.

## 4.2 The Virtual Persuasion Game

As we have seen, when there are no informational incentive constraints, the solution to the primal problem can be easily characterized. Informational incentive constraints complicate matters by interconnecting the signals in different states. However, we can integrate the welfare effects of incentive compatibility into the objective function using duality theory. The idea is that the set of communication equilibria is defined by a system of linear inequalities (see Remark 1) for which the dual variables can be used to define the sender's *virtual utility*. These virtual utility scales incorporate into the utility function the signalling costs associated with the incentive compatibility. Using the concept of virtual utility we can transform the original primal problem into a simplified problem without informational incentive constraints but with a different objective function.

Let  $\gamma(k' | k) \geq 0$  be the dual variable (or Lagrange multiplier) for the constraint that type  $k$  of the sender should not gain by reporting  $k'$  in the primal problem for  $p$ . Following Myerson (1991, sec. 10.5), we define the *virtual utility* of the sender from the action  $j$ , when his type is  $k$ , w.r.t. the prior  $p$  and the duals  $\gamma$  to be

$$\alpha_j^k(p, \gamma) = \frac{1}{p^k} \left[ \left( p^k + \sum_{k' \in K} \gamma(k' | k) \right) a_j^k - \sum_{k' \in K} \gamma(k | k') a_j^{k'} \right]. \quad (4.6)$$

In order to understand formula (4.6), we disentangle its components. The terms of the form  $\gamma(k' | k)$  measures the *information rent* that type  $k$  can extract by pretending to be type  $k'$ . On the other hand, the terms of the form  $\gamma(k | k')$  measure the *signaling cost* that type  $k$  must incur in order to reduce the misrepresentation of type  $k'$ . Virtual utility is thus defined as the actual utility plus the total information rents minus the total signaling costs. Notice that multiplying type  $k$ 's utility  $a_j^k$  by the positive constant  $\frac{1}{p^k} (p^k + \sum_{k' \in K} \gamma(k' | k))$  is decision-theoretically inessential. That is, the unique decision-theoretic difference between the real utility scale and the virtual utility scale is given by the signaling costs. Hence, the virtual utility of the sender is a distorted

utility that magnifies the difference between his actual utility and the utility of the types that would be tempted to imitate him.

In a situation where information is not verifiable, so that misrepresentation is possible, some types of the sender may get some information rents from having private information. Also, some types may be compelled to incur in signaling costs in an effort to distinguish themselves from the types that try to mimic them. This new compromise in the payoff maximization goals of the different types of the sender is described by the virtual utility and mathematically measured by the dual variables.

Let us assume that, as a consequence of the pressure that a type might feel in getting the receiver to trust him, the sender begins to act as if he were maximizing his virtual utility (Myerson (1991, sec. 10.8) refers to this idea as the *virtual utility hypothesis*). Thus, for some fixed prior beliefs  $p$  and signaling costs  $\gamma$ , consider the  $(p, \gamma)$ -virtual persuasion problem, a fictitious game that differs from the original persuasion game in the following. First, the sender's types are verifiable by the mediator (but not by the receiver), so that there are no informational incentive constraints. Second, the sender's payoffs are in the virtual utility scales  $(\alpha_j^k(p, \gamma))_{j \in J, k \in K}$  instead of  $(a_j^k)_{j \in J, k \in K}$ .

Let  $\hat{\alpha}(\cdot; p, \gamma)$  denote the non-revealing (virtual) payoff function of the  $(p, \gamma)$ -virtual persuasion problem. As already observed in Section 4.1, the value of persuasion in the  $(p, \gamma)$ -virtual game is given by  $\text{cav } \hat{\alpha}(p; p, \gamma)$ .

Although the  $(p, \gamma)$ -virtual game gives us some insights on how to simplify the sender's problem by removing the informational incentive constraints, it does not say anything about the "dual" relationship between the optimal value of the primal problem for  $p$ ,  $a^*(p)$ , and the value of persuasion in the virtual game. Furthermore, it leaves open the question of determining the optimal signaling costs incurred by the sender in order to distinguish himself from the types that are tempted to imitate him. However, there exists an answer to the first question that will make the second question redundant. The following result is a consequence of strong duality.

**Theorem 1.**

For any prior  $p \in \Delta(K)$  we have that

$$a^*(p) = \min_{\gamma \geq 0} \text{cav } \hat{\alpha}(p; p, \gamma). \tag{4.7}$$

We refer to the minimization problem in (4.7) as the *dual problem for p*.<sup>9</sup>

**Proof.** The Lagrangian of the primal problem for  $p$  is

$$\begin{aligned} \mathcal{L}(\delta, p, \gamma) &= \sum_{k \in K} p^k \sum_{j \in J} \delta_j^k a_j^k + \sum_{k \in K} \sum_{k' \in K} \gamma(k' | k) \left[ \sum_{j \in J} \delta_j^k a_j^k - \sum_{j \in J} \delta_j^{k'} a_j^{k'} \right] \\ &= \sum_{k \in K} p^k \alpha^k(\delta; p, \gamma) \end{aligned}$$

---

<sup>9</sup>We notice that the right-hand side of (4.7) is not, strictly speaking, the dual problem associated to (4.1). However, as shown in the proof of Theorem 1, its optimal value equals the optimal value of the dual.



where  $\gamma \geq 0$  and  $\delta \in \mathcal{D}(p)$ .

Then, the dual problem for  $p$ , associated to the primal problem for  $p$ , is given by

$$\begin{aligned} \min_{\gamma \geq 0} \max_{\delta \in \mathcal{D}(p)} \mathcal{L}(\delta, p, \gamma) &= \min_{\gamma \geq 0} \max_{\delta \in \mathcal{D}(p)} \sum_{k \in K} p^k \alpha^k(\delta; p, \gamma) \\ &= \min_{\gamma \geq 0} \text{cav } \hat{\alpha}(p; p, \gamma) \end{aligned}$$

By strong duality, the value of the primal problem equals the value of its dual<sup>10</sup>. Thus  $a^*(p) = \min_{\alpha \geq 0} \text{cav } \hat{\alpha}(p; p, \gamma)$ .  $\square$

Fix a prior  $p \in \Delta(K)$  and let  $\gamma^*(p)$  be an optimal solution of the dual problem for  $p$ . Then, the key implication of Theorem 1 is that the value of persuasion in the original game coincides with the value of persuasion in the  $(p, \gamma^*(p))$ -virtual game, namely,

$$a^*(p) = \text{cav } \hat{\alpha}(p; p, \gamma^*(p)).$$

Thus, instead of saying that incentive compatibility restricts the sender's ability to signal his information, we may say that he is compelled to modify his actual preferences from the real to the virtual scales  $(p, \gamma^*(p))$ .

**Definition 4 (Value of information).**

The *value of information* for the sender at the prior  $p$  is the difference between the value of persuasion at  $p$  and the non-revealing value at  $p$ , i.e.,  $a^*(p) - \hat{a}(p)$ .

We say that the sender benefits from his private information at  $p$  if the value of information at  $p$  is positive. Theorem 1 provides a necessary and sufficient condition for the sender to benefit from his information.

**Corollary 1.**

The sender benefits from his information at  $p$  if and only if for all  $\gamma \geq 0$ ,

$$\text{cav } \hat{\alpha}(p; p, \gamma) > \hat{a}(p).$$

The motivating example of Section 2 will help us to understand the meaning and significance of the virtual utility. Also, it will provide some interesting conclusions about the effects of informational incentive compatibility.

**Example 1** Consider again the motivating example studied in Section 2. The non-revealing value function  $\hat{a}$  joint with its concavification  $\text{cav } \hat{a}$  are depicted in Figure 3.

We fix the prior probability of type  $H$  to be  $p < \frac{1}{5}$  and we denote  $p_j$  the posterior belief about type  $H$  that the receiver will infer after receiving the recommendation to choose action  $j$ . Under

<sup>10</sup>The following characterization also results from strong duality. Let  $\alpha(\delta; p, \gamma)$  be the ex-ante expected payoff the sender gets in the  $(p, \gamma)$ -virtual game when he uses the communication device  $\delta \in \mathcal{D}(p)$ . Let us define, for every  $p \in \Delta(K)$ , the auxiliary zero-sum game  $G(p)$ , in which the maximizing player chooses  $\delta \in \mathcal{D}(p)$ , the minimizing player chooses  $\gamma \geq 0$  and the payoff is  $\alpha(\delta; p, \gamma)$ . Then the value of  $G(p)$  exists and equals  $a^*(p)$ .

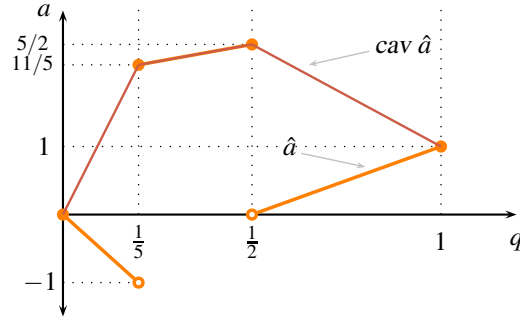


Figure 3: Function  $\hat{a}$  and its concavification

full commitment, (which is equivalent to assume that information is verifiable by the mediator), an optimal communication device must split the total probability  $p$  into the posteriors  $p_3 = 0$  and  $p_2 = \frac{1}{5}$  with probabilities  $\rho_3 = 1 - 5p$  and  $\rho_2 = 5p$ , respectively (see Figure 3). The value of persuasion under the verifiability assumption is

$$\text{cav } \hat{a}(p) = 0\rho_3 + \frac{11}{5}\rho_2 = 11p$$

Thus, the unique optimal communication device is given by

$$\delta_2^H = 1, \quad \delta_2^L = 1 - \delta_3^L = \frac{4p}{1-p}$$

This communication device is, however, not incentive compatible for the sender. This is so because type  $L$  would have incentives to report that he is type  $H$ . As a consequence, the sender cannot achieve the expected payoff  $\text{cav } \hat{a}(p)$  when information is not verifiable, yet he can do better than  $\hat{a}(p)$  as we will see in the sequel.

By solving the dual problem for  $p < \frac{1}{5}$ , we have that the optimal value of the dual variables (Lagrange multipliers) is

$$\gamma^*(H | L) = \frac{10p(1-p)}{3-11p} := \gamma^*(p), \quad \gamma^*(L | H) = 0$$

Because type  $L$  has incentives to lie, it is natural that  $\gamma^*(H | L) > 0$ . Since type  $H$  cannot take any advantage from his private information (lying is not profitable),  $\gamma^*(L | H) = 0$ . The  $(p, \gamma^*(p))$ -virtual utility game can thus be described by the following payoffs matrix:

$\alpha, b$	$j_1$	$j_2$	$j_3$
$H$	$1 + \frac{\gamma^*(p)}{p}, 3$	$3 - \frac{2\gamma^*(p)}{p}, 1$	$-5, -3$
$L$	$-\left(1 + \frac{\gamma^*(p)}{1-p}\right), -3$	$2\left(1 + \frac{\gamma^*(p)}{1-p}\right), -1$	$0, 0$

We notice that type  $L$ 's virtual utility is just a positive multiple of his actual utility. Therefore, both the virtual game and the actual persuasion problem are decision-theoretically equivalent

in state  $L$ . On the other hand, in state  $H$ , the sender's virtual utility magnifies the difference between his true type and the type that would be tempted to imitate him.

Figure 4 illustrates the non-revealing value function  $\hat{\alpha}$  of the virtual game and its concavification. According to Theorem 1, the optimal value of the primal problem for  $p < \frac{1}{5}$  is

$$a^*(p) = \text{cav } \alpha(p; p, \gamma^*(p)) = p + \gamma^*(p) < 11p = \text{cav } \hat{\alpha}(p)$$

We observe that ex-post inefficiencies are incurred in an optimal solution of the primal problem. The optimal value is, however, ex-post efficient in terms of the virtual utility scales. Thus, instead of saying that incentive compatibility forces the sender to incur in (ex-post) inefficiencies, we may say that incentive compatibility compels the sender to behave according to his virtual utilities (Myerson (1991, ch. 10) refers to this idea as the *virtual utility hypothesis*).

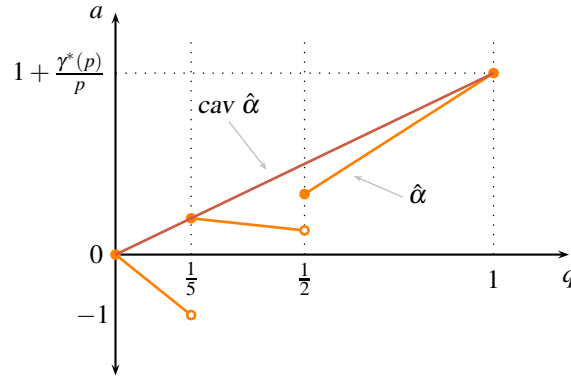


Figure 4: Function  $\hat{\alpha}$  and its concavification

A similar analysis for the case  $p > \frac{1}{2}$  shows that optimal value of the primal problem is

$$a^*(p) = \text{cav } \alpha(p; p, \gamma^{**}(p)) = p + 6\gamma^{**}(p) < 4 - 3p = \text{cav } \hat{\alpha}(p)$$

where,  $\gamma^{**}(p) := \gamma^{**}(L | H) = \frac{2p(p-1)}{5p-1}$  and  $\gamma^{**}(H | L) = 0$  are the optimal solutions of the dual problem for  $p$ .

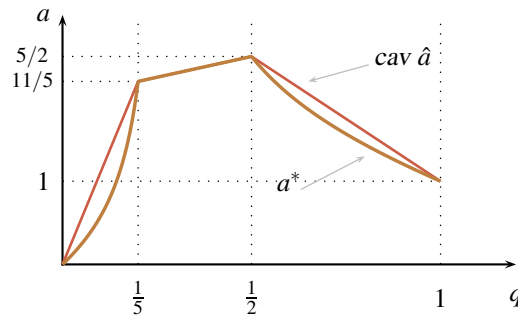


Figure 5: Functions  $a^*$  and  $\text{cav } \hat{\alpha}$  for Example 1

Whenever  $p \in [1/5, 1/2]$ , the receiver's optimal action is to choose  $j_2$ , which is the preferred action for both types of the sender. Thus,  $a^*(p) = \hat{\alpha}(p) = \text{cav } \hat{\alpha}(p)$ . The value function  $a^*$  looks like in Figure 5.

**Example 2** As a further illustration of our methodology, we study the following game, proposed by Forges (1990). It has been extensively analyzed in the literature of strategic information transmission. Payoffs for both players depend on the sender’s type and the receiver’s action as follows:

$a, b$	$j_1$	$j_2$	$j_0$	$j_3$	$j_4$
$H$	3,0	4,4	0,7	10,9	6,10
$L$	6,10	10,9	0,7	4,4	3,0

This example has a natural interpretation in terms of a job assignment scenario. An employer must decide whether to hire a candidate and, if so, to assign the employee to one of four possible jobs. The candidate may be one of two types. Type  $L$  performs better in job 1 but prefers job 2; he is bad at job 3, and even worse at job 4. Type  $H$  is similar but with jobs reversed. Sender’s prior probability of type  $H$  is  $p \in (0, 1)$ .

By performing a similar analysis as in Example 1, it can be shown that the value of persuasion in this game (depicted in Figure 6) is given by

$$a^*(p) = \begin{cases} 6 + 14p + \frac{28p(5p-1)}{4-15p}, & \text{if } p < \frac{1}{5} \\ \frac{44}{5}, & \text{if } \frac{1}{5} \leq p \leq \frac{4}{5} \\ 20 - 14p - \frac{28(1-p)(5p-4)}{15p-11}, & \text{if } p > \frac{4}{5} \end{cases}$$

In particular, we have that for any  $p < 1/5$  (resp.  $p > 4/5$ ) only type  $L$  (resp.  $H$ ) has incentives to lie, so that  $\gamma(H | L) > 0$  (resp.  $\gamma(L | H) > 0$ ).

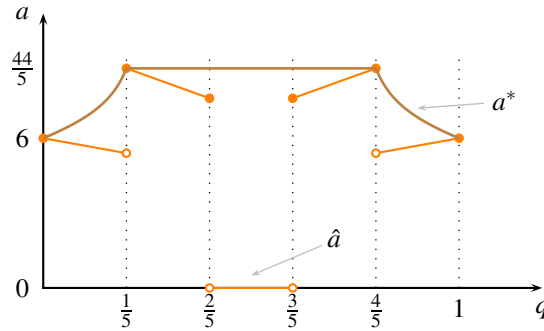


Figure 6: Functions  $a^*$  and  $\hat{a}$  for example 2

The nature of this game is similar to that of Example 1, except that here there is an outside option: not to hire the candidate, i.e., action  $j_0$ .

### 4.3 Optimal Mediators

So far we have focused on the optimal value of the primal problem. Our aim now is to characterize its optimal solutions. For that, let us start considering the case in which the optimal value of the dual variables is zero at some prior  $p$ , so that incentive constraints are not essential. In such a situation virtual utilities coincide with real utilities and the value of persuasion at  $p$  is  $\text{cav } \hat{a}(p)$ . Then, according to (4.3b), the posterior beliefs  $\{p_i\}_{i \in I}$  induced by any optimal communication

device correspond to the points on the domain of  $\hat{a}$  for which the convex combination of their images yields  $\text{cav } \hat{a}(p)$ . The corresponding distribution of posteriors  $\rho = (\rho_i)_{i \in I}$  is the (unique) solution of a well determined system of linear equations given by (4.3a). Hence, given  $\{p_i\}_{i \in I}$  and  $\rho$ , an optimal communication device can be easily computed using formula (4.4).

Now consider the situation in which there are binding informational incentive constraints. On one hand, according to the virtual utility hypothesis, the sender might distort his preferences from the actual to the virtual scales, exaggerating the difference from the types that try to mimic him. On the other hand, as we have shown in the motivating example (Example 1), the sender may require to transmit more messages than his number of types. Then, the number of unknowns in (4.3a) (messages) increases while the number of equations (types) remains the same. The system in (4.3a) may become underdetermined and therefore infinitely many distributions of posteriors may be consistent with the same prior probability. Thus, additional conditions are required to characterize any optimal communication device.

To understand this issue, let us return to Example 1. Let  $p < \frac{1}{5}$  and consider the  $(p, \gamma^*(p))$ -virtual game. According to Figure 4 and condition (4.3b), the optimal value of the primal problem can be achieved by splitting the total probability  $p$  in either of the following collection of posteriors:

(i)  $p_3 = 0, p_2 = \frac{1}{5}$ , or

(ii)  $p_3 = 0, p_1 = 1$ , or

(iii)  $p_3 = 0, p_2 = \frac{1}{5}, p_1 = 1$ .

In case (i), Bayes plausibility implies that  $(\rho_2, \rho_3) = (5p, 1 - 5p)$  and thus formula (4.4) yields  $\delta_2^H = 1 - \delta_3^H = 1$  and  $\delta_2^L = 1 - \delta_3^L = \frac{4p}{1-p}$ . But this communication device is not incentive compatible for the sender. In case (ii) we have that Bayes plausibility implies that  $(\rho_1, \rho_3) = (p, 1 - p)$  and therefore formula (4.4) yields  $\delta_1^H = \delta_3^L = 1$ . This communication device is a FR communication equilibrium giving an expected payoff to the sender equal to  $p$  which is strictly lower than  $a^*(p)$ . Finally, in case (iii), Bayes plausibility does not uniquely identify a distribution of posteriors. In particular, any probability vector  $(\rho_1, \rho_2, \rho_3)$  satisfying  $\rho_1 + \frac{\rho_2}{5} = p$  is a feasible distribution of posteriors. An additional condition is thus required in order to identify the correct distribution of posteriors.

Duality theory implies a relationship between the primal and dual problems that is known as *complementary slackness*. Specifically, it says that if a dual variable is positive, then the associated informational incentive constraint must be binding. Conversely, if a constraint fails to bind, then the associated dual variable must be zero. Complementary slackness provides us the additional equations we needed. Consider again Example 1 with  $p < \frac{1}{5}$ . As we have already shown, the optimal value of  $\gamma^*(H | L)$  is strictly positive. Then, according to the complementary slackness, the constraint asserting that the type  $L$  should not gain by reporting  $H$  is binding, i.e.,

$$2\delta_2^L - \delta_1^L = 2\delta_2^H - \delta_1^H$$

The previous equality joint with formula (4.4) yield the additional restriction

$$5\rho_1(1 - p) = 2\rho_2(1 - 5p)$$

This equality together with Bayes plausibility implies that the optimal distribution of posteriors is

$$\rho_1 = \frac{2p(1-5p)}{3-11p}, \quad \rho_2 = \frac{5p(1-p)}{3-11p}, \quad \text{and} \quad \rho_3 = 1 - \rho_1 - \rho_2$$

Given these posteriors, formula (4.4) gives the optimal communication device solving the primal problem for  $p$ . We conclude that, in order to achieve the optimal value  $a^*(p)$ , the sender requires to induce a split of the total prior probability  $p$  into posterior beliefs  $p_3 = 0$ ,  $p_2 = \frac{1}{5}$  and  $p_1 = 1$ . Hence, an optimal communication device transmits 3 ( $> |K| = 2$ ) different messages (recommendations) with positive probability. The sender transmits as much information as when misrepresentation is not problematic (i.e., he induces posteriors  $p_3 = 0$  and  $p_2 = \frac{1}{5}$ ), however he also requires to send the message  $j_1$  to make such revelation credible to the receiver.

Drawing on the virtual utility hypothesis, the increased number of messages can be justified by the fact that incentive compatibility obliges type  $H$  to use an additional costly signal, from which he gains positive virtual utility. The optimal probability of sending message  $j_1$  (i.e.,  $\rho_1$ ) is determined by the minimization of the signaling costs incurred in recommending  $j_1$ .

Optimality conditions from strong duality theory imply the following result:

**Proposition 1.**

Let  $\delta$  be a communication device satisfying the informational incentive constraints for the sender. Then,  $\delta$  is an optimal solution of the primal problem for  $p$  if and only if there exists a vector  $\gamma \geq 0$  such that

$$\gamma(k' | k) \left[ \sum_{j \in J} (\delta_j^k - \delta_j^{k'}) \alpha_j^k \right] = 0, \quad \forall k, k' \in K \quad (4.8)$$

and

$$\sum_{k \in K} p^k \alpha^k(\delta; p, \gamma) = \text{cav } \hat{\alpha}(p; p, \gamma) \quad (4.9)$$

Condition (4.9) is the counterpart of conditions (4.3a) and (4.3b). It says that the optimal communication device induces a distribution of posterior beliefs giving the sender an ex-ante expected virtual payoff equal to the concavification of the non-revealing virtual payoff function  $\alpha(\cdot; p, \gamma)$  evaluated at the prior distribution  $p$ . Condition (4.8) is the complementary slackness.

Following Myerson (1991, sec. 10.5), we say that a type  $k$  *jeopardizes* another type  $k'$  at the prior  $p$  if the optimal value of the dual variable  $\gamma(k' | k)$  at  $p$  is positive.

#### 4.4 Extreme Communication Equilibria and the Number of Signals

Our aim in this section is to provide an upper-bound on the number of recommended actions required to achieve the value of persuasion. For that, we exploit the geometric properties of the set of communication equilibria.

Recall that, for any fixed  $p \in \Delta(K)$ , the feasible set of the primal problem for  $p$  is a convex polytope (bounded polyhedron). Then, the sender's expected payoff achieves its maximum at an extreme point of  $\mathcal{D}^*(p)$  (or a convex combination of them).

**Definition 5 (Extreme communication equilibrium).**

The communication device  $\delta$  is an *extreme communication equilibrium* of  $\Gamma(p)$  if it is an extreme point of  $\mathcal{D}^*(p)$ .

Using a basic result from the theory of linear programming, it is possible to characterize the number of messages sent with positive probability in any extreme communication equilibrium. A solution of a system of linear inequalities is an extreme point of the corresponding feasible set if and only if it can be obtained as the unique solution to a system of equations derived from equality constraints by setting a subset of variables equal to zero (see for instance Schrijver (1998)). These are called *basic feasible solutions*. As a consequence, the number of non-zero components in any extreme point is no greater than the number of binding constraints. Thus, a way to identify an upper-bound on the number of actions with positive probability in an extreme communication equilibrium is to determine how many incentive constraints can be binding.

The previous insight was applied by Forges (1994) to show that whenever the sender has only two types (i.e.,  $|K| = 2$ ), the number of recommended actions in an extreme communication equilibrium cannot exceed 4. This bound corresponds to the number of types ( $|K| = 2$ ) plus the number of informational incentive constraints. Unfortunately, the reasoning in the proof of this result relies strongly on the fact that  $|K| = 2$ . However, a similar statement can be proved for the general case  $|K| \geq 2$  by modifying the sender's problem<sup>11</sup>. The idea is as follows: let  $\bar{\delta}$  denote a solution of the primal problem for  $p$ . Now, replace each  $\delta_j^k$  in the definition of the primal for  $p$  by the variable  $\theta_j \bar{\delta}_j^k$ , with  $\theta_j \geq 0$  and add  $|K|$  constraints of the form  $\sum_j \theta_j \bar{\delta}_j^k = 1$  for all  $k \in K$ . By keeping fixed  $\bar{\delta}$  and  $p$ , we obtain a linear programming problem on  $\theta$ . For this problem, the strategic incentive constraints are redundant, thus we end up with  $|K|^2 (= |K| + |K|(|K| - 1))$  constraints. Then, applying the previous insights, there exist a solution of the modified problem, denoted  $\tilde{\theta}$ , with at most  $|K|^2$  positive components. By construction, the communication device  $\tilde{\delta}$  defined by  $\tilde{\theta}_j \bar{\delta}_j^k$  is also an optimal solution of the primal problem for  $p$ . Since, all actions  $j$  for which  $\theta_j = 0$  have zero probability in  $\tilde{\delta}$ , we are able to find an upper bound on the number of signals.

**Proposition 2.**

For any  $p \in \Delta(K)$ , there exists a solution of the primal problem for  $p$  for which the number of actions with positive probability does not exceed  $|K|^2$ .<sup>12</sup>

**Proof.** Let  $p \in \Delta(K)$  and assume that only  $m (\leq |K|(|K| - 1))$  informational incentive constraints are linearly independent in the primal problem for  $p$ . Then, there exist a set  $M \subseteq K \times K$

<sup>11</sup>The same method is also applied by Bester and Strausz (2007).

<sup>12</sup>As it is inferred from the proof, we can also establish a result somewhat stronger than Proposition 2. Suppose that it is possible to establish that at most  $m$  informational incentive constraints are binding, for instance by showing that some of them can be written as linear combinations of the others. Then, there is a solution of the primal problem for which the number of actions with positive probability does not exceed  $|K| + m$ .

such that  $|M| = m$  and any solution of the programming problem

$$\begin{aligned}
& \max_{\delta \geq 0} \sum_{k \in K} p^k \sum_{j \in J} \delta_j^k a_j^k & (4.10) \\
& \text{s.t.} \sum_{j \in J} \delta_j^k a_j^k \geq \sum_{j \in J} \delta_j^{k'} a_j^k, \quad \forall (k, k') \in M \\
& \sum_{k \in K} \delta_j^k p^k b_j^k \geq \sum_{k \in K} \delta_j^k p^k b_i^k, \quad \forall i, j \in J \\
& \sum_{j \in J} \delta_j^k = 1, \quad \forall k \in K
\end{aligned}$$

is a solution of the primal problem for  $p$ , and viceversa. We need to prove that there exists a solution of the primal problem for  $p$  for which no more than  $|K| + m$  actions are recommended with positive probability. Let  $\bar{\delta}$  be a solution of (4.10). Let  $N = \{j \in J \mid \pi_j(\bar{\delta}) > 0\}$  and  $n = |N|$ . If  $n \leq |K| + m$ , there is nothing to prove. Then, assume that  $n > |K| + m$ . Consider the linear programming problem

$$\begin{aligned}
& \max_{\theta \geq 0} \sum_{k \in K} p^k \sum_{j \in J} \theta_j \bar{\delta}_j^k a_j^k & (4.11) \\
& \text{s.t.} \sum_j \theta_j \bar{\delta}_j^k a_j^k \geq \sum_j \theta_j \bar{\delta}_j^{k'} a_j^k, \quad \forall (k, k') \in M \\
& \sum_{j \in J} \theta_j \bar{\delta}_j^k = 1, \quad \forall k \in K
\end{aligned}$$

Because  $\bar{\delta}$  is a solution of (4.10), the vector  $\bar{\theta} \geq 0$  defined by  $\bar{\theta}_j = 1$  for all  $j \in N$  and  $\bar{\theta}_j = 0$  for all  $j \in J \setminus N$  solves the linear program (4.11). By a fundamental result of linear programming (see Schrijver (1998)), we can always find a solution of (4.11) among the extreme points of its feasible set. Therefore, since (4.11) has  $|K| + m$  constraints, it has a basic feasible solution  $\tilde{\theta} \geq 0$  with no more than  $|K| + m$  strictly positive components. For every  $j \in J$  and  $k \in K$ , we define  $\tilde{\delta}_j^k = \tilde{\theta}_j \bar{\delta}_j^k \geq 0$ . Let  $\tilde{N} = \{j \in J \mid \pi_j(\tilde{\delta}) > 0\}$ . Then,  $|\tilde{N}| \leq |K| + m$ ,

$$\sum_{j \in J} \tilde{\delta}_j^k = \sum_{j \in J} \tilde{\theta}_j \bar{\delta}_j^k = 1, \quad \forall k \in K$$

$$a(\tilde{\delta}; p) = \sum_{k \in K} p^k \sum_{j \in J} \tilde{\theta}_j \bar{\delta}_j^k a_j^k = \sum_{k \in K} p^k \sum_{j \in J} \tilde{\delta}_j^k a_j^k = a(\bar{\delta}; p)$$

and

$$\sum_{k \in K} \tilde{\delta}_j^k p^k b_j^k = \tilde{\theta}_j \sum_{k \in K} \bar{\delta}_j^k p^k b_j^k \geq \tilde{\theta}_j \sum_{k \in K} \bar{\delta}_j^k p^k b_i^k = \sum_{k \in K} \tilde{\delta}_j^k p^k b_i^k, \quad \forall i, j \in J$$

Then,  $\tilde{\delta}$  is a solution of (4.10) for which no more than  $|K| + m$  actions are recommended with positive probability.  $\square$

We now provide an example that shows that the bound in Proposition 2 is actually tight. This means that without any further knowledge on the number of binding informational incentive constraints, the lowest possible upper bound on the number of recommended actions is  $|K|^2$ .



**Example 3** Payoffs for both players depend on the sender's type and the receiver's action as indicated in the following matrix:

$a, b$	$j_1$	$j_2$	$j_3$	$j_4$	$j_5$
$H$	-2,0	2,4	1,7	$\frac{3}{2}, 9$	0,10
$L$	0,10	$\frac{3}{2}, 9$	1,7	2,4	-2,0

Let  $p = 1/2$  be the prior probability of type  $H$ . Then, the optimal solution of the dual problem for  $p$  is  $\gamma(H | L) = \gamma(L | H) = \frac{17}{42}$ . Complementary slackness implies that both informational incentive constraints are binding. Therefore, we expect an optimal solution of the primal problem to involve 4 messages. Indeed, the unique optimal solution is

$$\delta_2^H = \delta_4^L = \frac{8}{21}, \quad \delta_4^L = \delta_2^L = \frac{4}{7}, \quad \delta_5^H = \delta_1^L = \frac{1}{21},$$

which induces posterior probabilities  $p_1(\delta) = 0$ ,  $p_2(\delta) = \frac{2}{5}$ ,  $p_4(\delta) = \frac{3}{5}$ ,  $p_5(\delta) = 1$ . Thus, actions  $j_1$ ,  $j_2$ ,  $j_4$  and  $j_5$  are recommended with positive probability. Qualitatively similar results are obtained for any prior probability  $p \in (\frac{2}{5}, \frac{3}{5})$ .

## 5 Discussions

### 5.1 Cheap-Talk Implementation

In some environments, plain conversation between the players is more natural than mediated communication. Is it possible to achieve any communication equilibrium payoff by means of cheap-talk? Forges (1990) shows that there may exist communication equilibrium payoffs that cannot be implemented as *Nash* equilibrium payoffs of any long cheap-talk extension of  $\Gamma(p)$ . However, communication equilibrium payoffs can be implemented as *correlated* equilibrium payoffs (in the sense of Aumann (1974)). More precisely, a *feasibility theorem* holds: the set of all correlated equilibrium payoffs of all cheap-talk extensions of  $\Gamma(p)$  coincides with the set of all communication equilibrium payoffs. In fact, any communication equilibrium payoff can be achieved as a correlated equilibrium payoff of a cheap-talk extension of  $\Gamma(p)$  with only one stage of information transmission (see Forges (1985)).

### 5.2 Information Design Problems

The fundamental result we use to make the sender's problem more tractable is the revelation principle for general Bayesian games. This principle states that without loss of generality, the sender may restrict attention to communication equilibria, which are described by a set of linear inequalities. Therefore the sender's problem can be formulated as a linear programming problem. One advantage of this approach is that it does not depend on the number of receivers or the fact that they are uninformed. Indeed, the revelation principle also applies for a persuasion problem with an arbitrary number of privately informed receivers. In this general setting, the mediator (communication device) first asks all informed players (sender and receivers) to simultaneously and confidentially reveal their individual types. Then he privately recommends

an action to each receiver. In a communication equilibrium of this game, all informed players always report their types truthfully and the receivers always follow the prescribed recommendation. Here again, the set of communication equilibria is a convex polyhedron, thus our analytical framework readily extends to more general information design problems.

Unfortunately, our main results concerning the dual properties of the primal problem (Theorem 1 and Proposition 1) do not extend to general information design problems. The main difficulty comes from the fact that, because of the strategic externalities, each receiver's optimal action depends not only on the posterior beliefs she infers after receiving a recommendation, but also on the *non observed* recommendations made to the other receivers.

Our approach is reminiscent of a recent methodology developed by Taneva (2016) for the study of information design problems. She considers a basic game in which: (i) a set of multiple receivers have symmetric uncertainty about an unobserved payoff-relevant state with a commonly known prior distribution; and (ii), an information designer has preferences that depend on the state and the actions taken by the agents. The designer, before observing the realization of the state, commits to an information structure (i.e., a set of signals together with a signaling strategy). His problem then is to find an information structure which, for the given basic game, supports a Bayesian equilibrium that maximizes his expected payoff. Using the concept of *Bayes correlated equilibrium*, introduced by Bergemann and Morris (2016), Taneva characterizes the set of all Bayesian equilibria associated with all possible information structures for a given basic game. By doing so, she equivalently reformulates the designer's problem as a linear programming problem. The notion of Bayes correlated equilibrium considers only the strategic incentive constraints related to the "obedience" of the receivers. For that reason, Taneva's approach assumes full commitment on the part of the designer. In contrast, the concept of communication equilibrium captures the idea that players can strategically manipulate their information by imposing additional truth-telling incentive constraints. As a consequence, Taneva's formulation can be seen as a particular case of our analytical framework, one in which the receivers are uninformed (symmetric uncertainty) and the private information of the designer is verifiable by the mediator.



# Bibliography

1. AUMANN, R. (1960). "Linearity of unrestrictedly transferable utilities". *Naval Research Logistics Quarterly*, 7, pp. 281-284.
2. AUMANN, R. (1974). "Subjectivity and correlation in randomized strategies". *Journal of Mathematical Economics*, 1, pp. 67-96.
3. AUMANN, R. (1994). Economic applications of the Shapley value. In: *Game Theoretic Methods in General Equilibrium Analysis*. J.-F. Mertens and S. Sorin (Eds.). Kluwer Academic Publishers, Dordrecht, pp. 121-133.
4. AUMANN, R. AND MASCHLER, M. (1995). *Repeated Games With Incomplete Information*, MIT Press.
5. BERGEMANN, D. AND MORRIS, S. (2016). "Bayes correlated equilibrium and the comparison of information structures in games". *Theoretical Economics*, 11, pp. 487-522.
6. BESTER, H. AND STRAUZ, R. (2001). "Contracting with imperfect commitment and the revelation principle: The single agent case". *Econometrica*, 69, pp. 1077-1098.
7. BESTER, H. AND STRAUZ, R. (2007). "Contracting with imperfect commitment and noisy communication". *Journal of Economic Theory*, 136, pp. 236-259.
8. CRAWFORD, V. AND SOBEL, J. (1982). "Strategic information transmission". *Econometrica*, 50, pp. 1431-1451.
9. D'ASPREMONT, C. AND GÉRARD-VARET, L.-A. (1979). "Incentives and incomplete information". *Journal of Public Economics*, 11, pp. 25-54.
10. D'ASPREMONT, C. AND GÉRARD-VARET, L.-A. (1982). "Bayesian incentive compatible beliefs". *Journal of Mathematical Economics*, 10, pp. 83-103.
11. DE CLIPPEL, G. (2005). "Values for cooperative games with incomplete information: An eloquent example". *Games and Economic Behavior*, 53 pp. 73-82.
12. DE CLIPPEL, G. (2008). "Equity, envy and efficiency under asymmetric information". *Economic Letters*, 99, pp. 265-267.
13. DE CLIPPEL, G. (2012). "Egalitarianism in mechanism design". Working paper, Department of Economics, Brown University.

14. DE CLIPPEL, G. AND MINELLI, E. (2004). "Two-person bargaining with verifiable information". *Journal of Mathematical Economics*, 40, pp. 799-813.
15. DE CLIPPEL, G., PÉREZ-CASTRILLO, D. AND WETTSTEIN, D. (2012). "Egalitarian equivalence under asymmetric information". *Game and Economic Behavior*, 75, pp. 413-423.
16. FARRELL, J. (1993). "Meaning and credibility in cheap-talk games". *Games and Economic Behavior*, 5, pp. 514-531.
17. FORGES, F. (1985). "Correlated equilibria in a class of repeated games with incomplete information". *International Journal of Game Theory*, 14, pp. 129-150.
18. FORGES, F. (1986). "An approach to communication equilibria". *Econometrica*, 54, pp. 1375-1385.
19. FORGES, F. (1990). "Equilibria with communication in a job market example". *The Quarterly Journal of Economics*, 105, pp. 375-398.
20. FORGES, F. (1994). Non-zero sum repeated games and information transmission. In *Essays in Game Theory in Honor of Michael Maschler*, 65-95. (Ed.) Nimrod Megiddo, Springer-Verlag, New York.
21. FORGES, F., MERTENS, J.-F. AND VOHRA, R. (2002). "The ex-ante incentive compatible core in the absence of wealth effects". *Econometrica*, 70 pp. 1865-1892.
22. FORGES, F., MINELLI, E. AND VOHRA, R. (2002). "Incentives and the core of an exchange economy: a survey". *Journal of Mathematical Economics*, 38 pp. 1-41.
23. FORGES, F. AND SERRANO, R. (2013). "Cooperative games with incomplete information: Some open problems". *International Game Theory Review*, 15 pp. 298-302.
24. GOLTSMAN, M., HÖRNER, J., PAVLOV, G. AND SQUINTANI, F. (2009). "Mediation, arbitration and negotiation". *Journal of Economic Theory*, 144, pp. 1397-1420.
25. HARSANYI, J. (1963). "A simplified bargaining model for the  $n$ -person cooperative game". *International Economic Review*, 4, pp. 194-220.
26. HARSANYI, J. (1967-8). "Games with incomplete information played by Bayesian players". *management Sciences*, 14, pp. 159-189, 320-334, 486-502.
27. HART, S. (1985a). "Nontransferable utility games and markets: Some examples and the Harsanyi solution". *Econometrica*, 53, pp 1445-1450.
28. HART, S. (1985b). "An axiomatization of Harsanyi's nontransferable utility solution". *Econometrica*, 53, pp 1295-1313.
29. HART, S. (2004). "A comparison of non-transferable utility values". *Theory and Decisions*, 56, pp 35-46.

30. HART, S. AND MAS-COLELL, A. (1996). "Bargaining and the value". *Econometrica*, 64, pp. 357-380.
31. HOLMSTRÖM, B. AND MYERSON, R. (1983). "Efficient and durable decision rules with incomplete information". *Econometrica*, 51, pp 1779-1819.
32. IMAI, H. (1983). "On Harsanyi's solution". *International Journal of Game Theory*, 12, pp. 161-179.
33. KALAI, E., SAMET, D. (1985). "Monotonic solutions to general cooperative games". *Econometrica*, 53, pp. 307-327.
34. KALAI, A. AND KALAI, E. (2013). "Cooperation in strategic games revisited". *The Quarterly Journal of Economics*, 128, pp. 917-966.
35. KAMENICA, E. AND GENTZKOW, M. (2011). "Bayesian persuasion". *American Economic Review*, 101, pp. 2590-2615.
36. KOLOTILIN, A., LI, M., MYLOVANOV, T. AND ZAPECHELNYUK, A. (2015). Persuasion of a privately informed receiver. Working paper.
37. MITUSCH, K. AND STRAUZ, R. (2005). "Mediation in situations of conflict and limited commitment". *The Journal of Law, Economics and Organization*, 21, pp. 467-500.
38. MYERSON, R. (1979). "Incentive compatibility and the bargaining problem". *Econometrica*, 47, pp. 61-73.
39. MYERSON, R. (1980). "Conference structures and fair allocation rules". *International Journal of Game Theory*, 9, pp. 169-182.
40. MYERSON, R. (1981). "Optimal auction design". *Mathematics of Operations Research*, 6, pp. 58-73.
41. MYERSON, R. (1983). "Mechanism design by an informed principal". *Econometrica*, 51, pp. 1767-1797.
42. MYERSON, R. (1984a). "Two-person bargaining problems with incomplete information". *Econometrica*, 52, pp. 461-488.
43. MYERSON, R. (1984b). "Cooperative games with incomplete information". *International Journal of Game Theory*, 13, pp. 69-96.
44. MYERSON, R. (1985). Analysis of two bargaining problems with incomplete information. In: *Game Theoretic Models of Bargaining*. A.E. Roth (Ed.). Cambridge University Press, Cambridge, pp. 115-147.
45. MYERSON, R. (1986). "Multi-stage games with communication". *Econometrica*, 54, pp. 323-358.

46. MYERSON, R. (1991). *Game Theory: Analysis of Conflict*, Harvard University Press.
47. MYERSON, R. (1992). Fictitious-transfers solutions in cooperative game theory. In: *Rational Interaction*. R. Selten (Ed.). New York: Springer-Verlag, pp. 13-33.
48. MYERSON, R. (2007). "Virtual utility and the core for games with incomplete information". *Journal of Economic Theory*, 136, pp. 260-285.
49. MYERSON, R. AND SATTERTHWAITE, M. (1983). "Efficient mechanisms for bilateral trading". *Journal of Economic Theory*, 29, pp. 265-281.
50. NASH, J. (1950). "The bargaining problem". *Econometrica*, 18, pp. 155-162.
51. NASH, J. (1953). "Two-person cooperative games". *Econometrica*, 21, pp. 128-140.
52. NEHRING, K. (2004). "The veil of public ignorance". *Journal of Economic Theory*, 119, pp. 247-270.
53. OWEN, G. (1972). "A value for games without side payments". *International Journal of Game Theory*, 1, pp. 95-109.
54. PELEG, B AND SUDHÖLTER, P. (2007). *Introduction to the Theory of Cooperative Games*, Second Edition. Springer-Verlag Berlin Heidelberg.
55. PRESCOTT, E. AND TOWNSEND, R. (1984). "General competitive analysis in an economy with private information". *International Economic Review*, 25, pp. 1-20.
56. ROTH, A. (1980). "Values for games without side-payments: Some difficulties with current concepts". *Econometrica*, 48, pp. 457-465.
57. SALAMANCA, A. (2016). A generalization of the Harsanyi NTU value to games with incomplete information. HAL Working paper 01579898.
58. SALAMANCA, A. (2017). A comparison of NTU values in a cooperative game with incomplete information. HAL Working paper 01500966.
59. SCHRIJVER, A. (1998). *Theory of Linear and Integer Programming*, John Wiley and sons.
60. SHAPLEY, L. (1953). A Value for  $n$ -person Games. In: *Contributions to the Theory of Games II*. H. W. Kuhn and A. W. Tucker (Eds.). Princeton: Princeton University Press, pp. 307-317.
61. SHAPLEY, L. (1969). Utility comparisons and the theory of games. In: *La Decision*. Paris: Editions du CNRS, pp. 251-263.
62. TANEVA, I. (2016). "Information design". Working paper available at <https://sites.google.com/site/itaneva13/research>.