Essays in Game Theory

A thesis presented

by

Giuseppe G. L. Cappelletti

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Bocconi University

Milano, italy

January, 2005

Contents

Preface

Voglio ringraziare mia madre, Barbara, Federico per l'incondizionata fiducia e per il continuo supporto che mi hanno offerto in tutti questi anni. Professor Pierpaolo Battigalli, Professor Sandro Brusco e Professor Marco Mariotti per i preziosi consigli, la voglia e la disponibilità che mi hanno sempre dato.

Introduction

These thesis contains chapters which deal with very different issues. The first two chapters study the existence if a so-called universal type space, usinge a purely logical perspective. From the Harsanyi's seminal contribution the analysis of game with incomplete information relies on the so-called Universal Type Space. The first chapter investigates whether a Universal Type Space exists and provides a general framework which allows to rejoin previous results. We take a purely linguistic perspective in order to determine sufficient conditions for the Universal Type Space to exist and we characterize some of its properties. The construction of the Universal Type Space has been pursued by using two different approaches, Semantic and Syntactic. The second chapter investigates whether Semantic and Syntactic approach are equivalent. We take a purely linguistic perspective where Semantic and Syntactic approach are two possible interpretation of a formal language namely, first order logic which is able to fully describe players' epystemic characteristics.

The following two chapters relying on recent result on the implication of rationality and common knowledge of rationality investigates when information disclosure occurs. The third chapter analyzes a model in which a informed agent sends a cheap-talk message to an uniformed party, subsequently, takes an action that determines the utility of both. We assume full rationality, a certain degree of alignment of interest and that agent have a propensity to believe others. Pursuing a fully fledged non-equilibrium analysis we point out sufficient and tight conditions for full disclosure to occur. We, also, apply our framework to previous results in the literature. The forth chapter is a short note on the characterization of $\Delta-$*rationalizability* using the notion of iterated dominance. The concept of rationalizability, which has introduced by Bernheim and Pearce, has been widely accepted in the study of normal form games. There are a situation where is natural to introduce assumptions on players' belief. In this settings it is natural to consider the implications of rationality, common certainty of rationality and common certainty of some restrictions on players beliefs. This leads to an extension of the rationalizability solution concept to static and dynamic games of incomplete information, which takes as given some exogenous restrictions on players' beliefs. In this chapter we want to characterize the set of rationalizable actions in terms of dominance ralations.

The last chapter reviews recent results on aggregation of preferences when individuals face uncertainty.

Part I

Foundation of Game Theory

Chapter 1

On the Existence of a Universal Type Space

On the Existence of a Universal Type Space

Giuseppe G. L. Cappelletti
Bocconi University

January 2^{nd}, 2005

Abstract

From the Harsanyi's seminal contribution the analysis of game with incomplete information relies on the so-called Universal Type Space. This essay investigates whether a Universal Type Space exists and provides a general framework which allows to rejoin previous results. We take a purely linguistic perspective in order to determine sufficient conditions for the Universal Type Space to exist and we characterize some of its properties.

1.1 Introduction

This paper investigates whether a so-called universal type space exists. The perspective pursued is quite original but it will prove to be very insightful and fruitful. Namely, we will follow a purely linguistic point of view.

Harsanyi's (1967-68) proposed a new theory for the analysis of games with incomplete information, i.e. games where the relationship between actions and payoffs is not commonly known by the agents. Naturally, the analysis of this kind of games gives rise to an infinite regress in reciprocal expectation. For example, a player's strategy choice will depend on what he expects to be his opponents' preferences. This expectation is called *first-order* expectation. But his choice will also depend on what he expects to be his opponents' expectation about his own preferences. This is called *second-order* expectation. Indeed, player's strategy choice will also depend on what he expects to be his opponents' *second-order* expectation.[1] This is called *third-order* expectation and we can proceed *ad infinitum.*[2]
Harsanyi encodes both player's preferences and his first-order and higher-order beliefs concerning opponents in the so called player's type.[3] Given any game, the set of all possible types of each player is defined the universal type space. The universal type space is the basis for a more manageable framework to investigate interactive interaction with incomplete information. Moreover, the existence the a universal type space justifies the use of a Bayesian approach to model these situations (see Tan and Werlag (1988) [54]).

The existence of a universal type space has been proved in several papers under different assumptions (see Mertens and Zamir (1985) [39], Brandenburger and Dekel (1987)[9] and Heifetz and Samet (1998) [28] among others). If there is not such space we should stick to the explicit description of agents' hierarchical beliefs. But, also some negative results appeared (refer to Brandenburger and Keisler (2003), Heifetz and Samet (1998) [29]). The main issues raised were twofold. On the one hand there are measure theoretic problems usually studied using the semantic approach (refer to Heifetz and Samet (1999) [29]). On the other hand there are purely logical ones. In this category, we mention Branden-

[1] That is, what the player thinks that his opponents think that he thinks about their own preferences.

[2] In the following section it will become clear how tricky and essential is the expression "*ad infinitum*" in the construction of a universal type space.

[3] In the original paper Harsanyi used the expression "*information vector*". In order to distinguish payoff or preference related aspects (also called payoff-type) from beliefs related ones (also called epistemic-type) the plain expression "Harsanyi-type" has been used.

burger and Keisler (2003), who proved an impossibility theorem for a specific type space, namely a complete possibility structure.

One can doubt whether a definitive or comprehensive answer on the existence of a universal type space can be given. Our attempt is to give at least a partial answer to this question. Although, we will take a long route in order to do that.

All the previous results started from two different[4] perspectives: a purely set theoretic approach, also called semantic approach, and a purely logical or linguistic approach, also called syntactic approach.
In the so-called semantic approach, the epistemic characteristics of the players are represented by a structure consisting of a set Ω of states of the world, together with a partition Π_i of Ω, for each player i. The cell of Π_i containing an element ω of the set Ω, represents player i knowledge in state ω. This approach is purely set theoretic.
The syntactic approach hinges on a purely logic construction. This construction is constituted by a set of propositions built with a formal language, which is expressive enough to describe the epistemic properties of the players. Logical relations between the various propositions are described by formal rules of inference.

The equivalence between the semantic and syntactic approaches has been studied by Aumann ([2]), Fagin et al. ([19]), who analyzed it by means of modal logic (Chellas (1998)), and Kaneko et al. ([34]), who were interested in the broader issue of the characterization of the concept of Common Knowledge in various logics.

In our paper we will show that the existence of a universal type space. As a preliminary step we will choose the formal language.[5] In economics the standard choice has been epistemic logic[6], instead we prefer to use first order logic.

Our choice can be justified with two motives: first, we can rely on the results obtained in mathematical logic[7], and specifically model theory[8]; second, there has been a huge debate on whether modal logic is a completely reliable tool

[4] Quite opposite.

[5] This language could be used by the players for communication or by the economists to analyze their strategic interaction.

[6] As a a matter of fact, epistemic logic has been widely used in economics for dealing with the rationales of knowledge and beliefs when multiple agents interact.

[7] We mention that the distinction between expression and their intended meanings is the very starting point in logic and not by chance this fact is stated as syntax versus semantics of a language.

[8] Model theory is a branch of mathematical logic that considering which mathematical models satisfy a certain theory.

(refer to Quine (1964)[9]).
Moreover, first-order logic will prove to be a more flexible and natural language for describing the interactive contexts. By means of first order logic, we can explicitly define arbitrarily high order of reasoning and common knowledge without any hidden assumption.
This choice will be crucial for stating the equivalence between the two approaches that are commonly used in the economic literature for modelling games with incomplete information.[10]

In order to investigate of the existence of a universal type space we will be the natural candidate as Universal Type Space. But, we will single out an essential problem on the uniqueness of the Universal Types Space. In order to overcome this problem and for plainly defining concepts as common knowledge we will enrich the initial formal language. Then, we will state sufficient conditions for the Universal Type Space to exist and derive some of its properties.

The essay is organized as follows. Section 2 introduces the general framework and the preliminary structures needed. Namely, the formal language used to describe player reasoning and a natural candidate as Universal Type Space. In Section 3 we can point out an essential problem for the existence of a universal type space. In order to investigate this problem we need to enrich the formal language considered. Namely, we allow for expression of infinite length. In this richer framework we can state sufficient and tight conditions for the universal type space to exist. Lastly, we will identify some characteristics of the universal type space i.e., it is compact and it contains at least one self-evident event.

1.2 The framework

Starting from a set N of agents, labelled by $1, 2, ..., n$, we assume that they want to reason about an external setting, which can be described by a non-empty countable set of characteristics. Instead of using modal logic to describe interactive reasoning of the players we will consider a first-order language, which has been proved to be as rich and expressive as modal logic (see Fagin et al. [19] and Cappelletti (2003)[11]).
Formally a first-order language is given by specifying the following data:

[9] Quine, "The Problem of Interpreting Modal Logic" (1947)

[10] Moreover, we clearly separate the two types of analysis (semantic and syntactic), in order to disentangle purely logic issues (see Brandenburger and Keisler (2003)) from purely semantic or set-theoretic ones (see Heifetz and Samet (1998) [29]) and we find conditions, in purely linguistic terms, that exclude both.

1. a set of function symbols $\mathcal{F}$ and positive integers n_f for each $f \in \mathcal{F}$;

2. a set of relation symbols $\mathcal{R}$ and positive integers n_R for each $R \in \mathcal{R}$;

3. a set of constant symbols $\mathcal{C}$.

For our purpose, $\mathcal{R}$ is constituted by the countable set of unary relation symbols $\Phi = \{P_1, P_2, ...\}$, where each element P_k corresponds to a primitive charcteristics of the external setting, and by the set of binary relation symbols $\{R_1, R_2, ..., R_n\}$, one for each player. Our language is purely relational and it is composed by $\langle \Phi, R_1, R_2, ..., R_n \rangle$.

Unary relations describe basic facts about the external world, for example "it is snowing in Milan", "the price of oil reaches 20$" or "Player j chooses action a". To express statements like "Player i knows p", where p is an arbitrary element of Φ, we use the binary ralation in order to have an formula equivalent to the modal or epistemic operators $\{K_i\}_{i=1,2,...,n}$. Namely, we introduce the following abbreviation for expressing that an agent i know an arbritrary formula ϕ denoted as $K_i(\phi)$:

$$K_i(\phi) := \forall x_2 (R_i(x_1, x_2) \Rightarrow P_k(x_2)).^{11}$$

For example, $(K_3\phi)$ should be read as "Player 3 knows ϕ".
Now, we need to define well-formed formulas in this formal language. A formula is any string of symbols built using symbols of the language, variables symbols x_1, x_2.., the Boolean connectives "and" (in symbol $\wedge$), "or" (in symbol $\vee$), "not" (in symbol $\neg$) and the existence and universal quantifiers (in symbols $\exists$, $\forall$). Proceeding step by step we define atomic formulas and then formulas.[12]

Definition 1 *ϕ is an atomic formula if ϕ is either*

1. $x_1 = x_2$ where x_1 and x_2 are variables or;

2. $R_i(x_1, x_2)$ or $P_k(x_3)$, where x_1, x_2, x_3 are variables.

Definition 2 *The set of (well-formed) formulas is the smallest set $\mathcal{L}$ containing all atomic formulas and such that*

1. if ϕ is in $\mathcal{L}$ then $\neg\phi$ is in $\mathcal{L}$,

[12] Note that given any formulas ϕ^* and φ^* in $\mathcal{L}^*$, $\phi^* \vee \varphi^*$ is an abbreviation of $\neg(\neg\phi^* \wedge \neg\varphi^*)$. Moreover given any formula $\phi^*(x)$ with free variable x, the expression $\forall x \phi^*$ is an abbreviation of $\neg\forall x \neg\phi^*$. Therefore, wlog we can consider just the conjunction symbol and the existence quantifiers.

2. *if ϕ and λ are in $\mathcal{L}$ then $\phi \wedge \lambda$ is in $\mathcal{L}$, and*

3. *if ϕ is in $\mathcal{L}$ then $\exists x \phi$ is in $\mathcal{L}$.*

A variable x can occur freely or it can be bound by a quantifier. A formula with no free variable is called a *sentence*. We denote the set of formulas of our language as $\mathcal{L}$.[13]
For notation convenience, we will drop the free variable notation whenever this is not relevant for the exposition.[14]

Given a first order language, we need a semantic or $\mathcal{L}$-structure to give substance to the different formulas of $\mathcal{L}$.

Definition 3 *An $\mathcal{L}$-structure, $M = \left\langle U, \{P_k^M\}_{k=1}^K, \{R_i^M\}_{i=1}^n \right\rangle$, is given by the following elements:*

1. *a non-empty set U called universe of M;*

2. *a set $P_k^M \subseteq U$ for each unary relation P_k;*

3. *a set $R_i^M \subseteq U \times U$ for each binary relation R_i.*

If a language has constant symbols $c \in \mathcal{C}$ then their interpretations are just elements c^M of U. If ϕ is a formula in $\mathcal{L}$ with free variables $x_1, ..., x_n$ we can think of ϕ as expressing a property of elements of $U^n := U \times U \times ... \times U$, hence we must define what it means for $\phi(x_1, ..., x_n)$ to hold at $(a_1, ..., a_n) \in U^n$. Once more we will use the recursive nature of the well-formed formulas for a formal language.

Definition 4 *Let ϕ be a formula with free variables from $x = (x_1, ..., x_m)$ and let $\bar{a} = (a_1, ..., a_m) \in U^m$. We inductively define "M satisfies $\phi(\bar{a})$" or "$\phi(\bar{a})$ is true in M" (in symbols $M \models \phi[\bar{a}]$) as follows:*

1. *if ϕ is $R_i(x_1, x_2)$ then $M \models \phi[\bar{a}]$ if $\bar{a} = (a_1, a_2) \in R_i^M$;*

2. *if ϕ is $P_k(x)$ then $M \models \phi[\bar{a}]$ if $\bar{a} = (a) \in P_k^M$;*

3. *if ϕ is $(K_i p_k)$ then $M \models \phi[\bar{a}]$ if for every b, such that $(\bar{a}, b) \in R_i^M$, $M \models P_k[b]$;*[15]

[13] We adopt the convention that greek letters denote formulas, while latin letters denote primitive expressions. Starred greek letters denote first order logic formulas.

[14] For example we will often denote a formula ϕ^* without remarking whether there are free variables. Complete notations would be $\phi^*(x)$ where $x = \{x_1, ...x_n\}$ are n free variables.

[15] Let us recall that $(K_i p_k)^* = \forall x_2 R_i(x_1, x_2) \Rightarrow P_k(x_2)$

4. *if* ϕ *is* $\neg\psi$ *then* $M \models \phi[\bar{a}]$ *if* $M \not\models \psi[\bar{a}]$;

5. *if* ϕ *is* $(\psi \wedge \varphi)$ *then* $M \models \phi[\bar{a}]$ *if* $M \models \psi[\bar{a}]$ *and* $M \models \varphi[\bar{a}]$;

6. *if* ϕ *is* $\exists x\psi(\bar{a}, x)$, *then* $M \models \phi[\bar{a}]$ *if there is* $b \in U$ *such that* $M \models \psi[\bar{a}, b]$ *holds.*

7. *if* ϕ *is* $\forall x\psi(\bar{a}, x)$, *then* $M \models \phi[\bar{a}]$ *if for every* $b \in U$, $M \models \psi[\bar{a}, b]$ *holds.*

Now, we consider a set of sentences that describe some properties of the epistemic operator, namely a theory. Given a purely syntactic object as a theory we can investigate the class of $\mathcal{L}$-structures, that satisfy all the sentences of the theory.
For example, given the sentence $\forall x\phi(x)$, where $\phi(x)$ is a formula with just one free variable, and the $\mathcal{L}$-structure $M = \langle U, P^M, R^M \rangle$, if all the elements of U satisfy the formula $\phi(x)$, in symbols $M \models \phi[u]$, then $\langle U, P^M, R^M \rangle$ is a model for the theory consisting of the single sentence $\forall x\phi(x)$ (in symbols $M \models \forall x\phi(x)$).
We define any set of sentences a $\mathcal{L}$-theory, denoted as $\mathbf{T}$.

Definition 5 *A* $\mathcal{L}$*-structure,* M*, is a model of* $\mathbf{T}$ *(in symbols* $M \models \mathbf{T}$*) if and only if* $M \models \sigma$ *for all sentences* σ *belonging to* $\mathbf{T}$.

We will focus on a particular theory, denoted $\mathbf{T}^{rst}$, which corresponds to the Knowledge operator[16] in modal logic. The five axioms that characterize the epistemic operator and constitute our theory are the following $\mathcal{L}$-sentences:

$$\forall x_1 (K_i\varphi)(x_1) \wedge (K_i(\varphi \Rightarrow \psi))(x_1) \Rightarrow (K_i\psi)(x_1) \qquad \text{(Distribution Axiom)}$$

[17]

$$\text{From } \phi \text{ infer } (K_i\phi): \qquad \text{(Knowledge Generalization Rule)}$$

$$\forall x_1\phi(x_1) \Rightarrow \forall x_1 (K_i\phi)(x_1)$$

[16] Note that we restrict ourselves to $S5$ system. The same way of reasoning applies to different axiomatization of the knowledge operator.

[17] The plain expression is:

$$\forall x_2 \{[\forall x_1 (R_i(x_1, x_2) \Rightarrow \varphi^*(x_2)) \wedge \forall x_1 (R_i(x_1, x_2) \Rightarrow (\varphi^*(x_2) \Rightarrow \psi^*(x_2)))] \Rightarrow \forall x_1 (R_i(x_1, x_2) \Rightarrow \psi^*(x_2))\}$$

[18]

$$\forall x_1 \left(K_i\phi\right)(x_1) \Rightarrow \phi(x_1) \qquad \text{(Truth Axiom)}$$

[19]

$$\forall x_1 \left(K_i\phi\right)(x_1) \Rightarrow \left(K_iK_i\phi\right)(x_1) \qquad \text{(Positive Introspection)}$$

[20]

$$\forall x_1 \left(\neg K_i\phi\right)(x_1) \Rightarrow \left(K_i\neg K_i\phi\right)(x_1) \qquad \text{(Negative Introspection)}$$

[21]

$$\forall x_1 \left[\varphi(x_1) \wedge (\varphi \Rightarrow \psi)(x_1) \Rightarrow \psi(x_1)\right] \qquad \text{(Modus Ponens)}$$

[22]

In a first order language a $\mathcal{L}$-sentence, φ, is a logical consequence of a set of sentences, $\mathbf{T}$, if and only if it is true in any model of $\mathbf{T}$. We denote this fact as $\mathbf{T} \models \varphi$.

[18] The plain expression is:

$$\forall x_1 \phi^*(x_1) \Rightarrow \forall x_1 \left[\forall x_2 \left(R_i(x_1, x_2) \Rightarrow \phi^*(x_2)\right)\right]$$

[19] The plain expression is :

$$\forall x_1 \left[\forall x_2 \left(R_i(x_1, x_2) \Rightarrow \phi^*(x_2)\right)\right] \Rightarrow \phi^*(x_1)$$

[20] The plain expression is:

$$\forall x_1 \left[\forall x_2 \left(R_i(x_1, x_2) \Rightarrow \phi^*(x_2)\right)\right] \Rightarrow \left[\forall x_3 \left(R_i(x_3, x) \Rightarrow (K_i\phi)^*(x_3)\right)\right]$$
$$\forall x_1 \left[\forall x_2 \left(R_i(x_1, x_2) \Rightarrow \phi^*(x_2)\right)\right] \Rightarrow \left\{\forall x_3 R_i(x_3, x_1) \Rightarrow \left[\forall x_2 \left(R_i(x_3, x_2) \Rightarrow \phi^*(x_2)\right)\right]\right\}$$

[21] The plain expression is :

$$\forall x_1 \neg \left[\forall x_2 \left(R_i(x_1, x_2) \Rightarrow \phi^*(x_2)\right)\right] \Rightarrow \left[\forall x_3 \left(R_i(x_3, x_1) \Rightarrow \neg (K_i\phi)^*(x_3)\right)\right]$$
$$\forall x_1 \neg \left[\forall x_2 \left(R_i(x_1, x_2) \Rightarrow \phi^*(x_2)\right)\right] \Rightarrow \left\{\forall x_3 R_i(x_3, x_1) \Rightarrow \neg \left[\forall x_2 \left(R_i(x_3, x_2) \Rightarrow \phi^*(x_2)\right)\right]\right\}$$

[22] The plain expression is:

$$\forall x_1 \left[\varphi^*(x_1) \wedge \varphi^*(x_1) \Rightarrow \psi^*(x_1)\right] \Rightarrow \psi^*(x_1)$$

1.2.1 The Henkin construction and the Canonical Model

We need to determine whether there exists a model where all occurrences compatible with the theory $\mathbf{T}^{\mathbf{rst}}$ are satisfiable. Given a model M, for every possible description of individual characteristics, which does not contradict $\mathbf{T}^{\mathbf{rst}}$, there must be an element $s \in S$ such that the description holds at s. We can build such model using the so-called Henkin construction (see Chang and Keisler (1998) [13] or Marker (2002) [38]).

Given our first-order language $\mathcal{L}$ and the theory $\mathbf{T}^{\mathbf{rst}}$ we can consider a richer language $\mathcal{L}_1 \supseteq \mathcal{L}$, constructed adding a constant symbol for each $\mathcal{L}$-formula with one free variable. Namely, for a generic $\mathcal{L}$-formula, $\phi(x)$, there is an $\mathcal{L}_1 - constant$, c_ϕ, such that the following implication holds in our theory: $\exists x \phi(x) \Rightarrow \phi(c_\phi)$.[23] For every formula in $\mathcal{L}$ which is consistent with $\mathbf{T}^{\mathbf{rst}}$, we add a constant symbol which represents the possibility that the formula holds.

For example, let us take the formula $(K_i\varphi)(x)$, meaning that player i knows φ, then we consider a constant symbol $c_{K_i\varphi}$ such that the sentence $(K_i\varphi)(c_{K_i\varphi})$ is true. The intuition is that the constant witnesses the possibility that a player i could know φ. It is apparent that the Henkin construction resembles the universal type space built from expressions used, for example, by Heifetz and Samet (1998) [28].

Thus, the new language is $\mathcal{L}_1 := \mathcal{L} \cup \{c_\phi : \phi(x) \text{ an } \mathcal{L} - formula \text{ with one free variable}\}$. If for each $\mathcal{L}$-formula $\phi(x)$ we denote the corresponding $\mathcal{L}$-sentence $\Gamma_\phi := [\exists x \phi(x) \Rightarrow \phi(c_\phi)]$, then the new $\mathcal{L}_1$-theory is defined as:

$$\mathbf{T}_1 := \mathbf{T}^{\mathbf{rst}} \cup \{\Gamma_\varphi : \phi(x) \text{ is an } \mathcal{L} - formula \text{ with one free variable}\}.$$

Remark 1 *Given that any $\mathcal{L}$-structure corresponding to some Kripke structure M_{rst} is a model of $\mathbf{T}^{rst}$, any finite subset of sentences of $\mathbf{T}^{rst}$ has a model. Therefore any finite subset of $\mathbf{T}_1$ has a model.*

If a theory satisfies this properties it is said that the theory is "*finitely satisfiable*".

Definition 6 *An $\mathcal{L}$-theory $\mathbf{T}$ is finitely satisfiable if every finite subset of $\mathbf{T}$ admits a model or equivalently every finite subset of $\mathbf{T}$ is satisfiable.*

---

[23] In symbols, $\mathbf{T}^{\mathbf{rst}} \models \exists x \varphi^*(x) \Rightarrow \varphi^*(c_\varphi)$

This property has a major role in logic because it allows to prove that a not-finite theory has a model thanks to the compactness of many standard formal languages.

Now, we iteratively construct a sequence of languages, $\mathcal{L} \subseteq \mathcal{L}_1 \subsetneq \mathcal{L}_2, \ldots$, and a sequence of finitely satisfiable $\mathcal{L}_i$-theories, $\mathbf{T}^{\mathbf{rst}} \subseteq \mathbf{T}_1 \subseteq \mathbf{T}_2 \subseteq \ldots$. If $\phi(x)$ is a $\mathcal{L}_i$-formula, then there is a constant symbol $c_\phi \in \mathcal{L}_{i+1}$. c_ϕ is such that $\mathbf{T}_{i+1} \models (\exists x \phi(x)) \Rightarrow \phi(c_\phi)$.
We can define $\mathcal{L}_\infty := \bigcup_{i=1}^{\infty} \mathcal{L}_i$ and $\mathrm{T}_\infty := \bigcup_{i=1}^{\infty} \mathbf{T}_i$. If we take any $\mathcal{L}_\infty$-formula $\phi(x)$ with one free variable x then, by construction, there is a constant symbol $c_\phi \in \mathcal{L}_\infty$ such that, by construction, $\mathrm{T}_\infty \models (\exists x \phi(x) \Rightarrow \phi(c))$. This property is called the witness property of the theory.

Definition 7 *An $\mathcal{L}$-theory* $\mathbf{T}$ *has the* witness property *if whenever $\phi(x)$ is a $\mathcal{L}$-formula with one free variable x, then there is a constant symbol* $c \in \mathcal{L}$ *such that* $\mathrm{T} \models (\exists x \phi(x) \Rightarrow \phi(c))$.

If a formula is consistent with the theory then there is a constant that represents it. $\mathbf{T}_\infty$ has at most a countable number of sentences because the set of well-formed formula is at most countable. Moreover, any finite subset of $\mathbf{T}_\infty$ admits a model.

We would like that any sentence is either part of the considered theory or is incompatible with it.

Definition 8 *An $\mathcal{L}$-theory* $\mathbf{T}$ *is* maximal *if for any sentence ϕ either $\phi \in \mathbf{T}$ or* $\neg\phi \in \mathbf{T}$.

We can ensure that there is a finitely satisfiable $\mathcal{L}_\infty$-theory, $\mathbf{T}'_\infty \supseteq \mathbf{T}_\infty$ such that for every $\mathcal{L}_\infty$-sentence, ϕ, either $\phi \in \mathbf{T}'_\infty$ or $\neg\phi \in \mathbf{T}'_\infty$ i.e., there exists a maximal finitely satisfiable $\mathcal{L}_\infty$-theory.[24]

Once we have defined this fully encompassing language and theory, we can consider the so called "canonical model", denoted as M'_∞, of $\mathbf{T}'_\infty$. This model is a $\mathcal{L}_\infty$-structure and its universe is $U := \mathcal{C}/\sim$, where $\mathcal{C}$ is the set of constant symbols of $\mathcal{L}_\infty$. For any $c, d \in \mathcal{C}$, $c \sim d$ if and only if $\mathbf{T}'_\infty \models c = d$.

[24] This is a consequence of the following theorem (refer to Marker (2002)[38]).

Theorem 1 *If* $\mathbf{T}^*$ *is a finitely satisfiable $\mathcal{L}^*$-theory then there is a maximal finitely satisfiable $\mathcal{L}^*$-theory* $\mathbf{T}' \supseteq \mathbf{T}^*$.

Thus, the universe is the set of equivalence classes of $\mathcal{C}$ with respect to $\sim$. The interpretation of the constant symbol $c \in \mathcal{C}$ is $c^{M'_\infty} = \bar{c}$.[25]

The interpretation of the relations symbols of $\mathcal{L}_\infty$ are respectively:

$$
\begin{aligned}
P_k^{M'_\infty} &= \{\bar{c}_1 : P_k(\bar{c}_1) \in \mathbf{T}'_\infty\} \text{ for } k = 1, 2, ... \\
R_i^{M'_\infty} &= \{(\bar{c}_1, \bar{c}_2) : R_i(\bar{c}_1, \bar{c}_2) \in \mathbf{T}'_\infty\} \text{ for } i = 1, 2, ..., n
\end{aligned}
$$

$P_k^{M'_\infty}$ and $R_i^{M'_\infty}$ are well-defined.[26]

Therefore, we can define a $\mathcal{L}_\infty$-structure, called canonical model of $\mathbf{T}'_\infty$:

$$
M'_\infty := \left\langle U, \left\{P_k^{M'_\infty}\right\}_{k \geq 1}, \left\{R_i^{M'_\infty}\right\}_{i=1}^{n} \right\rangle .
$$

By construction, the canonical model includes any conceivable combination of epistemic characteristics for the n players. Therefore, it is the natural candidate for being a universal type space.

1.3 Existence of a universal type space and its properties

First, we need to relate the concept of type space, commonly used in the economic literature, and the concept of $\mathcal{L}$-structure. Then, we can show an essential problem for the existence of the universal type space and propose our solution.

1.3.1 Existence of a universal type: an essential problem

There has been two kinds of definition of universal type space: one constructive (see [39]) and the other implicit (see [28]). The former relies on the informal description of player higher order of uncertainty leading to a hierarchical construction in a set theoretic framework. Note that this construction leads to a structure which is equivalent to the canonical model in a logical framework. In order to be more precise about the second definition we need some preliminary definitions.

---

[25] Where c^* denotes the equivalence class of c.

[26] If $c^* \sim d^*$ then $P(c^*) \in \mathbf{T}'_\infty$ if and only if $P(d^*) \in \mathbf{T}'_\infty$. Note that if $c^* \sim d^*$ then $c^* = d^*$ belongs to $\mathbf{T}'_\infty$. Hence if $P(c^*) \in \mathbf{T}'_\infty$ then $P(d^*)$ is the logical consequence of $c^* = d^*$ and $P(c^*) \in \mathbf{T}'_\infty$. By maximality of $\mathbf{T}'_\infty$, it implies that $P(d^*) \in \mathbf{T}'_\infty$. A similar reasoning applies to the binary relations.

Definition 9 *Let $\langle T,m\rangle$ and $\langle T',m'\rangle$ be two type spaces on Φ, let $(\varphi_i)_{i\in I_0}$ I_0-tuple of measurable functions $\varphi_i : T_i \to T_i'$. The induced function $\varphi : T \to T'$ is called a type morphism if:*

1. *φ_0 is the identity on Φ;*
2. *for each $i \in I$, $m' \circ \varphi_i = (m_i \circ \varphi^{-1})$.*

In other words, a type morphism embeds $\langle T,m\rangle$ in $\langle T',m'\rangle$ and preserves the epistemic characteristics of the players. If we look at type spaces in terms of $\mathcal{L}$-structures, then a type morphism is just a function between two $\mathcal{L}$-structures.[27] As already mentioned, a type morphism preserve the property of any epistemic operator.

Definition 10 *A type space $\langle T,m\rangle$ on Φ is universal if for every type space $\langle T',m'\rangle$ on Φ there is a unique type morphism from T to T'.*

Remark 2 *Note that the canonical model is, by construction, a universal type space.*

Heifetz and Samet (1998) [28] proved that for any measurable space Φ there exists a unique universal type space on Φ. In a companion paper Heifetz and Samet (1999) [29] gave an example of a coherent hierarchy of beliefs that could not be extended to a belief on the space of all coherent hierarchies. The parallelism, therefore, between the explicit and the implicit construction of epistemic characteristics of players breaks down. In fact the universal type space still exists but it does not coincide with the hierachical construction.

Similar results were obtained by the same authors for the so-called knowledge type space on Φ (Heifetz and Samet (1998) and (1999) [27] [30]).[28]

Moreover, Heifetz and Samet (1998) [27] and Fagin (1994) provided an example where the complete hierarchical description of the epistemic properties of agents is unbounded. Namely, if there are at least two players and at least two states of nature then there is no universal knowledge space.[29]

---

[27] In the appendix we will be more precise and use the technical definition of $\mathcal{L}^*$-elementary embedding between the two $\mathcal{L}^*$-structures.

[28] **Definition:** A (knowledge) type space on Φ is the triple $\langle \Omega, \Theta, (t_i)_{i\in I}\rangle$, where Ω is a non-empty set whose elements are called *states of the world*, $\Theta : \Omega \to \Phi$ specifies for each state of the world the state of nature that prevails there, and for each player $i \in I$, t_i is a type function from Ω to $\Delta(\Omega)$ such that the support of $t_i(\omega)$ is $\Pi_i(\omega)$.
Note that Π_i is a partition of Ω, hence $\Pi_i(\cdot)$ maps Ω to a subset of the power set of Ω.

[29] Thus, the knowledge spaces, as ordinal numbers (Halmos (1974) [22]), are unbounded, and in general there is no knowledge space in which we can embed any conceivable knowledge space.

We would like to analyze these results employing our framework. The language used to describe the state of nature and players' reasoning will be the building block. As a first result, thanks to a classic result in model theory we will show an essential problem for the existence and uniqueness of a universal type space.

In the previous section we have stated plainly the equivalence between a type space and $\mathcal{L}$-structure. Therefore we can apply a classic result in model theory, the Upward Löwenheim-Skolem Theorem in order to prove the following proposition.

Proposition 1 *Given any type space* $M = \langle (T_i)_{i \in I_0}, (m_i)_{i \in I} \rangle$ *there is another type space, N, such that any formula satisfiable in M is satisfiable in N. Moreover N has cardinality at least* $|M| + \omega$.

Proof. See Appendix 1.A. ■

Given any infinite type space $\langle (T_i)_{i \in I_0}, (m_i)_{i \in I} \rangle$ there is another type space, that satisfies the same formulas and it has cardinality of at least $|M|+\omega$. Where $|M|$ is the cardinality of the type space, or equivalently of the structure.
This result strengthens Theorem 2.3 of Heifetz and Samet ([28]). In general the minimal embedding model of M has cardinality equal to $|M| + \omega$.

Hence, the universal type space itself can be strictly embedded in another type space. From a logic point of view this might not be an issue, because theorems provable in universal type space are still valid in the "larger" model. But from a game theoretical perspective this result casts some doubts on the existence of a universal type space and on the claim that the universal type space is common knowledge.

In the next section we will generalize our framework for discussing the problem raised in the economic literature and by Proposition 1. Next, we will provide sufficient conditions for the existence of the universal type and we will single out some of its properties.

1.3.2 Main Results

First, we need to make some remarks. If the external state is uncountable then we need an uncountable set of unary relation symbols and this drives us away from standard first-order logic. Instead, if we assume that the state of nature is definable with a finite set of properties measure theoretic problems, which are the basis of the results of Heifetz and Samet (1999) [29], are excluded.

Similarly, if we want to study interactive reasoning and common knowledge we need at least a countable conjunction of formulas, hence we have another reason to depart from standard first-order languages. But, if we consider just first order languages, then unbounded regression for describing players' reasoning, is excluded by definition . Actually, if we focus on the linguistic perspective there is obviously a complete description of players' epistemic characteristics, because well-formed formulas are at most countable[30].
Therefore, one obvious solution is to consider only standard first - order languages. This implies that a "canonical" type space exists[31], but Proposition 1 implies that a universal type space does not exist.

We will enrich our language and refer to the infinitary logic (see Keisler (1991) [35]). Infinitary logic allows us to express concept like common knowledge without introducing new symbols.[32]
As first guess we would need a language which admits countable conjunctions of formulas, referring for example to the definition of common knowledge in modal logic. We will show that this guess is wrong and we need an even richer language.

In the previous section we have defined the equivalent formula for the modal expression $K_i\phi$, that is $(K_i\phi)(x_1) := \forall y (R_i(x_1, x_2) \Rightarrow \phi(x_2))$ or more explicitly $(K_i\phi)(x_1) := \forall x_2 (\neg R_i(x_1, x_2) \vee \phi(x_2))$. If we want to say that player i knows that player j knows the proposition ϕ, in symbols $K_iK_j\phi$, the following $\mathcal{L}$-formula is needed :

$$\begin{aligned} (K_iK_j\phi)(x_1) &= \forall x_2 \{R_i(x_1, x_2) \Rightarrow (K_j\phi)(x_2)\} = \\ &= \forall x_2 \{\neg R_i(x_1, x_2) \vee [\forall x_3 \neg R_i(x_2, x_3) \vee \phi(x_3)]\} \end{aligned}$$

Given this simple example, it becomes apparent that we need a language that allows for countable conjunction and countable quantification, in particular we will use the logic denoted with $\mathcal{L}_{\omega_1,\omega_1}$[33], which is built from first order language by allowing countably[34] infinite disjunctions, conjunctions and quantifiers.

[30] Sophistication can be expressed up to an at most countable order, but if we look at the semantic counterpart we might go "deeper".

[31] The existence of a universal type space and its uniqueness remain problematic.

[32] We will be able to determine whether there is some connections between properties of the knowledge operator and properties of common knowledge

[33] ω_1 is the first ordinal number bigger than ω.

[34] To be precise we should allow for uncountable disjunctions and conjunctions to capture completely the result on knowledge space, but we it will become clearer below why this is

Therefore, we will encounter all the difficulties related to this kind of formal languages, for example the lack of compactness[35], existence of a model and completeness.

Given the our language $\mathcal{L} = \left\langle \{P_k\}_{k\geq 1}, \{R_n\}_{n\in N} \right\rangle$, where P_k is a unary relation symbol, corresponding to one of the "external" state characteristic and R_n is a binary relation symbol, one for each player $n \in N$.
The language $\mathcal{L}_{\omega_1,\omega_1}$ has the same logic symbols as $\mathcal{L}$ but the conjunction symbol, $\wedge$, can be applied to a countable or finite set of formulas, and, similarly, the quantification symbols, $\forall$ and $\exists$, can be applied to a countable or finite set of variables.

Definition 11 *We denote with $\mathcal{F}_{\omega_1,\omega_1}$ the smallest class of formulas such that:*

1. *$R_n(x,y)$ and $P_k(z)$, where x, y, z are variable symbols, belong to $\mathcal{F}_{\omega_1,\omega_1}$;*

2. *if ϕ belongs to $\mathcal{F}_{\omega_1,\omega_1}$ and the at most countable set of variable symbols*[36] *$\{x_\alpha\}_{\alpha\in A}$ then $\neg\phi$ and $\forall\{x_\alpha\}\phi$ belong to $\mathcal{F}_{\omega_1,\omega_1}$;*

3. *if $\Phi = \{\phi, \phi_2, ...\}$ is a finite or countable, non-empty, subset of $\mathcal{F}_{\omega_1,\omega_1}$ then $\bigwedge_{i\geq 1} \phi_i$ belongs to $\mathcal{F}_{\omega_1,\omega_1}$.*

Once we have defined the set of well-formed formulas $\mathcal{F}_{\omega_1,\omega_1}$, we need to enrich the rules of inference, similarly to what is done for the language $\mathcal{L}_{\omega_1,\omega}$ (see [35]):[37]

- from φ and $\varphi \Rightarrow \theta$ imply θ (Modus Ponens in $\mathcal{L}_{\omega_1,\omega_1}$);[38]

- from $\varphi \Rightarrow \theta$ infer $\varphi \Rightarrow \forall x\theta(x,...)$ where the variable x does not occur free in θ (Generalization);

- from $\varphi \Rightarrow \phi$ for all ϕ belonging to the at most countable set of formulas Φ infer $\bigwedge \Phi$ ($\mathcal{L}_{\omega_1,\omega_1}$–inference).

Using this formal language, we are able to define explicitly mutual knowledge $(E\varphi)(x) := \bigwedge_{n\in N} K\varphi(x)$ and common knowledge $(CK\varphi)(x) := \bigwedge_{k\geq 1} [(E\varphi)]^k(x)$

unnecessary.

[35] In a compact language if a theory is finitely satisfiable then it is satisfiable.

[36] The index set, A, is at most countable.

[37] $\mathcal{L}^*_{\omega_1,\omega}$ is a formal language where it is not allowed for countable quantification.

[38] Note that φ and θ belong t $\mathcal{F}^*_{\omega_1,\omega_1}$, hence in a trivial sense this is a generalization of the Modus Ponens.

with well-formed formulas in the richer language $\mathcal{L}_{\omega_1,\omega_1}$.

Now, we turn to the question of whether there exists a universal type space. We could pursue a constructive derivation. Namely, we construct the canonical model, similarly to what we have done for the finitary case. Unfortunately, this is not possible without additional assumptions, because our infinitary language is incomplete. In $\mathcal{L}_{\omega_1,\omega_1}$ there are valid formulas that are not deducible from the above axiom scheme.[39] Hence the construction of the so called canonical model stops at its very beginning because we are not able to determine whether a sentence follows from our theory $\mathbf{T}^{\mathrm{rst}}$.

Proposition 2 *There is not a canonical model such that each state of the world represents the maximal set of consistent formulas.*

Proof. See Appendix 1.B. ■

A possible solution is to take a set of all the formulas of $\mathcal{L}_{\omega_1,\omega}$, which is a subset of $\mathcal{F}_{\omega_1,\omega_1}$, and consider its closure with respect to the following two axioms: [40]

$$\forall x \bigwedge_{k\geq 1} (K\phi_k)(x) \Rightarrow \left(K \bigwedge_{k\geq 1} \phi_k\right)(x) \qquad \text{(Epistemic Continuity)}$$

$$\text{for any } \{\phi_k\}_{k\geq 1} \text{ belonging to } \mathcal{L}_{\omega_1,\omega}$$

and

$$\forall x \bigvee_{k\geq 1} \neg (K\phi_k)(x) \Rightarrow \neg \left(K \bigwedge_{k\geq 1} \phi_k\right)(x) \qquad \text{(Epistemic Monotonicity)}$$

$$\text{for any } \{\phi_k\}_{k\geq 1} \text{ belonging to } \mathcal{L}_{\omega_1,\omega}$$

We denote this fragment of $\mathcal{F}_{\omega_1,\omega_1}$ as $\mathcal{A}_{\omega_1,\omega_1}$. Once we have restricted our language to a transitive closure of $\mathcal{L}_{\omega_1,\omega}$, we can proceed with a construction a' la Henkin and obtain a canonical space. This implies that a canonical model exists. Moreover, a universal type space exists, given that Proposition 1 does not apply in $\mathcal{L}_{\omega_1,\omega}$. We formalize this result in the following Theorem.

[39] $\mathcal{L}^{*}_{\omega_1,\omega_1}$ is incomplete as a consequence of Scott's indefinability theorem (see [Bell 2000]) and the set of valid sentences for $\mathcal{L}^{*}_{\omega_1,\omega}$ is not recursively enumerable as in first order logic.

[40] The two axioms are strictly related to the Barcan property (see [34])

Theorem 2 *Given Axioms Epistemic Continuity and Epistemic Monotonicity a universal type space exists and it is unique.*

Proof. See Appendix 1.D. ∎

We can refer to Axioms Epistemic Continuity and Epistemic Monotonicity to conclude that the set of all the finite descriptions of epistemic properties of a player determine completely his epistemic type. Therefore, we can conclude that the canonical model is our universal type space which includes any other conceivable type space and which can not be embedded in a strictly bigger type space.[41]

Now, we can investigate some properties of the universal type space.

Lemma 1 *The universal or canonical type space is compact.*

Proof. See Appendix 1.C. ∎

Given the previous results, Axioms Epistemic Continuity and Epistemic Monotonicity not only guarantee the existence of a universal type space but they justify the widespread use of a compact space (see [39]).

At last, the following proposition makes precise in what sense the partitions could be considered common knowledge (see Aumann (1967)).

Lemma 2 *Given the Epistemic Continuity Axiom there exists a formula $\varphi(x)$ such that:*[42]

$$\forall x \phi(x) \Rightarrow (K\phi)(x)$$

Proof. Take $(CK\phi)(x)$ then by definition of common knowledge and the truth axiom $\forall x ([E\phi])^{k+1}(x) \Rightarrow \left[K(E\phi)^k\right](x)$. Therefore $\forall x \bigwedge_{k\geq 1} ([E\phi])^{k+1}(x) \Rightarrow \bigwedge_{k\geq 1} K([E\phi])^k(x)$ by the third inference rule that we have introduced in (1.3.2) and finally by the Epistemic Continuity Axiom (Epistemic Continuity) we can conclude that $\forall x \bigwedge_{k\geq 1} K([E\phi])^k(x) \Rightarrow K\left(\bigwedge_{k\geq 1}([E\phi])^k\right)(x)$.

Hence $\forall x (CK\phi)(x) \Rightarrow (K(CK\phi))(x)$. ∎

We can conclude that the canonical space is well-behaved. By construction, it is sufficient for describing epistemic condition of players. Moreover, there is at lest one self-evident event.

[41] Note that we have followED a strictly constructive procedure in a purely syntactic framework, which seems the appropriate "place" to do that.

[42] We suppress the player's index for the knowledge operator because the following sentence holds for any player.

1.4 Discussion

In the economic literature it is an important question whether or not there exists a universal type space and whether it is well defined, because it is the starting point for "the Bayesian foundations of solution concepts of games" ([54]). The major results were found by Boge and Eisele (1979), Mertens and Zamir (1985) [39], Heifetz and Samet (1998) [28], Brandenburger and Dekel (1995) among others. Most of the results are based on a hierarchical construction that finds its intuition in the linguistic description of all order of uncertainty, that players may face upon reasoning on a strategic situation. From our point of view, the length of the reasoning, made by players, is constrained by the language used by them or by the researcher. Once we have decided the language to use, the language itself will be the guidance for building the space of uncertainty and therefore the universal space.

First, we recall some standard concept used in game theory for modelling incomplete information settings. Given a set of states of nature Φ and a set of players $I = \{1, 2, ..., n\}$, we denote with I_0 the set of players including the external state (see [29]).

Definition 12 *A (belief) type space on* Φ *is a pair* $\langle (T_i)_{i\in I_0}, (m_i)_{i\in I}\rangle$ *or for short* $\langle T, m\rangle$ *where:*

1. $T_0 = \Phi$ *and* T_i*, for* $i \in I$ *is a measurable space;*
2. *for each* $i \in I$*,* m_i *is a measurable function* $m_i : T_i \to \Delta(T)$ *and* m_0 *is the identity map;*
3. *for each* $i \in I$ *and* $t_i \in T_i$*, the marginal of* $m_i(t_i)$ *on* T_i *assigns a probability equal to 1 to* t_i*.*

A type space could be associated to a Kripke structure, or equivalently a $\mathcal{L}$-structure.[43] Indeed, given a type space $\langle (T_i)_{i\in I_0}, (m_i)_{i\in I}\rangle$, let T be the product set $\prod_{i\in I} T_i$, and let $\Pi_i : T \to T_i$ be the natural $i^{th}-projection$ function, then we can define $m_i \circ \Pi_i : T \to \Delta(T)$. For each $t \in T$ we consider the support of the probability measure $m_i \circ \Pi_i(t)$. The t's, belonging to the support of $m_i \circ \Pi_i(t)$, are those elements $t' \in T$ that player i thinks are possible given t. If we consider

[43] Refer to Cappelletti (2003)[11].

the graph of the correspondence $\mathbf{Supp}\,(m_i \circ \Pi_i)\,(\cdot) : T \to 2^T$, we can define the binary relation:

$$\mathcal{K}_i := \{(t,t') \in T \times T : t' \in \mathbf{Supp}\,(m_i \circ \Pi_i)\,(t)\}$$

for each player i. In order to define a $\mathcal{L}$-structures we need an interpretation function, $\pi(t) : \Phi \to \{True, False\}$ for each $t \in T$, but we can think of a degenerate version of it i.e., $\pi(t) : \Phi \to \{True\}$. [44]

Once we have stated plainly this equivalence, we can turn to the question of the existence of a universal type. In the following section we will briefly discuss previous results in the literature and then apply our framework in order to answer this question.

1.5 Conclusions

Using techniques of model theory[45], we single out that the existence a proper universal type space is problematic. We identify sufficient condition in order to prove the existence of a universal type space. Finally, we state the main properties of the resulting universal type space.

The conditions that we have identified are purely logical or linguistic. They are not related to any set theoretic framework. A part from the theoretical result, we were able to, first, state the role of interpretation to pin down the main characteristics of a universal type space and, second, identify the assumption that justify the use of the Bayesian framework for analyzing a incomplete information setting. The obtained universal type space is quite well-behaved. It's a compact set endowed with at least one self evident event.

[44] This recalls the distinction between Aumann type space, or Kripke frame, and the fully-fledged Kripke structure.

[45] Aumann ([2]) already proved this by means of modal logic.

1.A Proof of Proposition 1

We need some definitions.

Definition 13 *Two $\mathcal{L}$-structures M and N are elementary equivalent (in symbols $N \equiv M$) if*

$$M \models \phi \text{ iff } N \models \phi$$

for all $\mathcal{L}$−sentences ϕ.

By definition two elementary equivalent $\mathcal{L}$-structures have the same theory.

Definition 14 *If M and N are $\mathcal{L}$-structures, with universes U and W, respectively. An $\mathcal{L}$ − embedding $\eta : M \to N$ is a one to one map $\eta : U \to W$ that preserves the interpretation of all the symbols of $\mathcal{L}$:*

1. *$\eta\left(f^M\left(a_1, ..., a_{n_f}\right)\right) = f^N\left(\eta\left(a_1\right), ..., \eta\left(a_{n_f}\right)\right)$ for all function symbols $f \in \mathcal{F}$ and $a_1, ..., a_{n_f} \in M$;*

2. *$\left(a_1, ..., a_{n_R}\right) \in R^M$ iff $\left(\eta\left(a_1\right), ..., \eta\left(a_{n_R}\right)\right) \in R^N$ for all relation symbols $R \in \mathcal{R}$ and $a_1, ..., a_{n_R} \in M$;*

3. *$\eta\left(c^M\right) = c^N$ for all constant symbols $c \in \mathcal{C}$.*

If $\mathcal{L}$ − *embedding* η is bijective then η is called an $\mathcal{L}$ − *isomorphism.*

Any isomorphism preserves the validity of $\mathcal{L}$-sentences i.e., given a sentence holding in M then if we consider its image through η, the sentence still holds in N. The last definition that we need is the following:

Definition 15 *If M and N are $\mathcal{L}$-structures, then an $\mathcal{L}$ − embedding, $j : M \to N$ is called an elementary embedding if*

$$M \models \phi\left[a_1, ..., a_n\right] \text{ iff } N \models \phi\left[j\left(a_1\right), ..., j\left(a_n\right)\right]$$

for all $\mathcal{L}$-formula and all $a_1, ..., a_n \in M$. M and N are called elementary equivalent.

The cardinality (in symbols $|\mathcal{L}|$) of a first order language is equal to the number of well-formed formula in the language; similarly the cardinality of a $\mathcal{L}$-structure M (in symbols $|M|$) is the cardinality of its universe. Now we can state the Upward Löwenheim-Skolem Theorem.

Theorem 3 (Upward Löwenheim-Skolem Theorem) *Let M be an infinite $\mathcal{L}$-structure and k be an infinite cardinal $k \geq |\mathcal{L}| + |M|$. Then there is an $\mathcal{L}$-structure, N, of cardinality k and an $\mathcal{L}-$ embedding $j : M \to N$ that is elementary.*

As a direct implication, given any infinite type space $\langle (T_i)_{i \in I_0}, (m_i)_{i \in I} \rangle$ there is another type space, that satisfies the same formulas and it has cardinality of at least $|M| + \omega$. Where $|M|$ is the cardinality of the type space, or equivalently of the structure. We can state this result as a proposition.

Proposition 3 *Given any type space $M = \langle (T_i)_{i \in I_0}, (m_i)_{i \in I} \rangle$ there is another type space, N, such that any formula satisfiable in M is satisfiable in N. Moreover N has cardinality at least $|M| + \omega$.*

Proof. We know from the previous constructions that M is a $\mathcal{L}$-structure. Given that the cardinality of our language is ω (see Halmos [22])), i.e. $|\mathcal{L}| = \omega$, the thesis follows from the Upward Löwenheim-Skolem Theorem (Theorem 3) ■

1.B Proof of Proposition 2

First we state the Scoot's indefinability theorem.

Theorem 4 (Scott's Undefinability Theorem for $\mathcal{L}_{\omega_1,\omega_1}$) *The set of valid sentence for $\mathcal{L}_{\omega_1,\omega_1}$ is not definable in coding structure of hereditarily infinite sets $H(\omega_1)$ by any $\mathcal{L}_{\omega_1,\omega_1}$-formulas.*

Given that the canonical model is built by witnessing all the $\mathcal{L}_{\omega_1,\omega_1}$-formulas that are compatible with a theory, as a result of Scott's theorem there are formulas, more precisely sentences, for which we can not say whether they are compatible or not with any given theory.

Proposition 4 *There is not a canonical or universal model such that each state of the world represents the maximal set of consistent formulas.*

Proof. A sentence is a formulas without free variable, hence the statement follows from Scott's indefinability Theorem 4. ■

1.C Proof of Lemma 1

Recall that $\mathbf{T}_{+}^{\mathbf{rst}\prime}$ is the maximal, finitely satisfiable theory containing $\mathbf{T}_{+}^{\mathbf{rst}}$. We need some definition before we can prove Lemma 1.

Definition 16 *Let p be the set of $\mathcal{A}_{\omega_1,\omega_1}$-formulas in free variables $x_1, ..., x_n$, p is an $n-$type if $p \cup \mathbf{T}_{+}^{\mathbf{rst}\prime}$ is satisfiable.*
p is a complete $n-$type if for all $\mathcal{A}_{\omega_1,\omega_1}$-formula, φ, with free variables $x_1, ..., x_n$ either $\varphi \in p$ or $\varphi \notin p$.
Denote with S_n^T the set of all complete $n-$types of theory $\mathbf{T}_{+}^{\mathbf{rst}\prime}$.

Take the set of *complete $n-$types* of $\mathcal{A}_{\omega_1,\omega_1}$ of theory $\mathbf{T}_{+}^{\mathbf{rst}\prime}$ and for $\varphi \in \mathcal{A}_{\omega_1,\omega_1}$, we can define:

$$[\varphi] = \{p \in S_n^T : \varphi \in p\}$$

If p is complete type and $\phi \vee \gamma \in p$, then $\phi \in p$ and $\gamma \in p$. Thus we have $[\phi \vee \gamma] = [\phi] \cup [\gamma]$. Similarly $[\phi \wedge \gamma] = [\phi] \cap [\gamma]$.

Now we can define the so-called Stone topology on S_n^T.

Definition 17 *The Stone topology on S_n^T is the topology generated by taking the sets $[\phi]$ as basic open sets.*

By definition, if p is a complete types then exactly one of ϕ and $\neg\phi$ is in p. Thus $[\phi] = S_n^T \setminus [\neg\phi]$ is also closed. Then the Stones topology is composed by sets that are both open and closed, or briefly *clopen.*

Lemma 3 *S_n^T is compact.*

Proof. We need to prove that every cover of S_n^T, by open sets, has a finite subcover.

Suppose not, then let $C := \{[\varphi_i(v)] : i \in I\}$ be a cover of S_n^T by basic open sets without any finite subcover. Let $\Sigma = \{\neg\varphi_i(v) : i \in I\}$.

We claim that $\Sigma \cup \mathbf{T}_{+}^{\mathbf{rst}\prime}$ is finitely satisfiable

Given the assumption that there is no finite subcover of C, if I_0 is a finite subset of I then, there is a type p such that $p \notin \bigcup_{i \in I_0} [\varphi_i(v)]$. Let N be a model of $\mathbf{T}_{+}^{\mathbf{rst}\prime}$, such that there is an element u of its universe satisfying all the formulas of p, i.e. there exists u such that $N \models \phi[u]$ for any formula ϕ belonging to p.

Then $N \models \mathbf{T}_{+}^{\mathbf{rst}\prime} \cup \bigwedge_{i \in I_0} \neg\varphi_i(u)$, so Σ is finitely satisfiable. Hence by lemma 6 Σ is satisfiable.

Hence take a model of $\mathbf{T}_{+}^{\mathrm{rst}'} \cup \Sigma$ such that there is an element u' of its universe, such that $N' \models \phi[u']$ for any formula ϕ belonging to Σ.

Then the complete type $\{\varphi(v) \in \mathcal{A}_{\omega_1,\omega_1} : N \models \varphi(u')\} \in S_n^T \backslash \bigcup_{i \in I} [\varphi_i(v)]$, a contradiction, given the assumption that C is a cover of S_n^T. ■

Lemma 4 *The universal type space is compact.*

Proof. Take the set of *complete* $n-types$ of $\mathcal{A}_{\omega_1,\omega_1}$ and for $\varphi \in \mathcal{A}_{\omega_1,\omega_1}$ let

$$[\varphi] = \{p \in S_n^T : \varphi \in p\}$$

As already said, the Stone topology on S_n^T is the topology generated by taking the sets $[\varphi]$ as open sets. Following Lemma 3 we can conclude S_n^T is compact moreover by the Tychonoff theorem (see Dudley 2003) $\prod_{n \geq 1} S_n^T$ is compact relatively to the product topology. ■

1.D Proof of Theorem 2

We just restate the definition of maximal theory for this specific language:

Definition 18 *A theory $\mathcal{A}_{\omega_1,\omega_1}$-theory, $\mathbf{T}$, is $\mathcal{A}_{\omega_1,\omega_1}-$ maximal if and only if for every sentence, σ, belonging to $\mathcal{A}_{\omega_1,\omega_1}$ either $\sigma \in \mathbf{T}$ or $\neg\sigma \in \mathbf{T}$.*

Lemma 5 *Suppose that a $\mathcal{A}_{\omega_1,\omega_1}$-theory, $\mathbf{T}$, is $\mathcal{A}_{\omega_1,\omega_1}$-maximal and it has the witness property. If every finite subset of sentences in $\mathbf{T}$, has a model then $\mathbf{T}$ has a model.*

Proof. Let $\mathcal{C}$ be the set of constant symbols of $\mathcal{A}_{\omega_1,\omega_1}$. For any $c, d \in \mathcal{C}$ $c \sim d$ iff $T \models c = d$. Note that our theory now includes Axioms Epistemic Continuity and Epistemic Monotonicity, hence given a countable set of formulas, $\{\phi_k\}_{k \geq 1}$, belonging to $\mathcal{A}_{\omega_1,\omega_1}$ the formula $\left(K \bigwedge_{k \geq 1} \phi_k\right)(x)$ has as a witness constant that is $\sim -equivalent$ to the witness of $\left(\bigwedge_{k \geq 1} \phi_k\right)(x)$.

As in the finitary case, it is easy to verify that $\sim$ is an equivalence relation. The universe of our model will be $M = \mathcal{C}/\sim$, that is the equivalence classes of $\mathcal{C}$ modulus $\sim$.

With a little abuse of notation we denote the equivalence class with an arbitrary element c^M belonging to it, hence the interpretation of any constant symbol of $\mathcal{A}_{\omega_1,\omega_1}$ c is $c^M = c$. Next the binary relation, R_i, and the unary relation, P_k, symbols have interpretations defined as:

$$
\begin{aligned}
R_i^M &= \{(c_1, c_2) : R_i(c_1, c_2) \in T\} \\
P_k^M &= \{c_1 : P_k(c_1) \in T\}
\end{aligned}
$$

This completes the description of the structure $M = \left\langle \mathcal{C}/\sim, \{P_k^M\}_{k\geq 1}, \{R_i^M\}_{i\in N} \right\rangle$.

We need to prove that for all $\mathcal{A}_{\omega_1,\omega_1}$–formulas $\varphi(v_1, ..., v_n)$ and $c_1, c_2, ..., c_n \in \mathcal{C}$ $M \models \varphi(c_1, c_2, ..., c_n)$ if and only if $\varphi(c_1, c_2, ..., c_n) \in T$.

By induction on formulas we start from atomic formula: take $R_i(v_1, v_2)$, the binary relation with free variables v_1, v_2, if $M \models R_i(v_1, v_2)$ then by the witness property there exists $R_i(c_1, c_2) \in T$. The same reasoning applies for P_k.

Suppose that the claim holds for an arbitrary $\mathcal{A}_{\omega_1,\omega_1}$–atomic formulas $\varphi(v_1, v_2)$ and $c_1, c_2 \in \mathcal{C}$, if $M \models \neg\varphi(c_1, c_2)$ then $M \not\models \varphi(c_1, c_2)$. By the induction hypothesis $\varphi(c_1, c_2) \notin T$ and $\neg\varphi(c_1, c_2) \in T$, because T is maximal.

Conversely if $\neg\varphi(c_1, c_2) \in T$ then $\varphi(c_1, c_2) \notin T$, because T is finitely satisfiable. Thus, by inductive assumption $M \not\models \varphi(c_1, c_2)$ and $\mathcal{M} \models \neg\varphi(c_1, c_2)$. The same way of reasoning applies for $\wedge$ and countable conjunction of formulas in $\mathcal{A}_{\omega_1,\omega_1}$, thanks to the completeness of $\mathcal{L}_{\omega_1,\omega}$.

Hence we have a canonical model for our theory. ■

Using the language $\mathcal{L}_{\omega_1,\omega_1}$, we can not start from any theory and extend it to a theory that has the witness property and then build a model for that theory, because the language $\mathcal{L}_{\omega_1,\omega_1}$ is not compact. Therefore, we can not conclude that if all the finite subsets of a theory have a model then the full theory itself has a model.

If we want a compact language we should take a fragment or a subset of $\mathcal{F}_{\omega_1,\omega}$, but this precludes the possibility of defining explicitly for example common knowledge, because we should exclude formulas with countable quantification. Note that the canonical model is uncountable because the set of $\mathcal{L}_{\omega_1,\omega_1}$-formulas is uncountable and $\mathcal{A}_{\omega_1,\omega_1}$ contains it.

We consider the $\mathcal{A}_{\omega_1,\omega_1}$-theory, $\mathbf{T}_+^{\mathbf{rst}}$, which is obtained from $\mathbf{T}^{\mathbf{rst}}$ adding axioms (Epistemic Continuity) and (Epistemic Monotonicity). $\mathbf{T}_+^{\mathbf{rst}}$ is finite and admits a model; we need to consider a maximal finitely satisfiable $\mathcal{A}_{\omega_1,\omega_1}$-theory that includes $\mathbf{T}_+^{\mathbf{rst}}$, in symbols $\mathbf{T}_+^{\mathbf{rst}\prime} \supseteq \mathbf{T}_+^{\mathbf{rst}}$.

Let I be the set of all finitely satisfiable $\mathcal{A}_{\omega_1,\omega_1}$-theories containing $\mathbf{T}_+^{\mathbf{rst}}$. If $C \subseteq I$ is a chain, ordered by inclusion, of sets of $\mathcal{A}_{\omega_1,\omega_1}$-sentences, then we can define the set of sentences contained in the chain as $T_C := \bigcup\{\Sigma : \Sigma \in C\}$. If Δ is a finite subset of T_C then there is $\Sigma \in C$ such that $\Delta \subseteq \Sigma$. Hence T_C is a finitely satisfiable and $T_C \supseteq \Sigma$ for all $\Sigma \in C$.
Thus, every chain in I has an upper bound and by the Zorn Lemma there exists $\mathbf{T}_+^{\mathbf{rst}\prime}$ which is maximal with respect the partial order of inclusion.

Given $\mathbf{T}_+^{\mathbf{rst}\prime}$, which is finitely satisfiable, and a sentence ϕ in $\mathcal{A}_{\omega_1,\omega_1}$, either $\mathbf{T}_+^{\mathbf{rst}\prime} \cup \{\phi\}$ or $\mathbf{T}_+^{\mathbf{rst}\prime} \cup \{\neg\phi\}$ is finitely satisfiable. Suppose $\mathbf{T}_+^{\mathbf{rst}\prime} \cup \{\phi\}$ is not finitely satisfiable. Then, there is a finite $\Delta \subseteq \mathbf{T}_+^{\mathbf{rst}\prime}$ such that $\neg\phi$ is a logical consequence of Δ.
For any finite subset of $\mathbf{T}_+^{\mathbf{rst}\prime}$, $\Sigma \subseteq \mathbf{T}_+^{\mathbf{rst}\prime}$, $\Sigma \cup \Delta \subseteq \mathbf{T}_+^{\mathbf{rst}\prime}$ hence $\Sigma \cup \Delta$ is finitely satisfiable. Moreover, if $\neg\phi$ is a logical consequence of the theory obtained by the union of Σ and Δ, $\Sigma \cup \Delta$, then $\Sigma \cup \{\neg\phi\}$ is satisfiable. Thus, $\mathbf{T}_+^{\mathbf{rst}\prime} \cup \{\neg\phi\}$ is finitely satisfiable.

Therefore, for any sentence ϕ, either $\mathbf{T}_+^{\mathbf{rst}\prime} \cup \{\phi\}$ or $\mathbf{T}_+^{\mathbf{rst}\prime} \cup \{\neg\phi\}$ is finitely satisfiable. Hence, we have found a maximal theory that contains our original theory, $\mathbf{T}_+^{\mathbf{rst}}$. By Lemma 6 we can conclude that the canonical space exists. Moreover, the canonical model is also a universal type space because Theorem 3 does not hold in $\mathcal{L}_{\omega_1,\omega_1}$. The obtained universal type space is unique because the Theorem 3 does not apply. This proves Theorem 2.

Bibliography

[1] K. Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, 4:328–346, 1950.

[2] R. J. Aumann. Interactive epistemology i: Knowledge. *International Journal of Game Theory*, 28(3):263–300, 1999.

[3] P. Battigalli. Rationalizability in infinite, dynamic games with incomplete information. *Research in Economics*, 57(1):1–38, 2003. TY - JOUR.

[4] P. Battigalli. Rationalization in signaling games: Theory and applications. *mimeo*, 2003.

[5] P. Battigalli and M. Siniscalchi. Hierarchies of conditional beliefs and interactive epistemology in dynamic games. *Journal of Economic Theory*, 88(1):188–230, 1999.

[6] P. Battigalli and M. Siniscalchi. Strong belief and forward induction reasoning. *Journal of Economic Theory*, 106(2):356–391, 2002. TY - JOUR.

[7] P. Battigalli and M. Siniscalchi. Rationalization and incomplete information. *Advances in Theoretical Economics*, 3, 2003.

[8] J. P. Benoit and J. Dubra. Information revelation in auctions. *mimeo*, 2004.

[9] A. Brandenburger and E. Dekel. Rationalizability and correlated equilibria. *Econometrica*, 55:1391–1402, 1987.

[10] J. Broome. Utilitarianism and expected utility. *The journal of Philosophy*, 84:405–422, 1987.

[11] G. Cappelletti. On the equivalence between semantic and syntactic approach. 2003.

[12] G. G. L. Cappelletti. A note on the characterization of strong d-rationalizability in static and extensive form games. 2004.

[13] C. Chang and H. J. Keisler. *Model Theory*. North Holland, 1998.

[14] W. B. Charles Blackorby and D. Donaldson. HarsanyiŠs social aggregation theorem: A multi-profile approach with variable-population extensions. *mimeo*, 2003.

[15] V. Crawford. Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Economic Review*, 93(1):133–149, 2003.

[16] V. P. Crawford and J. Sobel. Strategic information transmission. *Econometrica: Journal of the Econometric Society*, 50(6):1431–1451, 1982.

[17] C. D'Aspremont and L. Gevers. Equity and the infromational basis of collective choice. *Review of Economic Studies*, 44:199–209, 1977.

[18] R. M. Dudley. *Real Analysis and Probability*. Cambridge Press, 2002.

[19] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about knowledge*. MIT Press, 1995.

[20] J. Farrell. Meaning and credibility in cheap-talk games. 5, 1993.

[21] C. Genest and J. V. Zidek. Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 1:114–135, 1986.

[22] P. R. Halmos. *Measure Theory*. Springer Verlag, 1974.

[23] P. J. Hammond. Equity in two person situations: Some consequences. *Econometrica*, 47:1127–1136, 1979.

[24] P. J. Hammond. Ex-ante and ex-post welfare optimality under uncertainty. *Economica*, 48:235–250, 1981.

[25] P. J. Hammond. Equity, arrow's conditions, and rawls' difference principle. *Econometrica*, 44:793–804, 76.

[26] J. Harsanyi. Cardinal welfare, individualistic ethics, and interpersonal comparison of utility. *Journal of Political Economics*, 63:309–321, 1955.

[27] A. Heifetz and D. Samet. Knowledge spaces with arbitrarily high rank. *Games and Economic Behavior*, 22(2):260–273, 1998. TY - JOUR.

[28] A. Heifetz and D. Samet. Topology-free typology of beliefs. *Journal of Economic Theory*, 82(2):324–341, 1998. TY - JOUR.

[29] A. Heifetz and D. Samet. Coherent beliefs are not always types. *Journal of Mathematical Economics*, 32(4):475–488, 1999. TY - JOUR.

[30] A. Heifetz and D. Samet. Hierarchies of knowledge: An unbounded stairway. *Mathematical Social Sciences*, 38(2):157–170, 1999. TY - JOUR.

[31] T.-W. Hu. General rationalizability and its robustness for strategic form games with incomplete information. 2004. Mimeo.

[32] A. Hylland and R. Zeckhauser. The impossibility of bayesian group decision making with separate aggregation of beliefs and values. *Econometrica*, 47:1321–1336, 1979.

[33] D. S. I. Gilboa and D. Schmeidler. Utilitarian aggregation of beliefs and tastes. *mimeo*, March 2001.

[34] M. e. A. Kaneko. A map of common knowledge logics. *Studia Logica*, 71:57–86, 2002.

[35] H. J. Keisler. *Model Theory for Infinitary Logic*. North Holland, 1991.

[36] M. C. Kemp and Y. Ng. On the existence of social welfare functions, social orderings and social decision functions. *Economica*, 43:59–66, 1976.

[37] V. Krishna and J. Morgan. The art of conversation: Eliciting information from experts through multi-stage communication. *mimeo*, 2003.

[38] D. Marker. *Model Theory: An Introduction*. Springer, 2002.

[39] J. Mertens and S. Zamir. Formulation of bayesian analysus for games with incomplete information. *International Journal of Game Theory*, 14:1–29, 1985.

[40] B. D. Meyer and P. Mongin. A note on affine aggregation. *Economics Letters*, 47:177–183, 1995.

[41] P. Mongin. Harsanyi's aggregation theorem: Multiple-profile version and unsettled questions. *Social Choice and Welfare*, 11:331–354, 1994.

[42] P. Mongin. Consistent bayesian aggregation. *Journal of Economic Theory*, 66:313–351, 1995.

[43] J. Nash. Non-cooperative game. *The Annals of Mathematics*, 54:286–295, 1951.

[44] M. Okuno-Fujiwara, A. Postlewaite, and K. Suzumura. Strategic information revelation. *The Review of Economic Studies*, 57(1):25–47, 1990. FLA 00346527 Society for Economic Analysis Limited Copyright 1990 The Review of Economic Studies Ltd.

[45] R. P. Parks. An impossiblility theorem for fixed preference: A dictatorial bergson-samuelson welfare function. *The Review of Economic Studies*, 43:447–4450, 1976.

[46] P. Pollak. Bergson - samuelson social welfare functions and the theory of social choice. *The Quarterly journal of Economics*, 93:73–90, 1979.

[47] M. Rabin. Communication between rational agents. *Journal of Economic Theory*, 51(1):144–170, 1990.

[48] H. Raiffa. *Decision Analysis.* Addion-Wesley, 1970.

[49] A. Rubinstein. The single profile analogues to multi profile theorems: Mathematical logic's approach. *International Economic Review*, 25:719–730, 1984.

[50] L. Savage. *The Foundation of Statistics.* Dover Publications, 1954.

[51] A. Sen. *Collective Choice and Social Welfare.* Holden Day, 1974.

[52] M. Shimoji and J. Watson. Conditional dominance, rationalizability, and game forms. *Journal of Economic Theory*, 83:161 195, 1998.

[53] J. B. K. T. Seidenfeld and M. Schervish. On the shared preferences of two bayesian decision makers. *The Journal of Philosophy*, 86:225–244, 1989.

[54] T. C. Tan and S. R. D. C. Werlang. The bayesian foundations of solution concepts of games. *Journal of Economic Theory*, (45), 1988.

Chapter 2

On the Equivalence between Semantic and Syntactic Approach

On the Equivalence between Semantic and Syntactic Approach

Giuseppe G. L. Cappelletti
Bocconi University

January 2nd, 2005

Abstract

From the Harsanyi's seminal contribution the analysis of game with incomplete information relies on the so-called Universal Type Space. The construction of the Universal Type Space has been pursued by using two different approaches, Semantic and Syntactic. This essay investigates whether Semantic and Syntactic approach are equivalent. We take a purely linguistic perspective where Semantic and Syntactic approach are two possible interpretation of a formal language namely, first order logic which is able to fully describe players' epystemic characteristics.

2.1 Introduction

Harsanyi (1967-68) proposed a new theory for the analysis of games with incomplete information, i.e. games where the relationship between actions and payoffs is not commonly known by the agents. Naturally, the analysis of this kind of games gives rise to an infinite regress in reciprocal expectation. For example, a player's strategy choice will depend on what he expects to be his opponents' preferences. This expectation is called *first-order* expectation. But his choice will also depend on what he expects to be his opponents' expectation about his own preferences. This is called *second-order* expectation. Indeed, player's strategy choice will also depend on what he expects to be his opponents' *second-order* expectation.[1] This is called *third-order* expectation and we can proceed *ad infinitum.*[2]
Harsanyi encodes both player's preferences and his first-order and higher-order beliefs concerning opponents in the so called player's type.[3] Given any game, the set of all possible types of each player is defined the universal type space. The universal type space is the basis for a more manageable framework to investigate interactive interaction with incomplete information. Moreover, the existence the a universal type space justifies the use of a Bayesian approach to model these situations (see Tan and Werlag (1988) [54]).

The existence of a universal type space has been proved in several papers under different assumptions (see Mertens and Zamir (1985) [39], Brandenburger and Dekel (1987)[9] and Heifetz and Samet (1998) [28] among others). If there is not such space we should stick to the explicit description of agents' hierarchical beliefs. But, also some negative results appeared (refer to Brandenburger and Keisler (2003), Heifetz and Samet (1998) [29]). The main issues raised were twofold. On the one hand there are measure theoretic problems usually studied using the semantic approach (refer to Heifetz and Samet (1999) [29]). On the other hand there are purely logical ones. In this category, we mention Brandenburger and Keisler (2003), who proved an impossibility theorem for a specific type space, namely a complete possibility structure.

[1] That is, what the player thinks that his opponents think that he thinks about their own preferences.

[2] In the following section it will become clear how tricky and essential is the expression "*ad infinitum*" in the construction of a universal type space.

[3] In the original paper Harsanyi used the expression "*information vector*". In order to distinguish payoff or preference related aspects (also called payoff-type) from beliefs related ones (also called epystemic-type) the plain expression "Harsanyi-type" has been used.

We have remarked that all the previous results started from two different[4] perspectives: a purely set theoretic approach, also called semantic approach, and a purely logical or linguistic approach, also called syntactic approach.
In the so-called semantic approach, the epistemic characteristics of the players are represented by a structure consisting of a set Ω of states of the world, together with a partition Π_i of Ω, for each player i. The cell of Π_i containing an element ω of the set Ω, represents player i knowledge in state ω. This approach is purely set theoretic.
The syntactic approach hinges on a purely logic construction. This construction is constituted by a set of propositions built with a formal language, which is expressive enough to describe the epistemic properties of the players. Logical relations between the various propositions are described by formal rules of inference.

The equivalence between the semantic and syntactic approaches has been studied by Aumann ([2]), Fagin et al. ([19]), who analyzed it by means of modal logic (Chellas (1998)), and Kaneko et al. ([34]), who were interested in the broader issue of the characterization of the concept of Common Knowledge in various logics.

In our paper we will show that twp approaches are equivalent proveing that they are just two iterpretation of a more general formal language namely, first order logic. As a preliminary step we will choose the formal language.[5] In economics the standard choice has been epistemic logic[6], instead we prefer to use first order logic. Our choice can be justified with two motives: first, we can rely on the results obtained in mathematical logic[7], and specifically model theory[8]; second, there has been a huge debate on whether modal logic is a completely reliable tool (refer to Quine (1964)[9]).
Moreover, first-order logic will prove to be a more flexible and natural language for describing the interactive contexts. By means of first order logic, we can explicitly define arbitrarily high order of reasoning and common knowledge

[4] Quite opposite.

[5] This language could be used by the players for communication or by the economists to analyze their strategic interaction.

[6] As a a matter of fact, epistemic logic has been widely used in economics for dealing with the rationales of knowledge and beliefs when multiple agents interact.

[7] We mention that the distinction between expression and their intended meanings is the very starting point in logic and not by chance this fact is stated as syntax versus semantics of a language.

[8] Model theory is a branch of mathematical logic that considering which mathematical models satisfy a certain theory.

[9] Quine, "The Problem of Interpreting Modal Logic" (1947)

without any hidden assumption.
This choice will be crucial for stating the equivalence between the two approaches that are commonly used in the economic literature for modelling games with incomplete information.[10] In order to establish this equivalence we need to build a so-called canonical model, which will be the natural candidate as Universal Type Space.

The essay is organized as follows. Section 2 introduces the syntactic and semantic approaches and then builds a formal language, which is equivalent to the syntactic approach. In Section 3, allowing only for a finite order of sophisticated reasoning, we prove the equivalence between the semantic and syntactic approach.

2.2 The framework

> "There are some obvious correspondences between the two approaches. Formulas (syntactic) correspond to events (semantic);...By "correspond", we mean "have similar substantive content. ... But substantively, they express the same thing".
>
> "This paper examines the relation between the two approaches, and shows that they are in a sense equivalent.". [Aumann (1998)] [2]

In the syntactic approach the players' reasoning is described explicitly starting from a formal language, which is fitted for describing exhaustively the epistemic characteristics of agents. Instead, in the semantic approach this description is obtained through a set-theoretic construction.[11]
In the economic literature modal logic has been used as the formal language for describing interactive situations. We will describe briefly how modal logic is used and then we will introduce a so-called first order logic and prove that it is equivalent to the former formal language.

[10] Moreover, we clearly separate the two types of analysis (semantic and syntactic), in order to disentangle purely logic issues (see Brandenburger and Keisler (2003)) from purely semantic or set-theoretic ones (see Heifetz and Samet (1998) [29]) and we find conditions, in purely linguistic terms, that exclude both.

[11] If we consider the semantic representation just as one possible interpretation of the expressions built with a formal language, the problem of the equivalence between syntactic and semantic approach is equivalent to the uniqueness of the interpretation of a formal language. Consequently, we will focus on the problem of interpretation of a language and we will borrow techniques from mathematical logic and model theory.

Starting from a set N of agents, labelled by $1, 2, ..., n$, we assume that they want to reason about an external setting, which can be described by a nonempty countable set of propositions, $\Phi = \{p_1, p_2, ...\}$. Usually these propositions describe basic facts about the external world, for example "it is snowing in Milan", "the price of oil reaches 20$" or "Player j chooses action a". To express statements like "Player i knows p", where p is an arbitrary element of Φ, we need to add the modal or epistemic operators $\{K_i\}_{i=1,2,...,n}$, one for each player, to our vocabulary. For example $K_3 p$ should be read as "Player 3 knows p".

From primitive propositions belonging to Φ we allow to build more complex formulas by taking negation (in symbol $\neg$), conjunction (in symbol $\wedge$), disjunction (in symbol $\vee$)[12] and composition with respect to epistemic operators. Thus, if p_1 and p_2 are primitive propositions, $\neg p_1$, $p_1 \wedge p_2$ and $K_i p_1$ are formulas of our language.[13] Inductively apply the previous rule of constructing more and more complex expression we can define the set of well-formed formulas, denoted as $\mathcal{L}_n(\Phi)$ or for short $\mathcal{L}$. Note that we will allow to form only finite sequences of symbols. In such a manner we have described the syntax of modal logic.

Instead of using modal logic to describe interactive reasoning of the players we will consider a first-order language, which will be proven to be as rich and expressive as modal logic (see [19]).

Formally a first-order language is given by specifying the following data:

1. a set of function symbols $\mathcal{F}$ and positive integers n_f for each $f \in \mathcal{F}$;
2. a set of relation symbols $\mathcal{R}$ and positive integers n_R for each $R \in \mathcal{R}$;
3. a set of constant symbols $\mathcal{C}$.

For our purpose, $\mathcal{R}$ is constituted by the countable set of unary relation symbols $\Phi^* \doteq \{P_1, P_2, ...\}$, where each element P_k corresponds to a primitive proposition p_k in Φ, and by the set of binary relation symbols $\{R_1, R_2, ..., R_n\}$, one for each player. Our language is purely relational and it is composed by $\langle \Phi^*, R_1, R_2, ..., R_n \rangle$.

Now, we need to define well-formed formulas in this formal language. A formula is any string of symbols built using symbols of the language, variables symbols $x_1, x_2..$, the Boolean connectives "and" (in symbol $\wedge$), "or" (in symbol $\vee$),

[12] Note that given any two primitive proposition p_1 and p_2 the expression $p_1 \vee p_2$ is equivalent to $\neg(\neg p_1 \wedge \neg p_2)$. Therefore, we can safely concentrate on the conjunction symbol.

[13] Moreover, we will use the following abbreviations: *True* for $p \vee \neg p$, *False* for $p \wedge \neg p$ and $\phi \Rightarrow \varphi$ for $\neg\phi \vee \varphi$.

"not" (in symbol $\neg$) and the existence and universal quantifiers (in symbols $\exists$, $\forall$). Proceeding step by step we define atomic formulas and then formulas.[14]

Definition 19 *ϕ^* is an atomic formula if ϕ^* is either*

1. *$x_1 = x_2$ where x_1 and x_2 are variables or;*

2. *$R_i(x_1, x_2)$ or $P_k(x_3)$, where x_1, x_2, x_3 are variables.*

Definition 20 *The set of (well-formed) formulas is the smallest set $\mathcal{L}^*$ containing all atomic formulas and such that*

1. *if ϕ^* is in $\mathcal{L}^*$ then $\neg\phi^*$ is in $\mathcal{L}^*$,*

2. *if ϕ^* and λ^* are in $\mathcal{L}^*$ then $\phi^* \wedge \lambda^*$ is in $\mathcal{L}^*$, and*

3. *if ϕ^* is in $\mathcal{L}^*$ then $\exists x\phi^*$ is in $\mathcal{L}^*$.*

A variable x can occur freely or it can be bound by a quantifier. A formula with no free variable is called a *sentence*. We denote the set of formulas of our language as $\mathcal{L}^*$.

In order to prove that this language is as expressive as modal logic, we need to construct a translation between the two languages. For each formula, $\phi \in \mathcal{L}$, we define a corresponding first-order formula $\phi^* \in \mathcal{L}^*$.[15]

We start from primitive propositions:

- for every primitive proposition p_k in Φ, let $p_k^* := P_k(x)$ be the corresponding first order logic unary-relation symbol with free variable x;

and then we proceed with the translation, considering negation, conjunction and epistemic expression of primitive propositions:

- for every formula in $\mathcal{L}$ of the form $\neg p_k$, let $(\neg p_k)^* := \neg P_k(x)$ be the corresponding first order formula;
- for every formula in $\mathcal{L}$ of the form $p_k \wedge p_l$, let $(p_k \wedge p_l)^* := P_k(x_1) \wedge P_l(x_2)$ be the corresponding first order formula;

[14] Note that given any formulas ϕ^* and φ^* in $\mathcal{L}^*$, $\phi^* \vee \varphi^*$ is an abbreviation of $\neg(\neg\phi^* \wedge \neg\varphi^*)$. Moreover given any formula $\phi^*(x)$ with free variable x, the expression $\forall x\phi^*$ is an abbreviation of $\neg\forall x\neg\phi^*$. Therefore, wlog we can consider just the conjunction symbol and the existence quantifiers.

[15] We adopt the convention that greek letters denote formulas, while latin letters denote primitive expressions. Starred greek letters denote first order logic formulas.

- for every formula in $\mathcal{L}$ of the form $K_i(p_k)$, let $(K_i p_k)^* := \forall x_2 (R_i(x_1, x_2) \Rightarrow P_k(x_2))$ be the corresponding first order formula, where x_2 is a new variable not appearing in $P_k(x_1)$ and $P_k(x_2)$ is the result of replacing all occurrences of x_1 in $P_k(x_1)$ by x_2.

We can complete the translation using the recursive nature of well-formed formula in both formal languages. For example let $K_i(\phi)$ be a modal logic formula, where ϕ belongs to $\mathcal{L}$, its translation in $\mathcal{L}^*$ is $(K_i\phi)^*(x_1)$, that is, $\forall x_2 (R_i(x_1, x_2) \Rightarrow \phi^*(x_2))$. For notation convenience, we will drop the free variable notation whenever this is not relevant for the exposition.[16]

Referring to modal logic, we need a semantics i.e., a model that can be used to determine whether a formula in $\mathcal{L}$ is true or false. Any language may have many possible semantics with different true assignments. A commonly used tool is the so-called Kripke structure. A Kripke structure M for n agents over Φ is a tuple $(S, \pi, \mathcal{K}_1, ..., \mathcal{K}_n)$, where S is a set of states, π is an interpretation that associates to each state in S a truth assignment for the primitive propositions (i.e. $\pi(s) : \Phi \to \{True, False\}$ for each $s \in S$) and $\mathcal{K}_i$ is a binary relation on S, for each $i \in \{1, 2, ..., n\}$.[17] We denote the class of all Kripke structures for n agents over Φ as $\mathcal{M}_n(\Phi)$ or briefly $\mathcal{M}$. The truthfulness of a formula in $\mathcal{L}$ depends on the state as well as the structure, hence we focus on the notion of a formula φ to be true at (M, s) (in symbols $(M, s) \models \phi$). The relation $\models$ is defined by induction on the length of formulas of modal language, starting from the primitive propositions.

If ϕ is a primitive proposition, i.e. $\phi = p \in \Phi$:

$$(M, s) \models \phi \text{ if and only if } \pi(s)(p) = True.$$

If ϕ is a conjunction or negation of formulas, namely:

$$\begin{aligned} \text{for } \phi &= \varphi \wedge \theta, \ (M, s) \models \phi \text{ if and only if } \pi(s)(\varphi) = True \text{ and } \pi(s)(\theta) = True \\ \text{for } \phi &= \neg\varphi, \ (M, s) \models \phi \text{ if and only if } \pi(s)(\varphi) = False \end{aligned}$$

For modal expression such as $K_i p$ the definition is:

[16] For example we will often denote a formula ϕ^* without remarking whether there are free variables. Complete notations would be $\phi(x)$ where $x = \{x_1, ...x_n\}$ are n free variables.

[17] Any binary relation $\mathcal{K}_k$ can be identified by a subset of $S \times S$.

for $\phi = K_i\varphi$, $(M,s) \models \phi$ if and only if $(M,t) \models \varphi$ for all t such that $(s,t) \in \mathcal{K}_i$

Similarly, given a first order language, we need a semantic or $\mathcal{L}^*$-structure to give substance to the different formulas of $\mathcal{L}^*$.

Definition 21 *An $\mathcal{L}^*$-structure, $M^* = \left\langle U, \{P_k^M\}_{k=1}^K, \{R_i^M\}_{i=1}^n \right\rangle$, is given by the following elements:*

1. *a non-empty set U called universe of M^*;*
2. *a set $P_k^M \subseteq U$ for each unary relation P_k;*
3. *a set $R_i^M \subseteq U \times U$ for each binary relation R_i.*

If a language has constant symbols $c \in C$ then their interpretations are just elements c^M of U.

If ϕ^* is a formula in $\mathcal{L}^*$ with free variables $x_1, ..., x_n$ we can think of ϕ^* as expressing a property of elements of $U^n := U \times U \times ... \times U$, hence we must define what it means for $\phi^*(x_1, ..., x_n)$ to hold at $(a_1, ..., a_n) \in U^n$. Once more we will use the recursive nature of the well-formed formulas for a formal language.

Definition 22 *Let ϕ^* be a formula with free variables from $x = (x_1, ..., x_m)$ and let $\bar{a} = (a_1, ..., a_m) \in U^m$. We inductively define "M satisfies $\phi^*(\bar{a})$" or "$\phi^*(\bar{a})$ is true in M" (in symbols $M \models \phi^*[\bar{a}]$) as follows:*

1. *if ϕ^* is $R_i(x_1, x_2)$ then $M \models \phi^*[\bar{a}]$ if $\bar{a} = (a_1, a_2) \in R_i^M$;*
2. *if ϕ^* is $P_k(x)$ then $M \models \phi^*[\bar{a}]$ if $\bar{a} = (a) \in P_k^M$;*
3. *if ϕ^* is $(K_i p_k)^*$ then $M \models \phi[\bar{a}]$ if for every b, such that $(\bar{a}, b) \in R_i^M$, $M \models P_k[b]$;*[18]
4. *if ϕ^* is $\neg\psi^*$ then $M \models \phi^*[\bar{a}]$ if $M \not\models \psi^*[\bar{a}]$;*
5. *if ϕ^* is $(\psi^* \wedge \varphi^*)$ then $M \models \phi^*[\bar{a}]$ if $M \models \psi^*[\bar{a}]$ and $M \models \varphi^*[\bar{a}]$;*
6. *if ϕ^* is $\exists x \psi^*(\bar{x}, x)$, then $M \models \phi^*[\bar{a}]$ if there is $b \in U$ such that $M \models \psi^*[\bar{a}, b]$ holds.*

[18] Let us recall that $(K_i p_k)^* = \forall x_2 R_i(x_1, x_2) \Rightarrow P_k(x_2)$

7. *if* ϕ^* *is* $\forall x \psi^* (\bar{x}, x)$, *then* $M \models \phi^* [\bar{a}]$ *if for every* $b \in U$, $M \models \psi^* [\bar{a}, b]$ *holds.*

Next, we consider the mapping, proposed by Fagin et al. (1995) [19], from the set of Kripke structures, $\mathcal{M}$, to the set of $\mathcal{L}^*$-structures. This provides us with an equivalent $\mathcal{L}^*$-structure, M^*, for each Kripke structure $M \in \mathcal{M}$.

Given a Kripke structure $M = (S, \pi, \mathcal{K}_1, ..., \mathcal{K}_n)$ the corresponding $\mathcal{L}^*$-structure, M^*, has a *universe* S and for each primitive proposition $p_k \in \Phi$, we define the interpretation of the corresponding proposition $p_k^* \in \Phi^*$ to be $P_k^{M^*} = \{s \in S : \pi(s)(p_k) = True\}$, and the interpretation of any binary relations to be $R_i^{M^*} = \mathcal{K}_i \subseteq S \times S$.

If a primitive proposition p_k holds at some subset $S_{p_k} \subseteq S$ then the corresponding first-order expression, p_k^*, has interpretation S_{p_k}. Similarly given $\mathcal{K}_i \subseteq S \times S$, which is the semantic counterpart of the modal knowledge operator, the corresponding first order expression R_i has interpretation $\mathcal{K}_i$.

Now, we need to prove that this $\mathcal{L}^*$-structure, M^*, is equivalent to the Kripke structure in terms of truth-values. Namely, if a modal logic formula holds in a given the Kripke structure M, then its translation in the first-order language holds in the corresponding $\mathcal{L}^*$-structure, M^*.

Proposition 5 *For each formula in modal logic* ϕ, $(M, s) \models \phi$ *if and only if* $M^* \models \phi^* [s]$ *where* ϕ^* *is the corresponding first-order language formula and* $s \in U^{n_\varphi}$.[19]

Proof. See Appendix ?? ■

Once, we have defined the language $\mathcal{L}^*$ and we have established this "valid" translation, we can consider a set of sentences that describe some properties of the epistemic operator. Given a purely syntactic object as a theory we can investigate the class of $\mathcal{L}^*$-structures, that satisfy all the sentences of the theory. For example, given the sentence $\forall x \phi^* (x)$, where $\phi^* (x)$ is a formula with just one free variable, and the $\mathcal{L}^*$-structure $M^* = \langle U, P^M, R^M \rangle$, if all the elements of U satisfy the formula $\phi^* (x)$, in symbols $M^* \models \phi^* [u]$, then $\langle U, P^M, R^M \rangle$ is a model for the theory consisting of the single sentence $\forall x \phi^* (x)$ (in symbols $M^* \models \forall x \phi^* (x)$).

We define any set of sentences a $\mathcal{L}^*$-theory, denoted as $\mathbf{T}$.

Definition 23 *A* $\mathcal{L}^*$*-structure,* M^*, *is a model of* $\mathbf{T}$ *(in symbols* $M^* \models \mathbf{T}$*) if and only if* $M^* \models \sigma^*$ *for all sentences* σ^* *belonging to* $\mathbf{T}$.

[19] n_φ is the number of free variables in φ^*

We will focus on a particular theory, denoted $\mathbf{T}^{\mathbf{rst}}$, which corresponds to the Knowledge operator[20] in modal logic. The five axioms that characterize the epistemic operator and constitute our theory are the following $\mathcal{L}^*$-sentences:

$$\forall x_1 \left(K_i\varphi\right)^* (x_1) \wedge \left(K_i \left(\varphi \Rightarrow \psi\right)\right)^* (x_1) \Rightarrow \left(K_i\psi\right)^* (x_1) \quad \text{(Distribution Axiom)}$$

[21]

From ϕ^* infer $\left(K_i\phi\right)^*$: (Knowledge Generalization Rule)

$$\forall x_1^* \phi (x_1) \Rightarrow \forall x_1 \left(K_i\phi\right)^* (x_1)$$

[22]

$$\forall x_1 \left(K_i\phi\right)^* (x_1) \Rightarrow \phi^* (x_1) \quad \text{(Truth Axiom)}$$

[23]

$$\forall x_1 \left(K_i\phi\right)^* (x_1) \Rightarrow \left(K_iK_i\phi\right)^* (x_1) \quad \text{(Positive Introspection)}$$

[24]

$$\forall x_1 \left(\neg K_i\phi\right)^* (x_1) \Rightarrow \left(K_i\neg K_i\phi\right)^* (x_1) \quad \text{(Negative Introspection)}$$

---

[20] Note that we restrict ourselves to $S5$ system. The same way of reasoning applies to different axiomatization of the knowledge operator.

[21] The plain expression is:

$$\forall x_2 \left\{\left[\forall x_1 \left(R_i (x_1, x_2) \Rightarrow \varphi^* (x_2)\right) \wedge \forall x_1 \left(R_i (x_1, x_2) \Rightarrow \left(\varphi^* (x_2) \Rightarrow \psi^* (x_2)\right)\right)\right] \Rightarrow \forall x_1 \left(R_i (x_1, x_2) \Rightarrow \psi^* (x_2)\right)\right\}$$

[22] The plain expression is:

$$\forall x_1 \phi^* (x_1) \Rightarrow \forall x_1 \left[\forall x_2 \left(R_i (x_1, x_2) \Rightarrow \phi^* (x_2)\right)\right]$$

[23] The plain expression is :

$$\forall x_1 \left[\forall x_2 \left(R_i (x_1, x_2) \Rightarrow \phi^* (x_2)\right)\right] \Rightarrow \phi^* (x_1)$$

[24] The plain expression is:

$$\forall x_1 \left[\forall x_2 \left(R_i (x_1, x_2) \Rightarrow \phi^* (x_2)\right)\right] \Rightarrow \left[\forall x_3 \left(R_i (x_3, x) \Rightarrow \left(K_i\phi\right)^* (x_3)\right)\right]$$

$$\forall x_1 \left[\forall x_2 \left(R_i (x_1, x_2) \Rightarrow \phi^* (x_2)\right)\right] \Rightarrow \left\{\forall x_3 R_i (x_3, x_1) \Rightarrow \left[\forall x_2 \left(R_i (x_3, x_2) \Rightarrow \phi^* (x_2)\right)\right]\right\}$$

[25]

$$\forall x_1 \left[\varphi^* (x_1) \wedge (\varphi \Rightarrow \psi)^* (x_1) \Rightarrow \psi^* (x_1)\right] \qquad \text{(Modus Ponens)}$$

[26]

Following the notation of Fagin et al. (1995) [19], we denote with $\mathcal{M}^{rst}$ the class of Kripke structures $(S, \pi, \mathcal{K}_1, ..., \mathcal{K}_n)$ that satisfy the $S5$ axioms. For our purpose, we just need to recall that, given the $S5$ axioms, K_i is a reflexive, transitive and symmetric binary relation (hence the superscript rst is justified). We want to make sure that any Kripke structure is a model of the theory $\mathbf{T^{rst}}$. Hence, we take a generic Kripke structure, M, and we prove that the corresponding $\mathcal{L}^*$-structure, M^*, is a model for the $\mathbf{T^{rst}}$.

Proposition 6 *Given the Kripke structure $M_{rst} \in \mathcal{M}^{rst}$ then the equivalent $\mathcal{L}^*$-structure, M^*_{rst}, is a model for $\mathbf{T^{rst}}$.*

Proof. See Appendix ?? ■

A modal logic formula, φ, is true for any Kripke structure if and only if it is true at every world, $s \in S$ of every Kripke structure $M = (S, \pi, \mathcal{K}_1, ..., \mathcal{K}_n) \in \mathcal{M}^{rst}$, we denote this property as $\mathcal{M}^{rst} \models \varphi$. Similarly, in a first order language a $\mathcal{L}^*$-sentence, φ^*, is a logical consequence of a set of sentences, $\mathbf{T}$, if and only if it is true in any model of $\mathbf{T}$. We denote this fact as $\mathbf{T} \models \varphi^*$.

If a modal logic formula is true in all Kripke structures satisfying the $S5$ axioms, then the corresponding first-order formula should be a logical consequence of the theory $\mathbf{T^{rst}}$. We state this fact as a proposition.

Proposition 7 *$\mathcal{M}^{rst} \models \varphi$ iff $\forall x \varphi^* (x)$ is a logical consequence of $\mathbf{T^{rst}}$, in symbols $\mathbf{T^{rst}} \models \varphi^*$*

Proof. This will be a corollary of Proposition 8. ■

---

[25] The plain expression is :

$$\forall x_1 \neg \left[\forall x_2 \left(R_i (x_1, x_2) \Rightarrow \phi^* (x_2)\right)\right] \Rightarrow \left[\forall x_3 \left(R_i (x_3, x_1) \Rightarrow \neg (K_i \phi)^* (x_3)\right)\right]$$
$$\forall x_1 \neg \left[\forall x_2 \left(R_i (x_1, x_2) \Rightarrow \phi^* (x_2)\right)\right] \Rightarrow \left\{\forall x_3 R_i (x_3, x_1) \Rightarrow \neg \left[\forall x_2 \left(R_i (x_3, x_2) \Rightarrow \phi^* (x_2)\right)\right]\right\}$$

[26] The plain expression is:

$$\forall x_1 \left[\varphi^* (x_1) \wedge \varphi^* (x_1) \Rightarrow \psi^* (x_1)\right] \Rightarrow \psi^* (x_1)$$

2.3 Semantic approach and Syntactic approach as $\mathcal{L}^*$-structures

We identify the *semantic approach* with the set of $\mathcal{L}^*$-structure, M^*_{rst}, corresponding to some Kripke structure $M_{rst} \in \mathcal{M}^{rst}$, and we denote this class as $\mathcal{M}^*_{rst}(\Phi)$ or shortly $\mathcal{M}^*_{rst}$.

While, the syntactic approach relies on a formal language and by logic deduction derive implication of the basic axioms. We need to determine whether the *syntactic approach* can be identified in term of $\mathcal{L}^*$-structures. This is possible using the so-called Henkin construction (see [13] or [38]).

2.3.1 The Henkin construction and the Canonical Model

Given our first-order language $\mathcal{L}^*$ and the theory $\mathbf{T}^{\mathbf{rst}}$ we can consider a richer language $\mathcal{L}^*_1 \supseteq \mathcal{L}^*$, constructed adding a constant symbol for each $\mathcal{L}^*$-formula with one free variable. Namely, for a generic $\mathcal{L}^*$-formula, $\phi^*(x)$, there is an $\mathcal{L}^*_1 -$ *constant*, c_ϕ, such that the following implication holds in our theory: $\exists x \phi^*(x) \Rightarrow \phi^*(c_\varphi)$.[27] For every formula in $\mathcal{L}^*$ which is consistent with $\mathbf{T}^{\mathbf{rst}}$, we add a constant symbol which represents the possibility that the formula holds.

For example, let us take the formula $(K_i\varphi)^*(x)$, meaning that player i knows φ^*, then we consider a constant symbol $c_{K_i\varphi}$ such that the sentence $(K_i\varphi)^*(c_{K_i\varphi})$ is true. The intuition is that the constant witnesses the possibility that a player i could know φ^*. It is apparent that the Henkin construction resembles the universal type space built from expressions used, for example, by Heifetz and Samet (1998) [28].

Thus, the new language is $\mathcal{L}^*_1 := \mathcal{L}^* \cup \{c_\phi : \phi^*(x) \text{ an } \mathcal{L}^* - formula \text{ with one free variable}\}$. If for each $\mathcal{L}^*$-formula $\phi^*(x)$ we denote the corresponding $\mathcal{L}^*_1$-sentence $\Gamma^*_\phi := [\exists x \phi(x) \Rightarrow \phi(c_\phi)]$, then the new $\mathcal{L}^*_1$-theory is defined as $\mathbf{T}_1 := \mathbf{T}^{\mathbf{rst}} \cup \{\Gamma^*_\varphi : \phi^*(x) \text{ is an } \mathcal{L}^* - formul$

Remark 3 *Given that any $\mathcal{L}^*$-structure corresponding to some Kripke structure M_{rst} is a model of T^{rst}, any finite subset of sentences of T^{rst} has a model. Therefore any finite subset of $\mathbf{T}_1$ has a model.*

If a theory satisfies this properties it is said that the theory is "*finitely satisfiable*".

Definition 24 *An $\mathcal{L}^*$-theory T^* is finitely satisfiable if every finite subset of T^* admits a model or equivalently every finite subset of T^* is satisfiable.*

[27] In symbols, $\mathbf{T}^{\mathbf{rst}} \models \exists x \varphi^*(x) \Rightarrow \varphi^*(c_\varphi)$

This property has a major role in logic because it allows to prove that a not-finite theory has a model thanks to the compactness of many standard formal languages.

Now, we iteratively construct a sequence of languages, $\mathcal{L}^* \subseteq \mathcal{L}_1^* \subseteq \mathcal{L}_2^*, ...,$ and a sequence of finitely satisfiable $\mathcal{L}_i^*$-theories, $\mathbf{T}^{rst} \subseteq \mathbf{T}_1 \subseteq \mathbf{T}_2 \subseteq ....$ If $\phi^*(x)$ is a $\mathcal{L}_i^*$-formula, then there is a constant symbol $c_\phi \in \mathcal{L}_{i+1}^*$. c_ϕ is such that $\mathbf{T}_{i+1} \models (\exists x \phi^*(x)) \Rightarrow \phi^*(c_\phi)$.

We can define $\mathcal{L}_\infty^* := \bigcup_{i=1}^{\infty} \mathcal{L}_i^*$ and $\mathbf{T}_\infty^* := \bigcup_{i=1}^{\infty} \mathbf{T}_i$. If we take any $\mathcal{L}_\infty^*$-formula $\phi^*(x)$ with one free variable x then, by construction, there is a constant symbol $c_\phi \in \mathcal{L}_\infty^*$ such that, by construction, $\mathbf{T}_\infty^* \models (\exists x \phi^*(x) \Rightarrow \phi^*(c))$. This property is called the witness property of the theory.

Definition 25 *An $\mathcal{L}^*$-theory $\mathbf{T}^*$ has the* witness property *if whenever $\phi^*(x)$ is a $\mathcal{L}^*$-formula with one free variable x, then there is a constant symbol $c \in \mathcal{L}^*$ such that $\mathbf{T} \models (\exists x \phi^*(x) \Rightarrow \phi^*(c))$.*

If a formula is consistent with the theory then there is a constant that represents it. $\mathbf{T}_\infty^*$ has at most a countable number of sentences because the set of well-formed formula is at most countable. Moreover, any finite subset of $\mathbf{T}_\infty^*$ admits a model.

We would like that any sentence is either part of the considered theory or is incompatible with it.

Definition 26 *An $\mathcal{L}^*$-theory $\mathbf{T}^*$ is* maximal *if for any sentence ϕ^* either $\phi^* \in \mathbf{T}^*$ or $\neg\phi^* \in \mathbf{T}^*$.*

We can ensure that there is a finitely satisfiable $\mathcal{L}_\infty^*$-theory, $\mathbf{T}'_\infty \supseteq \mathbf{T}_\infty^*$ such that for every $\mathcal{L}_\infty^*$-sentence, ϕ^*, either $\phi^* \in \mathbf{T}'_\infty$ or $\neg\phi^* \in \mathbf{T}'_\infty$ i.e., there exists a maximal finitely satisfiable $\mathcal{L}_\infty^*$-theory.[28]

Once we have defined this fully encompassing language and theory, we can consider the so called "canonical model", denoted as $\mathcal{M}^{canonical}$, of $\mathbf{T}'_\infty$. This model is a $\mathcal{L}_\infty^*$-structure and its universe is $U := \mathcal{C}/\sim$, where $\mathcal{C}$ is the set of constant symbols of $\mathcal{L}_\infty^*$. For any $c, d \in \mathcal{C}$, $c \sim d$ if and only if $\mathbf{T}'_\infty \models c = d$.

[28] This is a consequence of the following theorem (refer to Marker (2002)[38]).

Theorem 5 *If $\mathbf{T}^*$ is a finitely satisfiable $\mathcal{L}^*$-theory then there is a maximal finitely satisfiable $\mathcal{L}^*$-theory $\mathbf{T}' \supseteq \mathbf{T}^*$.*

Thus, the universe is the set of equivalence classes of $\mathcal{C}$ with respect to $\sim$. The interpretation of the constant symbol $c \in \mathcal{C}$ is $c^{M^{canonical}} = \bar{c}$.[29]

The interpretation of the relations symbols of $\mathcal{L}^*_\infty$ are respectively:

$$\begin{aligned} P_k^{M'_\infty} &= \{\bar{c}_1 : P_k(\bar{c}_1) \in \mathbf{T}'_\infty\} \text{ for } k = 1, 2, \dots \\ R_i^{M'_\infty} &= \{(\bar{c}_1, \bar{c}_2) : R_i(\bar{c}_1, \bar{c}_2) \in \mathbf{T}'_\infty\} \text{ for } i = 1, 2, \dots, n \end{aligned}$$

$P_k^{M'_\infty}$ and $R_i^{M'_\infty}$ are well-defined.[30]

Therefore, we can define a $\mathcal{L}^*_\infty$-structure, called canonical model of $\mathbf{T}'_\infty$:

$$M^{canonical} := \left\langle U, \left\{P_k^{M'_\infty}\right\}_{k\geq 1}, \left\{R_i^{M'_\infty}\right\}_{i=1}^{n} \right\rangle.$$

By construction, the canonical model includes any conceivable combination of epistemic characteristics for the n players. Therefore, it is the natural candidate for being a universal type space. Now, we need to establish the equivalence the equivalence between the semantic and syntactic approaches. This will be investigated plainly in the following subsection.

2.3.2 Equivalence between Semantic approach and Syntactic approach

Given an $\mathcal{L}^*$-structures M^* we denote with $Th(M^*)$ the maximal set of sentences which are satisfied by M^*. Using this notation we can state formally the equivalence between semantic and syntactic approaches.

Proposition 8 $\bigwedge_{M_{rst} \in \mathcal{M}^*_{rst}(\Phi)} Th(M_{rst}) = Th\left(M^{canonical}\right)$

Proof. One direction ($\subseteq$) is obvious once we have noted that $M^{canonical} \in \mathcal{M}^*_{rst}(\Phi)$. Given that $Th(M_{rst}) \supseteq Th\left(M^{canonical}\right)$ and $M^{canonical}$ is consistent with respect to M_{rst}, let Σ be a $\mathcal{L}$-sentence, if $Th\left(M^{canonical}\right) \models \Sigma$ then $Th(M_{rst}) \models \Sigma$. ■

[29] Where c^* denotes the equivalence class of c.

[30] If $c^* \sim d^*$ then $P(c^*) \in \mathbf{T}'_\infty$ if and only if $P(d^*) \in \mathbf{T}'_\infty$. Note that if $c^* \sim d^*$ then $c^* = d^*$ belongs to $\mathbf{T}'_\infty$. Hence if $P(c^*) \in \mathbf{T}'_\infty$ then $P(d^*)$ is the logical consequence of $c^* = d^*$ and $P(c^*) \in \mathbf{T}'_\infty$. By maximality of $\mathbf{T}'_\infty$, it implies that $P(d^*) \in \mathbf{T}'_\infty$. A similar reasoning applies to the binary relations.

Instead of proving that $M^{canonical}$ is a model[31] for $\mathbf{T}'_\infty$ and $\mathbf{T}$ we have proved that the canonical model, namely the syntactic approach, and the set of $\mathcal{L}^*$-structures, corresponding to some Kripke structures, namely the semantic approach, have the same theory.

Proposition 8 recalls Aumann's result and the result of soundness and completeness in modal logic. But, there is one aspect that makes this proof more easy to interpret. If we think of two game theorists who try to use one of the two approaches for analyzing a strategic situation and they want to avoid any hidden assumption, then Proposition 8 tells us that the two approaches are equivalent only if we consider all the possible models of our initial theory. As a matter of fact, Proposition 8 points out that if we consider a specific Kripke structure we have different theories from $Th\left(M^{canonical}\right)$.[32]

Moreover, it is clear that the universe in the canonical model is just one of the possible interpretations. This implies that the interpretation of the language can not be taken for granted when we try to model a situation where players can communicate using language $\mathcal{L}^*$.[33]
Moreover, from Proposition 8 we can deduce that the assumption that players share the same interpretation of the language is crucial for all "Agree to disagree" results.

2.4 Conclusion

Using techniques of model theory[34], we show that the semantic and syntactic approaches are equivalent if we want to describe the epistemic environment up to a finite order of sophistication. In order to prove this equivalence we built a structure (a canonical model) closely related to the so called universal type space.

---

[31] In symbols $M^{canonical} \models \mathbf{T}_\infty$.

[32] A model is a complete description of a situation since any conceivable expression is either true or false. But, it is not reliable for deriving conclusion on the characteristics of our theory because it is too restrictive. Namely, it satisfies some axioms which are not part of the original theory.

[33] Using Aumann's words what is still problematic is the "*vocabulary of our players*".

[34] Aumann ([2]) already proved this by means of modal logic.

Bibliography

[1] K. Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, 4:328–346, 1950.

[2] R. J. Aumann. Interactive epistemology i: Knowledge. *International Journal of Game Theory*, 28(3):263–300, 1999.

[3] P. Battigalli. Rationalizability in infinite, dynamic games with incomplete information. *Research in Economics*, 57(1):1–38, 2003. TY - JOUR.

[4] P. Battigalli. Rationalization in signaling games: Theory and applications. *mimeo*, 2003.

[5] P. Battigalli and M. Siniscalchi. Hierarchies of conditional beliefs and interactive epistemology in dynamic games. *Journal of Economic Theory*, 88(1):188–230, 1999.

[6] P. Battigalli and M. Siniscalchi. Strong belief and forward induction reasoning. *Journal of Economic Theory*, 106(2):356–391, 2002. TY - JOUR.

[7] P. Battigalli and M. Siniscalchi. Rationalization and incomplete information. *Advances in Theoretical Economics*, 3, 2003.

[8] J. P. Benoit and J. Dubra. Information revelation in auctions. *mimeo*, 2004.

[9] A. Brandenburger and E. Dekel. Rationalizability and correlated equilibria. *Econometrica*, 55:1391–1402, 1987.

[10] J. Broome. Utilitarianism and expected utility. *The journal of Philosophy*, 84:405–422, 1987.

[11] G. Cappelletti. On the equivalence between semantic and syntactic approach. 2003.

[12] G. G. L. Cappelletti. A note on the characterization of strong d-rationalizability in static and extensive form games. 2004.

[13] C. Chang and H. J. Keisler. *Model Theory*. North Holland, 1998.

[14] W. B. Charles Blackorby and D. Donaldson. HarsanyiŠs social aggregation theorem: A multi-profile approach with variable-population extensions. *mimeo*, 2003.

[15] V. Crawford. Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Economic Review*, 93(1):133–149, 2003.

[16] V. P. Crawford and J. Sobel. Strategic information transmission. *Econometrica: Journal of the Econometric Society*, 50(6):1431–1451, 1982.

[17] C. D'Aspremont and L. Gevers. Equity and the infromational basis of collective choice. *Review of Economic Studies*, 44:199–209, 1977.

[18] R. M. Dudley. *Real Analysis and Probability*. Cambridge Press, 2002.

[19] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about knowledge*. MIT Press, 1995.

[20] J. Farrell. Meaning and credibility in cheap-talk games. 5, 1993.

[21] C. Genest and J. V. Zidek. Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 1:114–135, 1986.

[22] P. R. Halmos. *Measure Theory*. Springer Verlag, 1974.

[23] P. J. Hammond. Equity in two person situations: Some consequences. *Econometrica*, 47:1127–1136, 1979.

[24] P. J. Hammond. Ex-ante and ex-post welfare optimality under uncertainty. *Economica*, 48:235–250, 1981.

[25] P. J. Hammond. Equity, arrow's conditions, and rawls' difference principle. *Econometrica*, 44:793–804, 76.

[26] J. Harsanyi. Cardinal welfare, individualistic ethics, and interpersonal comparison of utility. *Journal of Political Economics*, 63:309–321, 1955.

[27] A. Heifetz and D. Samet. Knowledge spaces with arbitrarily high rank. *Games and Economic Behavior*, 22(2):260–273, 1998. TY - JOUR.

[28] A. Heifetz and D. Samet. Topology-free typology of beliefs. *Journal of Economic Theory*, 82(2):324–341, 1998. TY - JOUR.

[29] A. Heifetz and D. Samet. Coherent beliefs are not always types. *Journal of Mathematical Economics*, 32(4):475–488, 1999. TY - JOUR.

[30] A. Heifetz and D. Samet. Hierarchies of knowledge: An unbounded stairway. *Mathematical Social Sciences*, 38(2):157–170, 1999. TY - JOUR.

[31] T.-W. Hu. General rationalizability and its robustness for strategic form games with incomplete information. 2004. Mimeo.

[32] A. Hylland and R. Zeckhauser. The impossibility of bayesian group decision making with separate aggregation of beliefs and values. *Econometrica*, 47:1321–1336, 1979.

[33] D. S. I. Gilboa and D. Schmeidler. Utilitarian aggregation of beliefs and tastes. *mimeo*, March 2001.

[34] M. e. A. Kaneko. A map of common knowledge logics. *Studia Logica*, 71:57–86, 2002.

[35] H. J. Keisler. *Model Theory for Infinitary Logic*. North Holland, 1991.

[36] M. C. Kemp and Y. Ng. On the existence of social welfare functions, social orderings and social decision functions. *Economica*, 43:59–66, 1976.

[37] V. Krishna and J. Morgan. The art of conversation: Eliciting information from experts through multi-stage communication. *mimeo*, 2003.

[38] D. Marker. *Model Theory: An Introduction*. Springer, 2002.

[39] J. Mertens and S. Zamir. Formulation of bayesian analysus for games with incomplete information. *International Journal of Game Theory*, 14:1–29, 1985.

[40] B. D. Meyer and P. Mongin. A note on affine aggregation. *Economics Letters*, 47:177–183, 1995.

[41] P. Mongin. Harsanyi's aggregation theorem: Multiple-profile version and unsettled questions. *Social Choice and Welfare*, 11:331–354, 1994.

[42] P. Mongin. Consistent bayesian aggregation. *Journal of Economic Theory*, 66:313–351, 1995.

[43] J. Nash. Non-cooperative game. *The Annals of Mathematics*, 54:286–295, 1951.

[44] M. Okuno-Fujiwara, A. Postlewaite, and K. Suzumura. Strategic information revelation. *The Review of Economic Studies*, 57(1):25–47, 1990. FLA 00346527 Society for Economic Analysis Limited Copyright 1990 The Review of Economic Studies Ltd.

[45] R. P. Parks. An impossiblility theorem for fixed preference: A dictatorial bergson-samuelson welfare function. *The Review of Economic Studies*, 43:447–4450, 1976.

[46] P. Pollak. Bergson - samuelson social welfare functions and the theory of social choice. *The Quarterly journal of Economics*, 93:73–90, 1979.

[47] M. Rabin. Communication between rational agents. *Journal of Economic Theory*, 51(1):144–170, 1990.

[48] H. Raiffa. *Decision Analysis*. Addion-Wesley, 1970.

[49] A. Rubinstein. The single profile analogues to multi profile theorems: Mathematical logic's approach. *International Economic Review*, 25:719–730, 1984.

[50] L. Savage. *The Foundation of Statistics*. Dover Publications, 1954.

[51] A. Sen. *Collective Choice and Social Welfare*. Holden Day, 1974.

[52] M. Shimoji and J. Watson. Conditional dominance, rationalizability, and game forms. *Journal of Economic Theory*, 83:161 195, 1998.

[53] J. B. K. T. Seidenfeld and M. Schervish. On the shared preferences of two bayesian decision makers. *The Journal of Philosophy*, 86:225–244, 1989.

[54] T. C. Tan and S. R. D. C. Werlang. The bayesian foundations of solution concepts of games. *Journal of Economic Theory*, (45), 1988.

Part II

Implications of Rationality and Exogenous Restrictions on Players' beliefs

Chapter 3

Rational Disclosure in Games with Incomplete Information

Rational Disclosure in Games with Incomplete Information

Giuseppe G. L. Cappelletti[1]
Bocconi University

January 2nd, 2005

Abstract

This paper analyzes a model in which a informed agent sends a cheap-talk message to an uniformed party, subsequently, takes an action that determines the utility of both. We assume full rationality, a certain degree of alignment of interest and that agent have a propensity to believe others. Pursuing a fully fledged non-equilibrium analysis we point out sufficient and tight conditions for full disclosure to occur. We, also, apply our framework to previous

[1]I wish to thank Pierpaolo Battigalli, Sandro Brusco and Marco Mariotti for very useful comments.

results in the literature.

Keywords: Information Disclosure, Incomplete Information
JEL Classifications: C72, D82

3.1 Introduction

In our paper, we will model strategic revelation of private information by adding a cheap talk communication stage to an incomplete information game. The analysis of the resulting game is pursued by a fully fledged non-equilibrium analysis. Starting from assumptions on preferences and beliefs held by agents and without assuming any sort of coordination, we assess when and how disclosure can rationally occur and what are the outcomes compatible with rationality and "common knowledge" of rationality.

The first conclusion we can extract is that alignment of interests between agents and the incentives by the sender to reveal truthfully information are jointly necessary for disclosure to occur, but they are not sufficient. We need to restrict the conceivable beliefs which agents can hold.

We apply our framework to previous results on rational information disclosure in the literature. Namely, the early analysis of Mathew Rabin (1990) [47] on communication games with cheap talk. Rabin's paper has been the first attempt to identify the set of messages that rational agents can exchange. We proceed on this line of research relying on the recent advances in the analysis of the implication of rationality in extensive form game with incomplete information (see Battigalli and Siniscalchi (1999) [5]).

Many economic situations are characterized by the presence of asymmetric information and by the possibility of communication before agents take an irrevocable decision. To the extent that such communication is believed by the other agents, the informational setting changes depending on the result of the communication stage. Therefore, agents with private information would exchange messages in a strategic manner.

Strategic information transmission has been investigated in economics for a long time. In the signaling literature, for example, Spence (1973) showed that workers may undertake costly education to signal their productivity to potential employers, even if education does not improve workers' productivity. Fujiwara et al. (1990) [44] investigated when an informed party, allowed to send a certifiable message, can decide to fully reveal his private information.

There are situations where agents neither get educated nor undertake costly actions (i.e. "burning banknotes") but they simply talk. This kind of communication is referred to as cheap talk.

From a theoretical perspective two opposite directions have been pursued in the literature. On the one hand, adding a communication stage to a game could, in principle, enlarges the set of possible outcomes. Intuitively cheap talk can foster coordination among agents with respect to the original game considered in isolation (refer to Crawford and Sobel (1980) [16] and Aumann and Hart (2003)). On the other hand, cheap talk may refine the set of outcomes (see Matthews, Okuno-Fujiwara and Postlewaite (1990) [44]). In the latter case, it is assumed that agents share a common language and messages have an intrinsic or literal meaning, understood by all agents and which agents can reason upon (refer to Farrell (1993) [20] and Rabin (1990) [47]). From a broader perspective communication is connected to the literature on signaling games, cheap talk and disclosure. Maybe surprisingly, disclosure appears to be intimately connected with mechanism design problems. In order to grasp the main intuition we can just refer to the well-known revelation principle (Myerson (1997). A direct mechanism associated to any game is a communication stage where truthful revelation has to be "pursued".

The remaining of the paper is organized as follows. Section 2 introduces the general framework and the definition of rationalizability in communication game (Battigalli (2003)) for given restrictions on beliefs. In Section 3 this framework is used to investigate a simple situation where the informed agent can reveal part of her information and subsequently the uninformed one chooses an action. In doing that, we are able to restate and generalize the results of Rabin (1990) [47].

3.2 General framework

A communication game is a two-stage game with incomplete information where the informed agent, evocatively denoted as the *S*ender, chooses a message and the uninformed party, evocatively denoted as the *R*eceiver, selects an action. Note that the message sent by S is payoff irrelevant, instead the decision, taken by the *R*eceiver, determines the payoffs of both agents. In order to investigate this setting we will borrow the framework proposed by Battigalli (2003) [4], who used it for signaling games.

Formally, a communication game is represented as the tuple

$$\Gamma = \langle \Theta, M, A, u_S, u_R \rangle$$

with the following interpretation:

- Θ is the (nonempty) set of conceivable *payoff-types* and coincides with the set of *states of nature*. An element $\theta \in \Theta$ represents the private information of the *S*ender;
- M is the (nonempty) set *feasible messages* by the Sender;
- A is the set of all *a priori* conceivable actions for the *R*eceiver;
- u_S and u_R are the von Neumann - Morgenstern utility functions of the players. As already anticipated, we assume that $u_S : \Theta \times A \rightarrow R$ and $u_R : \Theta \times A \rightarrow R$. Namely, messages do not influence players' payoffs.

The set of strategies for the *R*eceiver is denoted as

$$S_R := \{ s_R \in A^M : s_R(m) \in A \},$$

while Σ_S is the set of feasible pairs of payoff-types and messages for the sender:

$$\Sigma_S \subseteq \Theta \times M.$$

In other words, Σ_S is the set of graphs of the *S*ender's strategies. We, also, define the set $\Theta(m) := \{\theta \in \Theta : m \in M\}$ which contains the types compatible with the message m. This notation remark the genuine incomplete approach that we will employ. Therefore different types of the *S*ender might not have a commonly share prior belief on the *R*eceiver strategies.

In order to identify the set of rationalizable messages we need to consider the agents' beliefs on the possible state of the world and on the opponent's choice. Formally, we need to append to each payoff-type of each player an epistemic type (see Harsanyi (1967-68)).

The *Sender* moves just once, then her beliefs about the *Receiver* can be summarized by $\mu_S \in \Delta(S_R)$ or equivalently by a vector of probability measures $\pi \in \prod_{m\in M} \Delta^m(A)$ where:

$$\pi(a|m,\mu_S) = \mu_S(\{s_R \in S_R : s_R(m) = a\}).$$

The *Receiver* is uncertain about the *Sender*'s payoff type and the *Sender*'s behavior. Once she receives a message, she can update her belief about the true state of the world, using the Bayes rule whenever she has initially assigned a strictly positive probability to the message actually received. Therefore, the *Receiver*'s beliefs are represented by a system of conditional probability measure:

$$\mu_R = \left(\mu_R(\cdot|\phi), \mu_R(\cdot|m)_{m\in M}\right) \in \Delta(\Sigma_S) \times [\Delta(\Theta)]^M$$

(where ϕ is the "empty history" and $\mu_R(\cdot|\phi)$ is the initial belief of the Receiver) such that, for all $m \in M$ and for all $\theta \in \Theta$,

(i) $\mu_R(\Theta(m)|m) = 1$ (the Receiver believes what he observes)

(ii) if $\mu_R(\Theta \times \{m\}|\phi) > 0$ then

$$\mu_R(\theta|m) = \frac{\mu_R((\theta,m)|\phi)}{\mu_R(\Theta \times \{m\}|\phi)}$$

The set of conditional probability systems satisfying (3.2) and (3.2) is denoted as $\Delta^*(\Sigma_S)$.

Once we have introduced the set of possible first order beliefs of the two players we can define the best response correspondences that identify which actions are compatible with the assumption of rationality.

The best response correspondence for the *Sender* is $BR_S : \Theta \times \Delta(S_R) \to 2^M$ such that:

$$\forall \theta \in \Theta, \forall \mu_S \in \Delta(S_R), BR_S(\theta,\mu_S) := \arg\max_{m\in M} \left\{ \sum_{a\in A} u_S(\theta,a)\,\pi(a|m,\mu_S) \right\}.$$

The best response for the $Receiver$ is $BR_R : M \times \Delta^* (\Sigma_S) \to 2^A$ where:

$$\forall m \in M, \forall \mu_R \in \Delta^* (\Sigma_S), BR_R (m, \mu_R) := \arg\max_{a \in A} \left\{ \sum_{\theta \in \Theta} u_R (\theta, a) \mu_R (\theta | m) \right\}.$$

Thus, a sequentially rational $Receiver$ with conditional beliefs $\mu_R \in \Delta^* (\Sigma_S)$ follows the strategy s_R such that $s_R (m) \in BR_R (m, \mu_R (\cdot | m))$ for all $m \in M$.

Up to now, we have just assumed that agents are rational but it could be natural to restrict "exogenously" the set of conceivable beliefs that agents can have[2].

If we restrict the set of possible first order beliefs to a certain nonempty set $\Delta = (\Delta_S, \Delta_R)$ where $\Delta_S \subseteq \Delta (S_R)$ and $\Delta_R \subseteq \Delta^* (\Sigma_S)$, then we can define (see Battigalli (2003) [4]) an iterative procedure that captures a form of forward-induction reasoning based on the "rationalization" of the messages of the Sender given the restrictions Δ.

More specifically the procedure corresponds to the following assumptions:

(A1.S) The $Sender$ is rational and her beliefs belong to Δ_S;

(A1.R) The $Receiver$ is rational and has beliefs in Δ_R;

(A2.S) The $Sender$ is certain of (A1.R);

(A2.R) The $Receiver$ is certain of (A1.S) *whenever possible* (that is, he initially believes (A1.S) and continues to do so after each message m consistent with (A1.S));

....

(Ak+1.S) The Sender is certain of (A1.R),...,(Ak.R);

(Ak+1.R) The Receiver is certain of (A1.S),...,(Ak.S) *whenever possible*;

....

Given her beliefs and the set of all conceivable messages, the $Receiver$ try to rationalize a message without referring to any predetermined equilibrium' message profile.

Assumptions (A2.R), (A2.S) ... capture a notion of *forward induction*: even

[2] In this paragraph we are avoiding the term "rational" for belief, because exogenous restrictions on players' belief should not be justified by some pseudo rational motive.

if the *R*eceiver is surprised by a certain message he tries to rationalize it in a manner that is consistent with the *S*ender being as sophisticated as possible given the certain message.

Battigalli and Siniscalchi (2002) [6] formally express these assumptions (by means of "complete extensive-form type spaces") and show that the pairs of payoff-types and messages of the Sender, and the strategies of the Receiver consistent with assumptions (A1)-(Ak) are those and only those which belong, respectively, to the subsets $\Sigma_1(k,\Delta)$ and $S_2(k,\Delta)$ defined by the following procedure that iteratively deletes type-message pairs for the Sender and strategies for the Receiver:

- first, define $\Sigma_S(0,\Delta) = \Sigma_S$ and $S_R(0,\Delta) = S_R$ and
- for $k = 1, 2, ...$ let $\Theta(m, k-1, \Delta) := \{\theta : (\theta, m) \in \Sigma_S(k-1,\Delta)\}$ that is, the set of payoff-types consistent with message m and a level of sophistication of the *S*ender at least equal to k, then:

$$\begin{aligned}\Sigma_S(k,\Delta) \quad &: \; = \{(\theta,m) \in \Sigma_S(k-1,\Delta) : \exists \mu_S \in \Delta_S \text{ such that } m \in BR_S(\theta,\mu_S) \text{ and } \\ &\quad \mu_S(S_R(k-1,\Delta)) = 1\} \\ S_R(k,\Delta) \quad &: \; = \{s_R \in S_R(k-1,\Delta) : \exists \mu_R \in \Delta_R, \forall m, s_R(m) \in BR_R(m, \mu_R(\cdot|m)), \\ &\quad \text{and } \Theta(m,k-1,\Delta) \neq \emptyset \text{ implies } \mu_R(\Theta(m,k-1,\Delta)|m) = 1\}\end{aligned}$$

Definition 27 *(Battigalli 2003) Fix a pair of subsets of beliefs $\Delta = (\Delta_S, \Delta_R)$, where $\emptyset \neq \Delta_S \subseteq \Delta(S_2)$ and $\emptyset \neq \Delta_R \subseteq \Delta^*(\Sigma_S)$. The payoff/message pair (θ, m) (the strategy s_R) is strongly $(k,\Delta) -$ rationalizable if $(\theta,m) \in \Sigma_S(k,\Delta)$ $(s_R \in S_R(k,\Delta))$*

Definition 28 *(θ, m) (the strategy s_R) is strongly $\Delta -$ rationalizable if and only if $(\theta,m) \in \Sigma_S(\infty,\Delta) := \bigcap_{k\geq 1} \Sigma_S(k,\Delta)$ $(s_R \in S_R(\infty,\Delta))$.*

There is another concept that does not incorporate any forward reasoning called weak $\Delta - rationalizability$. Weak $\Delta - rationalizability$ characterizes the set of of type/message pairs and strategies where all the events are true:

(WA1.S) The *S*ender is rational and her beliefs belong to Δ_S;

(WA1.R) The *R*eceiver is rational and has beliefs in Δ_R;

(WA2) The *S*ender and the *R*eceiver are certain respectively of (WA1.R) and (WA1.S) at the beginning of the game i.e., at the empty history ϕ;

....

(WAk+1) The *S*ender, and the *R*eceiver are certain respectively of (WAk) at the beginning of the game;

....

For any set of type/message pairs $\hat{\Sigma}_S \subseteq \Sigma_S$ let us define:

$$\Lambda_R\left(\hat{\Sigma}_S, \Delta\right) = \left\{\mu_R \in \Delta_R : \mu_R\left(\hat{\Sigma}_S|\Sigma_S\right) = 1\right\}$$

and for any set of *R*eceiver's strategies $\hat{S}_R \subseteq S_R$ let us define:

$$\Lambda_S\left(\hat{S}_R, \Delta\right) = \left\{\mu_S \in \Delta_S : \mu_S\left(\hat{S}_R|S_R\right) = 1\right\}.$$

The set of pairs of payoff/message of *S*ender s and of strategies of the *R*eceiver consistent with assumptions (WA1)-(Ak) are those and only those which belong, respectively, to the subsets $\Sigma_1^W(k+1,\Delta)$ and $S_2^W(k+1,\Delta)$ defined by the following recursive procedure:

- first, define $\Sigma_S^W(0,\Delta) = \Sigma_S$ and $S_R^W(0,\Delta) = S_R$ and
- for $k = 1, 2, ...$ let

$$\begin{aligned}
\Sigma_S^W(k,\Delta) &: = \left\{(\theta, m) \in \Sigma_S^W(k-1,\Delta) : \exists \mu_S \in \Lambda_S\left(S_R^W(k-1,\Delta),\Delta\right) \text{ such that } m \in BR_S(\theta,\mu_S)\right\} \\
S_R^W(k,\Delta) &: = \left\{s_R \in S_R^W(k-1,\Delta) : \exists \mu_R \in \Lambda_R\left(\Sigma_S^W(k-1,\Delta),\Delta\right) \text{ such that} \right. \\
&\qquad \left. \text{for every } m \in BR_R(m,\mu_R)\right\}
\end{aligned}$$

Definition 29 *Fix a pair of subsets of beliefs $\Delta = (\Delta_S, \Delta_R)$, where $\emptyset \neq \Delta_S \subseteq \Delta(S_2)$ and $\emptyset \neq \Delta_R \subseteq \Delta^*(\Sigma_S)$. The payoff/message pair (θ, m) (the strategy s_R) is weakly (k,Δ) − rationalizable if $(\theta, m) \in \Sigma_S(k,\Delta)$ $(s_R \in S_R(k,\Delta))$*

Definition 30 *(θ, m) (the strategy s_R) is weakly Δ−rationalizable if and only if $(\theta, m) \in \Sigma_S(\infty,\Delta) := \bigcap_{k\geq 1} \Sigma_S(k,\Delta)$ $(s_R \in S_R(\infty,\Delta) := \bigcap_{k\geq 1} S_R(k,\Delta))$.*

Note that these assumptions are silent about how the *R*eceiver would change their beliefs if he observed a message which he believed impossible at the beginning of the game. Therefore, weak rationalizability cannot capture any kind of forward induction reasoning. For this reason we will apply the strong version of rationalizability for our main result. Instead, when we will analyze previous results in the literature we will mention also the weaker form of rationalizability.

3.3 An example

Let us give a very simple example of a communication game where a very weak restriction on Receiver's beliefs implies that full disclosure is the only possible message profile.

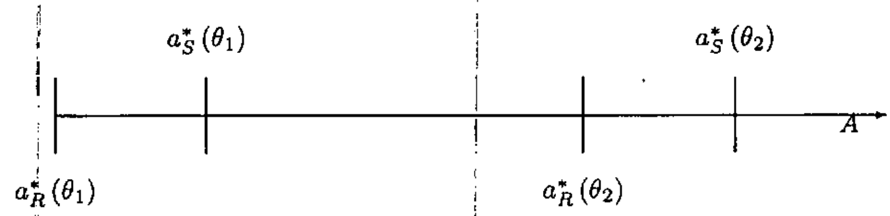


Figure 1

Suppose that preferences of the *S*ender and of the *R*eceiver are represented by the following utility functions $u_R(a,\theta) = -(a-\theta)^2$ and $u_S(a,\theta,b) = -(a-(\theta+b))^2$. Both utility functions depend on the random variable θ which is observed by the *S*ender. After having observed θ the *S*ender can send a message m to the *R*eceiver. After getting the message the *R*eceiver chooses an action a.

We can denote the message "θ is equal to $\bar{\theta}$" as $[\bar{\theta}]$ and for sake of simplicity we assume that $\Theta = \{\theta_1, \theta_2\} = \{0.1, 0.8\}$, $A = [0,1]$ and $b = 0.1$. If the *R*eceiver is certain that the state of the world is θ then her optimal action is $a_R^*(\theta) = \underset{a\in A}{\arg\max} u_R(a,\theta) = \theta$, instead at θ the *S*ender prefers $a_S^*(\theta,b) = \underset{a\in A}{\arg\max} u_S(a,\theta) = \theta + b$ (see Figure 1). Conditional on the generic message $[\theta]$ the *R*eceiver will choose $a_R^*[\theta] = \underset{a\in A}{\arg\max}\{u_R(a,\theta_1)\mu(\theta_1|[\theta]) + u_R(a,\theta_2)\mu(\theta_2|[\theta])\}$. Conditional on message $[\theta]$, if the conditional beliefs are not degenerated the optimal action belongs to the open interval $(a_R^*(\theta_1), a_R^*(\theta_2))$.[3]

Let the *R*eceiver think that the generic θ is more likely if the *S*ender says that his type is θ. Namely, if the *S*ender declares that his type is θ_1, sending message $[\theta_1]$, the *R*eceiver assigns at least probability $\frac{1}{2}$ that true type of the

[3] In symbols, $a_R^*[\theta] \in (a_R^*(\theta_1), a_R^*(\theta_2))$.

*S*ender is θ_1. Given this restriction on the conditional beliefs of the *R*eceiver, the message $[\theta_2]$ is strictly preferred by the *S*ender if his type is θ_2. The *R*eceiver knowing that the *S*ender is rational will deduce that conditional on the message $[\theta_1]$ the only conceivable payoff type of *S*ender is exactly θ_1. Moreover, if we assume that conditional on the vague message $[\Theta]$ the *R*eceiver will not exclude any possible payoff type of the *S*ender, we can conclude that the only rational message profile is $[\theta_1]$ if $\theta = \theta_1$ and $[\theta_2]$ if $\theta = \theta_2$. Under very weak assumptions on the beliefs of the *R*eceiver, even if the messages are cheap, full disclosure is the only rationalizable message profile.

3.4 Rational disclosure: one agent is informed and one agent acts

With the framework introduced in Section 3.2 we can investigate whether disclosure of information is justifiable. One of the early contributions on this topic is the article by Matthew Rabin (1990) [47]. He used a "non-equilibrium" approach and characterized the information that can be communicated between rational agents. We will pursue a non-equilibrium analysis in a genuine incomplete information without assuming that the players share the same beliefs on the set of the payoff parameters Θ.

A "non-equilibrium approach" has two different, although related, advantages with respect to an equilibrium approach: first, a message has literal meaning without referring to the induced *R*eceiver's action; second, in order to interpret a message the *R*eceiver should take as alternatives all the messages which are justifiable by some degree of sophistication of the *S*ender. Using an equilibrium analysis, the effect of possible deviations, namely sending different messages, is evaluated considering a putative equilibrium. The set of conceivable messages matters and its relation with counterfactual reasoning. Let us suppose that in a given equilibrium the sender has a message ("θ is in $\hat{\Theta} \subseteq \Theta$") which would induce the *S*ender to revise her expectation and choose a Pareto improving action. Then, if we used an equilibrium approach, we would conclude that the *R*eceiver should interpret the absence of that message as "θ is not in $\hat{\Theta}$".

The set of feasible messages is rich enough to distinguish any subset of Θ. Formally, the cardinality of M must be $2^{|\Theta|} - 1$.
The meaning of messages is commonly shared by the players. Therefore, we

can identify M with $2^{\Theta}/\{\emptyset\}$. In order to diffentiate messages with respect to subsets of Θ we introduce the following notation. The message "my payoff type belongs to $\hat{\Theta}$", where $\hat{\Theta}$ is a non-empty subset of Θ, is denoted as $\left[\hat{\Theta}\right]$, for example the message $[\theta]$ should be read as "my payoff-type is θ". Formally, we can define the interpretation isomorphism between the set of non empty subsets of Θ ($2^{\Theta}/\{\emptyset\}$) and the set of possible messages ($[\cdot] : 2^{\Theta}/\{\emptyset\} \rightarrow M$):

$$\left[\left\{\theta \in \Theta : \theta \in \hat{\Theta}\right\}\right] = \left[\hat{\Theta}\right] \in M \text{ for every nonempty subset of } \Theta \ (\emptyset \neq \hat{\Theta} \subseteq \Theta)$$

We remark that players assign to any message its literal meaning. This will avoid a trivial source of multiplicity of equilibria. Given any equilibrium in a communication game we can always "relabel" messages and obtain a new equilibrium. As Farrell (1993) [20] pointed out, this kind of multiplicity of equilibria seems quite artificial. If we consider a natural language a message as "the state of the world is θ" is hardly interpreted as "the state of the world is θ'", just because we have changed the vocabulary.[4]

In order to simplify our analysis we assume that the set of states of Nature, Θ, is an ordered finite set[5] and the utility functions are such that, if $\theta > \theta'$ then $\frac{\partial u_k(a,\theta)}{\partial a} > \frac{\partial u_k(a,\theta')}{\partial a}$ for $k \in \{S, R\}$.
The set of possible actions is an interval in $\mathbb{R}$ that is, A is a compact and convex subset of $\mathbb{R}$.
We can define the most preferred action at any given state of nature for the two players and the best response of the *R*eceiver given a generic message $\left[\hat{\Theta}\right]$ as:

$$\begin{aligned} a_S^*(\theta) &: = \arg\max_{a \in A} u_S(a, \theta) \\ a_R^*(\theta) &: = \arg\max_{a \in A} u_R(a, \theta) \\ a_R^*\left(\mu\left(\cdot \middle| \left[\hat{\Theta}\right]\right)\right) &: = \arg\max_{a \in A} \int_{\Theta} u_R(a, \theta)\, \mu\left(d\theta \middle| \left[\hat{\Theta}\right]\right). \end{aligned}$$

The utility function of the players are assumed to be continuous in both arguments and strictly concave in A. The *S*ender's utility can be written as a decreasing function of $d(a, a_S^*(\theta))$, which denotes the euclidean distance be-

[4] Given the assumption of a rich message space, for every information m that the *S*ender might want to convey, there is a message $\left[\hat{\Theta}\right] \in M$ whose literal meaning is that the *S*ender's payoff type is in $\hat{\Theta}$. As a consequence, the Δ *-rationalizable* strategy profile will be, by construction, robust to any conceivable *neologism* (see Farrell (1993) [20]).

[5] Results still hold if Θ is at most countable.

tween the action a and the most preferred action for the *S*ender when the state of Nature is θ:

$$u_S(a,\theta) = h\left(d\left(a, a_S^*(\theta)\right), \theta\right) \text{ where } h \text{ is decreasing}$$

and the marginal utility of the *R*eceiver $\frac{\partial u_R(\cdot,\cdot)}{\partial a}$ is such that:

$$d\left(\frac{\partial u_R(a,\theta)}{\partial a}, \frac{\partial u_R(\hat{a},\theta)}{\partial a}\right) \leq d(a,\hat{a}) \text{ for every } \theta \in \Theta$$

Focusing on more economic assumptions first, we assume a certain degree of alignment between the *R*eceiver's and the *S*ender's interests. This assumption guarantees the existence of an equilibrium different from the babbling one. Therefore, we restrict the set of possible payoff types Θ in such a way that:

$$\sup_{a\in A,\theta\in\Theta} d\left(\frac{\partial u_S(a,\theta)}{\partial a}, \frac{\partial u_R(a,\theta)}{\partial a}\right) \leq \sigma \qquad (\sigma - alignment')$$

where the distance considered in $\mathbb{R}$ is the euclidean distance.

Lemma 6 *If the conditions* $(\sigma - alignment')$ *holds then*

$$\sup_{\theta\in\Theta} d\left(a_S^*(\theta), a_R^*(\theta)\right) \leq \sigma \qquad (\sigma - alignment)$$

This latter condition is more easily interpretable. The two players' most preferred actions are always quite close. Hence, we call these two condition "*alignment*".

Second, different types of the *S*ender have different preferences:

$$\inf_{\theta',\theta''\in\Theta} d\left(a_S^*(\theta'), a_S^*(\theta'')\right) > \varepsilon \qquad (\varepsilon - sorting)$$

This assumption implies that if θ varies the most preferred action varies as well. This assumption can be justified on the basis of two rationales. First, the situation should not be trivial if preferences of the *S*ender remain constant she will always send the same message. Second, if the *R*eceiver would like to act differently in θ' and θ'' the preferences of the *S*ender should be different in the two states of the world.[6]

[6] This is resembles the incentive compatibility and the $\beta-$monotonicity condition in the mechanism design's literature.

Lemma 7 *If the conditions (σ − alignment) and (ε − sorting) hold then*

$$d\left(a_R^*(\theta'), a_S^*(\theta'')\right) \geq \varepsilon - \sigma$$

for every θ', θ'' belonging to Θ.

Proof. By the triangular inequality $d_A\left(a_S^*(\theta''), a_S^*(\theta')\right) \leq d_A\left(a_S^*(\theta''), a_R^*(\theta'')\right) + d_A\left(a_R^*(\theta''), a_S^*(\theta')\right)$, therefore $d_A\left(a_R^*(\theta''), a_S^*(\theta'')\right) \geq \varepsilon - \sigma$. ■

A part from preference' restrictions, we need to restrict the set of conceivable belief of the *Receiver*. Namely, we assume that the *Receiver* trust to a certain extent the messages sent by the *Sender*, this can be justified by considering the alignment of interests, Nevertheless, he never excludes the possibility of misreporting[7] if the message is not sharp.

Formally, we assume that her beliefs conditional on a generic message $\left[\hat{\Theta}\right]$ satisfy the following condition:

$$\mu_R\left(\hat{\Theta}|\left[\hat{\Theta}\right]\right) > \mu_R\left(\left(\hat{\Theta}^c|\left[\hat{\Theta}\right]\right)\right) \qquad \text{(Trust)}$$

where $\hat{\Theta}^c$ is the complementary w.r.t. Θ of $\hat{\Theta}$ (in symbols $\hat{\Theta}^c := \Theta \backslash \hat{\Theta}$)

Moreover, when the message is not sharp[8] the *Receiver* is quite cautious i.e. the conditional belief $\mu_R\left(\theta|\left[\hat{\Theta}\right]\right)$ is not a degenerated measure on Θ, namely:

$$\exists \theta \neq \theta' \text{ such that } \mu_R\left(\theta|\left[\hat{\Theta}\right]\right) > \kappa \text{ and } \mu_R\left(\theta'|\left[\hat{\Theta}\right]\right) > \kappa \qquad \text{(Cautiousness)}$$

where $\kappa > 0$ and $\sigma < \kappa\varepsilon$.

The two conditions describe two different attitudes of the *Receiver*: if the message is not precise the *Receiver* is incline to believe the message received but she will preserve some degree of cautiousness. These assumptions seem quite appealing given the assumptions we made on the preferences of the two players.

Formally, with conditions (Trust) and (Cautiousness) we are restricting the set of conceivable beliefs of the *Receiver*. Therefore, the set of conceivable first order beliefs is $\Delta = (\Delta_S, \Delta_R)$ where $\Delta_S = \Delta(S_R)$ and $\Delta_R := \left\{\mu \in \Delta^*(\Sigma_S) : \mu_R\left(\hat{\Theta}|\left[\hat{\Theta}\right]\right) > \mu_R\left(\left(\hat{\Theta}^c|\left[\hat{\Theta}\right]\right)\right) \text{ for every } \left[\hat{\Theta}\right] \in M \text{ and if } \left|\hat{\Theta}\right| > 1 \text{ then } \exists \theta \neq \theta' \text{ such that } \mu_R\left(\theta|\left[\hat{\Theta}\right]\right) > \kappa \text{ and } \mu_R\left(\theta'|\left[\hat{\Theta}\right]\right) > \kappa\right\}$.

---

[7] This second condition recalls the skepticism assumption made by Grossman - Hart (1980) and Fujiwara et al. (1990) [44].

[8] A message $\left[\hat{\Theta}\right]$ is not sharp whenever $\left|\hat{\Theta}\right| > 1$.

Definition 31 *(Battigalli (2003) [4]) The payoff/message pair (θ, m) (the strategy s_R) is $(k, \Delta) -$ rationalizable if $(\theta, m) \in \Sigma_S(k, \Delta)$ $(s_R \in S_R(k, \Delta))$*

Definition 32 *(θ, m) (the strategy s_R) is $\Delta -$ rationalizable if and only if $(\theta, m) \in \Sigma_S(\infty, \Delta) := \bigcap_{k \geq 1} \Sigma_S(k, \Delta)$ $(s_R \in S_R(\infty, \Delta))$. For any given message m, let*

$$k_\Delta(m) = \max\{k : \Theta(m, k, \Delta) \neq \emptyset\}$$

We say that $\Theta(m, k_\Delta(m), \Delta)$ is the best $\Delta -$ rationalization of a message m.

Now, we have to formally define when disclosure is rational.

Definition 33 *Partial truthful disclosure is rational if and only if for every strategy profile $(\sigma_S, s_R) \in M^\Theta \times A^M$, such that the corresponding pairs (θ, m) and the strategy s_R are $\Delta -$ rationalizable, there exists at least a payoff type θ such that $\sigma_S(\theta) = [\theta]$ and $s_R(\theta) = a_R^*(\theta)$.*
Truthful disclosure is rational if there exists a (unique) profile $(\sigma_S, s_R) \in M^\Theta \times A^M$ such that the corresponding payoff/strategy pairs are $\Delta -$ rationalizable and $\sigma_S(\theta) = [\theta]$, $s_R([\theta]) = a_R^*(\theta)$ for all $\theta \in \Theta$.

In the first case *S*ender reveals information only in some states of Nature and he is trusted by the *R*eceiver. In the second case, for every state of Nature the *S*ender rationally reveals his private information and the *R*eceiver believes him.
Let us consider the two slightly different situations: first, the *S*ender reveals only partially his information, for example sending a vague message; second, the *S*ender discloses completely the true state of Nature but the *R*eceiver does not believe him and chooses an action $a \neq a_R^*(\theta)$. According to our definition neither partial truthful disclosure nor Truthful disclosure has occurred in these situations. Now, we can state plainly the first main result.

Proposition 9 *If the conditions ($\sigma -$ alignment), ($\varepsilon -$ sorting), (Trust) and (Cautiousness) hold $\Delta -$ rationalizability implies that truthful disclosure is rational.*

Proof. See Appendix ■

Given our assumptions which are parameterized by σ, ε and κ, full disclosure and trust are the only possible results of a communication game. But if we look to the negative part of this result, we can conclude that in order for full disclosure to occur the condition needed are really strict. Not only interests must be aligned but we need to assume a certain degree of trust by the *R*eceiver. Mistrust is self enforcing, if the *R*eceiver is too much pessimistic the possibility of cooperation is void.

In the next subsection we will show that the assumptions made are tight.

3.4.1 Tightness

First, suppose that assumption Trust does not hold. We could say that the *R*eceiver does not trust the sender. Now, we can take a situation where players' interests are represented by the following figure (Figure 2). In this setting condition (Cautiousness) holds vacously. Therefore, the set of the *R*eceiver's belief is unrestricted i.e. $\Delta'_R = \Delta^*(\Sigma_S)$.

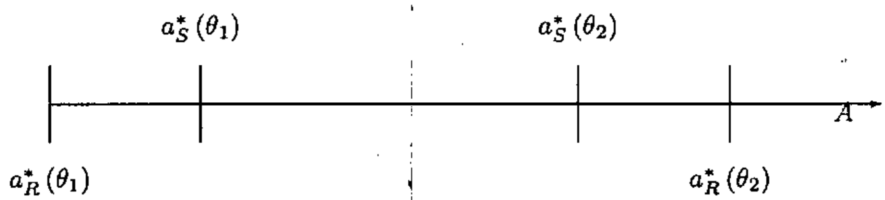


Figure 2

Given any message m, we define the set of conceivable best responses of the *R*eceiver, given that his beliefs belong to the set $\Delta'_R \subseteq \Delta^*(\Sigma_S)$ as:

$$A^*(m, \Delta'_R) := \{a \in A : \exists \mu \in \Delta_R \text{ s.t. } s_R \in BR_R(m, \mu) \text{ and } s_R(m) = a\}.$$

$A^*(m, \Delta'_R)$ represents the set of actions that, given message m, can be justified by some belief of the *R*eceiver.
Let us recall that $\Theta\left(\left[\hat{\Theta}\right], k-1, \Delta\right) := \left\{\theta : \left(\theta, \left[\hat{\Theta}\right]\right) \in \Sigma_S(k-1, \Delta)\right\}$, i.e. $\Theta\left(\left[\hat{\Theta}\right], k-1, \Delta\right)$ denotes the types for whom the message $\left[\hat{\Theta}\right]$ survives k step of iterative elimination of strictly dominated messages.

$\Theta([\theta], 0, \Delta) = \Theta$ for $[\theta] \in \{[\theta_1], [\theta_2]\}$. If *R*eceiver's beliefs belong to the set $\Delta_R = \Delta^*(\Sigma_S)$ then for any arbitrary message $[\theta]$ all the actions a belongs to the interval $[a_R^*(\theta_1), a_R^*(\theta_2)]$ are justifiable by some belief in Δ_R. For every a in $[a_R^*(\theta_1), a_R^*(\theta_2)]$ there exists a belief function $\mu_R(\cdot|[\theta])$ such that $a \in \arg\max_{a' \in A} \sum_{\theta \in \Theta} u_R(\theta, a') \mu_R(\theta|m)$. Therefore, $A^*[\theta_1, \theta_2] = A^*[\theta_1] = A^*[\theta_2]$ and no payoff type/message pair can be eliminated.

This example shows that when trust does not hold the remaining assumptions do not imply full disclosure.

Next, we drop condition (Cautiousness). We impose, only, the following restriction on the *R*eceiver's beliefs: $\Delta_R'' := \left\{\mu_R \in \Delta^*(\Sigma_S) : \mu_R\left(\hat{\Theta}|\left[\hat{\Theta}\right]\right) > \mu_R\left(\hat{\Theta}^c|\left[\hat{\Theta}\right]\right)\right\}$. Then the overall set of conceivable beliefs is denoted as $\Delta'' = (\Delta_R'', \Delta(S_R))$.

If the preferences of *S*ender and *R*eceiver are such that the following representation holds (Figure 3), then it is straightforward to prove that type θ_1 strictly prefers message $[\theta_1]$ to message $[\theta_3]$ and similarly type θ_3 strictly prefers message $[\theta_3]$ to message $[\theta_1]$.

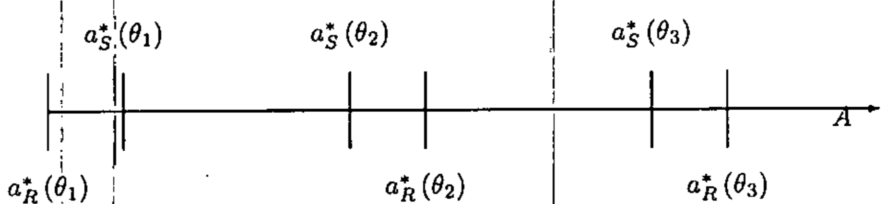


Figure 3

Hence $\Theta([\theta_1], 1, \Delta'') = \{\theta_1, \theta_2\}$ and $\Theta([\theta_3], 1, \Delta'') = \{\theta_2, \theta_3\}$. Moreover, at the second step of the iterative procedure we can prove that $\Theta([\theta_1, \theta_2], 2, \Delta'') = \{\theta_1, \theta_2\}$, $\Theta([\theta_2, \theta_3], 2, \Delta'') = \{\theta_2, \theta_3\}$ and $\Theta([\theta_1, \theta_3], 2, \Delta'') = \{\theta_1, \theta_2, \theta_3\}$.

Let's consider type θ_1, and assume that the *R*eceiver has conditional beliefs such that $\mu_R(\theta_1|[\theta_1]) = 1$ and $\mu_R(\theta_1|[\theta_1, \theta_2]) = 1$. Note that these conditional beliefs belong to Δ'' and $\mu_R(\Theta([\theta_1], 1, \Delta)|[\theta_1]) = 1$, $\mu_R(\Theta([\theta_1, \theta_2], 2, \Delta)|[\theta_1]) = 1$. The sets of possible best replies by the *R*eceiver associated to these conditional beliefs overlap. Formally, $\bigcup_{\mu_R \in \Delta_R''} BR_R([\theta_1], \mu_R) \cap \bigcup_{\mu_R \in \Delta_R''} BR_R([\theta_1, \theta_2], \mu_R) \neq \emptyset$. The same is true for θ_2 if we consider messages $[\theta_2]$ and $[\theta_1, \theta_2]$. Therefore, these two types can not credibly signal themselves and $\Delta-$*rationalizability* does not imply truthful disclosure

If we do not impose the other conditions, pertaining to preferences, it is easy to show that our result does not apply. If $\sigma-$*alignment* does not hold we can refer to Lemma 2 of Sobel and Crawford (1984) [16]. Similarly, if $\varepsilon - sorting$ does not hold we can refer to classical results in mechanism design literature.

3.5 Application: Rabin(1990)

In the previous section we have found sufficient conditions for full disclosure to occur. Now we want to investigate and generalize previous results in the literature, stating clearly the epistemic assumptions needed to those results to hold. Specifically, we will relate our results to the work by Rabin (1990). In order to investigate the relation between the $\Delta - rationalizability$ and the concept of "Credible Message Profile" we need to state some preliminary definitions used by Rabin.

Let us assume that a communication game has "*no relevant tie*" if the following holds: for all pair of outcomes (θ, a'), (θ, a'') if a', $a'' \in A$ such that $a' \neq a''$ then $u_R(\theta, a') \neq u_R(\theta, a'')$. Moreover, wlog we can assume that for every $\theta' \neq \theta''$ either $u_S(\theta', \cdot) \neq u_S(\theta'', \cdot)$ or $u_R(\theta', \cdot) \neq u_R(\theta'', \cdot)$.[9]

Definition 34 *(Rabin 1990) A type profile is a list of exclusive, not necessarily exhaustive, nonempty subsets of types of player S,* $\mathcal{X} = \left\{\hat{\Theta}_1, \hat{\Theta}_2, ..., \hat{\Theta}_D\right\}$. *Let* $\Theta_{\mathcal{X}} := \left\{\theta \in \Theta : \exists \hat{\Theta}_k \in \mathcal{X} \text{ such that } \theta \in \hat{\Theta}_k\right\}$ *be the set of types which are in some subset belonging to the type profiles. Let* $M(\mathcal{X}) = \left\{\left[\hat{\Theta}_1\right], \left[\hat{\Theta}_2\right], ..., \left[\hat{\Theta}_D\right]\right\}$ *be the set of messages* $\left[\hat{\Theta}_k\right]$ *associated to the subset of* Θ, $\hat{\Theta}_k$.[10]

We define with $\Theta^c_{\mathcal{X}} := \Theta \backslash \Theta_{\mathcal{X}}$ the set of types that do not belong to $\Theta_{\mathcal{X}}$. Note that a type profile is equivalent to a specific set of strategy profiles of the *S*ender. Each payoff-type $\theta \in \hat{\Theta}_k$ chooses the signal $\left[\hat{\Theta}_k\right]$ and $\theta \in \Theta^c_{\mathcal{X}}$ can choose any message. Formally the associated strategy is $\sigma_S : \Theta \rightarrow M$ such that $\sigma_S(\theta) = \left[\hat{\Theta}_k\right]$ if $\theta \in \Theta^c_{\mathcal{X}}$, $\sigma_S(\theta) \in M$ otherwise.

Rabin assumed that the prior belief of the *R*eceiver is common knowledge. Let us denote the initial belief of the *R*eceiver as $p(\cdot)$. To avoid trivialities we assume that $p(\cdot)$ is strictly positive.

Moreover, the conditional beliefs of the *S*ender are such that $\mu\left(\theta | \left[\hat{\Theta}\right]\right) = p(\theta)$

[9] Note that if there exists θ' and θ'' such that $u_S(\theta', \cdot) = u_S(\theta'', \cdot)$ and $u_R(\theta', \cdot) = u_R(\theta'', \cdot)$ we can merge them or simply drop one of the two.

[10] A type profile is equivalent to a specific set of strategy profiles of the *S*ender. Each payoff-type $\theta \in \hat{\Theta}_k$ chooses the signal $\left[\hat{\Theta}_k\right]$ and $\theta \in \Theta \backslash \Theta_{\mathcal{X}}$ can choose any message. Formally the associated strategy is $\sigma_S : \Theta \rightarrow M$ such that $\sigma_S(\theta) = \left[\hat{\Theta}_k\right]$ if $\theta \in \hat{\Theta}_k$, $\sigma_S(\theta) \in M$ otherwise.

if $\theta \in \hat{\Theta}$ and $\mu\left(\theta| \left[\hat{\Theta}\right]\right) = 0$ otherwise. Formally, we can restate the previous assumption as restrictions on the *R*eceiver's belief, i.e.:[11]

$$\begin{aligned} \Delta'_R \; = \; & \left\{ \left(\mu_R(\cdot|\phi), \mu_R(\cdot|m)_{m \in M}\right) \in \Delta^*(\Sigma_S) : \operatorname{marg}_{\Theta}\mu_R(\cdot|\phi) = p(\cdot) \right. \\ & \text{and } \mu_R\left(\theta| \left[\hat{\Theta}\right]\right) = \frac{p(\theta)}{\sum_{\theta' \in [\hat{\Theta}]} p(\theta')} \text{ if } \theta \in \left[\hat{\Theta}\right], \mu_R\left(\theta| \left[\hat{\Theta}\right]\right) = 0 \text{ otherwise for every } \left.\left[\hat{\Theta}\right] \in M \right\} \end{aligned} \qquad (\Delta')$$

We can define the set of optimal action by the *R*eceiver conditional on message $\left[\hat{\Theta}\right]$ as $A^*\left(\left[\hat{\Theta}\right]\right) := \left\{ a^* \in A : a^* \in \arg\max_{a \in A} \sum_{\theta \in \hat{\Theta}} p(\theta)\, u_R(a, \theta) \right\}$. $A^*\left(\left[\hat{\Theta}\right]\right)$ is the set of actions that the *S*ender considers possible after sending message $\left[\hat{\Theta}\right]$, under the assumption that her opponent is rational, she is certain of this fact and of the conditional beliefs of the *R*eceiver. Formally, one should note that $\left\{ s_R \in S_R : s_R\left(\left[\hat{\Theta}\right]\right) \in A^*\left(\left[\hat{\Theta}\right]\right) \text{ for } \left[\hat{\Theta}\right] \in M \right\} = S_R(1, \Delta')$ when $\Delta' = \Delta(S_R) \times \Delta'_R$.

The type in $\Theta^c_{\mathcal{X}}$ could choose to lie, sending a message $\left[\hat{\Theta}_k\right] \in M(\mathcal{X})$, or to say the truth, sending the residual message $[\Theta^c_{\mathcal{X}}]$. Rabin denoted with $\Theta\left(\left[\hat{\Theta}_k\right], \mathcal{X}\right)$, the set of types who have has some incentives to lie by choosing a specific message $\left[\hat{\Theta}_k\right]$.[12]

Definition 35 *Let $\Theta\left(\left[\hat{\Theta}_k\right], \mathcal{X}\right) \subseteq \Theta^c_{\mathcal{X}}$ be the set of types such that at least one of the following conditions DOES NOT hold:*

(NC1) $A^*\left(\left[\hat{\Theta}_k\right]\right) = \left\{ a^* : a \in \arg\min_{a \in A^*} u_S(a, \theta) \right\}$

(NC2) $\exists \left[\hat{\Theta}_j\right] \in M(\mathcal{X})$ *such that* $u_S(a^*, \theta) < u_S(\tilde{a}, \theta)$ *for every* $a^* \in A^*\left(\left[\hat{\Theta}_k\right]\right)$ *and* $\tilde{a} \in A^*\left(\left[\hat{\Theta}_j\right]\right)$.

---

[11] The *R*eceiver maximization problem conditional on message $\left[\hat{\Theta}\right]$ can be equivalently represented as $\max_{a \in A} \sum_{\theta \in \hat{\Theta}} p(\theta)\, u_R(a, \theta)$ or $\max_{a \in A} \sum_{\theta \in \hat{\Theta}} \frac{p(\theta)}{\sum_{\theta' \in [\hat{\Theta}]} p(\theta')} u_R(a, \theta)$

[12] According to Definition 35, a type θ could choose a message $\left[\hat{\Theta}_k\right]$, even if it implies the lowest outcome, just because there is no other message in $M(\mathcal{X})$ leading to a strictly better outcome.

Even worse, this definition implies that a type θ could send a strongly dominated message (given restriction Δ') just because it induces an action which does not minimize his payoff.

Note that $\Theta\left(\left[\hat{\Theta}_k\right], \mathcal{X}\right) \supseteq \Theta\left(\left[\hat{\Theta}_k\right], \mathbf{3}, \Delta'\right)$ because θ could belong to $\Theta\left(\left[\hat{\Theta}_k\right], \mathcal{X}\right)$ even if the message $\left(\theta, \left[\hat{\Theta}_k\right]\right) \notin \Sigma_S(2, \Delta')$, because condition (NC2) does not hold.

Moreover, one can note that the following stronger inclusion holds: $\Theta\left(\left[\hat{\Theta}_k\right], \mathcal{X}\right) \supseteq \Theta\left(\left[\hat{\Theta}_k\right], \mathbf{1}, \Delta'\right)$.

Definition 36 *Let $A^{**}\left(\left[\hat{\Theta}_k\right], \mathcal{X}\right)$ be the set of actions a^* such that $\exists \pi : \Theta \to [0,1]$ satisfying the following conditions:*

1. $\pi(\theta) = p(\theta)\ \forall \theta \in \hat{\Theta}_k$,
2. $\pi(\theta) \in [0, p(\theta)]\ \forall \theta \in \Theta\left(\hat{\Theta}_k, \mathcal{X}\right)$
3. $\pi(\theta) = 0$ *else*

   such that $a^* \in \arg\max_{a \in A^*} \sum_{\theta \in \Theta} \pi(\theta) u_R(a, \theta)$

$\pi(\cdot)$ is not normalized to sum to one and $A^{**}\left(\hat{\Theta}_k, \mathcal{X}\right)$ identifies the set of optimal actions that the *Receiver* could take "*if he were sure he was facing all types in Θ_k and were not sure which types he was facing in $\Theta\left(\hat{\Theta}_k, \mathcal{X}\right)$*" (see Rabin (1990)[47]).

Formally, we can characterize $A^{**}\left(\left[\hat{\Theta}_k\right], \mathcal{X}\right)$ in term of justifiable action by the *Receiver* conditional on message $\left[\hat{\Theta}_k\right]$ imposing the following restrictions on his beliefs:

$$\begin{aligned}\Delta''_R = & \left\{\left(\mu_R(\cdot|\phi), \mu_R(\cdot|m)_{m\in M}\right) \in \Delta^*(\Sigma_S) : \text{marg}_\Theta \mu_R(\cdot|\phi) = p(\cdot)\right. \\ & \left.\text{and for every } \left[\hat{\Theta}\right] \in M\ \mu_R\left(\theta|\left[\hat{\Theta}\right]\right) \geq p(\theta) \text{ if } \theta \in \left[\hat{\Theta}\right],\ \mu_R\left(\theta|\left[\hat{\Theta}\right]\right) \leq p(\theta) \text{ otherwise }\right\}\end{aligned}$$

Therefore, $\left\{s_R \in S_R : s_R\left(\left[\hat{\Theta}_k\right]\right) \in A^{**}\left(\left[\hat{\Theta}_k\right], \mathcal{X}\right) \text{ for every } \hat{\Theta}_k \in \mathcal{X}\right\} \supseteq S_R(2, \Delta'')$ where $\Delta'' = \Delta(S_R) \times \Delta''_R$.[13]

Definition 37 *$\mathcal{X}$ is a Credible Message Profile (henceforth CMP) if and only if $\forall \hat{\Theta}_k \in \mathcal{X}$,*

(C1) $\forall \theta \in \hat{\Theta}_k$, $A^*\left(\left[\hat{\Theta}_k\right]\right) = \arg\max_{a \in A^*} u_S(a, \theta)$ *and*

[13] It could happen that $\left\{s_R \in S_R : s_R\left(\left[\hat{\Theta}_k\right]\right) \in A^{**}\left(\left[\hat{\Theta}_k\right], \mathcal{X}\right) \text{ for every } \hat{\Theta}_k \in \mathcal{X}\right\} \supsetneq S_R(1, \Delta')$.

(C2) $A^*\left(\left[\hat{\Theta}_k\right]\right) = A^{**}\left(\left[\hat{\Theta}_k\right], \mathcal{X}\right)$.[14]

A CMP, $\mathcal{X}$, can be identified with the set of strategy profiles, (σ_S, s_R), such that $\sigma_S(\theta) = \begin{cases} \left[\hat{\Theta}_k\right] & \text{if } \theta \in \hat{\Theta}_k \\ M \backslash \left[\hat{\Theta}_k\right] & \text{if } \theta \notin \left[\hat{\Theta}_k\right] \cup \Theta\left(\left[\hat{\Theta}_k\right], \mathcal{X}\right) \\ m(\theta) & \text{otherwise} \end{cases}$ where $m(\theta)$ is a generic message in M depending on θ. $s_R\left(\left[\hat{\Theta}_k\right]\right) \in A^*\left(\left[\hat{\Theta}_k\right]\right)$ for every $\left[\hat{\Theta}_k\right] \in M$. It is worth to remember that a generic strategy σ_S can be associated to set of payoff/message pairs $\Sigma_S^{\sigma_S} := \left\{\left(\theta, \left[\hat{\Theta}\right]\right) \in \Theta \times M : \sigma_S(\theta) = \left[\hat{\Theta}\right]\right\}$.
With a slight abuse of expression we will define a strategy profile, (σ_S, s_R) or $\Sigma_S^{\sigma_S} \times s_R$ to be credible if and only if the associated message profile is credible. These strategy profiles are characterized by the fact that the *R*eceiver acts as if the message is truthful and each type of the *S*ender in $\Theta_{\mathcal{X}}$ gets the most preferred outcome. CMP exists only if we assume a certain degree of alignment of interest. Otherwise, $A^*(X_j) = \arg\max_{a \in A^*} u_S(a, \theta)$ will imply $A^*\left(\hat{\Theta}_j\right) \neq A^{**}\left(\hat{\Theta}_j, \mathcal{X}\right)$.

Rabin (1990) introduced the following last definition which formalizes an iterative procedure which is equivalent to an iterative elimination of dominated strategies.

Definition 38 *Fix a type profile $\mathcal{X}$. The Message Profile Theory (MPT) with respect to $\mathcal{X}$, denoted $MPT(\mathcal{X})$, is the set of strategy pairs $\{(\sigma, s) : (\sigma, s) \in \Sigma_S^{\mathcal{X}} \times S_R^{\mathcal{X}}\}$, where $\Sigma_S^{\mathcal{X}} \times \Sigma_R^{\mathcal{X}}$ is constructed as follows:*
$S_R(0) = \left\{s \in S_R : \forall \hat{\Theta}_k \in \mathcal{X},\ s\left(\left[\hat{\Theta}_k\right]\right) = A^*\left(\left[\hat{\Theta}_k\right]\right) \text{ and } \forall \left[\hat{\Theta}\right] \notin \mathcal{X},\ s\left(\left[\hat{\Theta}\right]\right) \in A^*\right\}$
$\Sigma_S(0) = \left\{\sigma \in \Sigma_S : \forall \hat{\Theta}_k \in \mathcal{X}, \forall \theta \in \hat{\Theta}_k,\ \sigma(\theta) = \left[\hat{\Theta}_k\right] \text{ and } \forall \theta \notin \left[\hat{\Theta}_k\right] \cup \Theta\left(\left[\hat{\Theta}_k\right], \mathcal{X}\right)\ \sigma(\theta) \in M \backslash \left[\hat{\Theta}_k\right]\right\}$.[15]
Then recursively:
$S_R(n+1) = \{s \in S_R(n) : s \text{ is not strongly dominated with respect to } \Sigma_S(n) \text{ by any } s' \in \Sigma_R(n)\}$
$\Sigma_S(n+1) = \{\sigma \in \Sigma_S(n) : \sigma \text{ is not strongly dominated with respect to } S_R(n) \text{ by any } \sigma' \in \Sigma_S(n)\}$
Taking the limit we can define $S_R^{\mathcal{X}} := \lim_{n\to\infty} S_R(n)$ *and* $\Sigma_S^{\mathcal{X}} = \lim_{n\to\infty} \Sigma_S(n)$.

This definition takes only an ex ante perspective. Moreover, it does not take into account the exogenous restrictions on players' beliefs. These two remarks suggest that the definition 38 is not stricter or is not related with the definition of $\Delta - rationalizable$ messages.

[14] Given that $\Delta_R'' \supseteq \Delta_R'$, condition 37 implies that $S_R(2, \Delta'') = S_R(2, \Delta')$.

[15] Remember that $\Theta^c := \Theta \backslash \bigcup_{i=1}^{D} \hat{\Theta}_i$.

Before stating our next results we need to recall one last definition.
A strategy σ_S, belonging to the set of pure strategy Σ_S, is weakly dominated by a mixed strategy $\sigma'_S \in \Delta(\Sigma_S)$ for type θ with respect to S_R if

$$\forall s_R \in S_R,\ u_S(\theta, \sigma_S, s_R) \leq u_S(\theta, \sigma'_S, s_R)$$

and

$$\exists s_R \in S_R,\ u_S(\theta, \sigma_S, s_R) < u_S(\theta, \sigma'_S, s_R)$$

The definition of strict dominance is analogous except that all weak inequalities are replaced by strict inequalities. The same definitions can be applied to *S*ender's strategy. Shimoji and Watson (1998) characterized rationalizability in extensive form game in term of iterative elimination of conditionally dominated strategies but they do not consider exogenous restrictions on players' belief.[16]

For any given subset $B \subseteq \Sigma_S \times S_R$ let $\mathcal{W}(B)$ $(\mathcal{S}(B))$ denote the set of $(\sigma_S, s_R) \in \Sigma_S \times S_R$ such that for $\sigma_S(\theta)$ is not weakly (strictly) dominated for θ in S_R and s_R is not weakly (strictly) dominated in Σ_S. Let $\mathcal{SW}(B) := \mathcal{S}(B) \cap \mathcal{W}(\Sigma_S \times S_R)$. The iterated operator $\mathcal{SW}^n$ is defined in the usual way: $\mathcal{SW}^n(B) = \mathcal{SW}(\mathcal{SW}^{n-1}(B))$ where $\mathcal{SW}^0(B) = B$.
The iterative procedure defined by Rabin, the Message Profile Theory, is equivalent to the set $\mathcal{S}^\infty\left(\hat{\Sigma}_S \times \hat{S}_R\right)$, where:
$\hat{\Sigma}_S = \left\{(\theta, m) \in \Sigma_S : \text{if } \theta \in \hat{\Theta} \text{ where } \hat{\Theta} \in \mathcal{X} \text{ then } m = \left[\hat{\Theta}\right] \text{ and if } \theta \notin \hat{\Theta}_k \cup \Theta\left(\left[\hat{\Theta}_k\right], \mathcal{X}\right) \text{ then } m \neq \left[\hat{\Theta}_k\right]\right\}$ and
$\hat{S}_R := \left\{s_R \in S_R : s_R\left(\left[\hat{\Theta}\right]\right) \in A^*\left(\left[\hat{\Theta}\right]\right) \text{ for every } \left[\hat{\Theta}\right] \in \mathcal{X} \text{ and } s_R\left(\left[\hat{\Theta}\right]\right) \in A^* \text{ otherwise}\right\}$.

Proposition 1 by Rabin state the main property of the set of strategy profiles obtained using his definition of rationale message profile.

Proposition 10 *(Rabin 1990) If $\mathcal{X}$ is a CMP then $MPT(\mathcal{X}) = \Sigma_S^{\mathcal{X}} \times S_R^{\mathcal{X}}$ satisfies: $\forall \sigma \in \Sigma_S^{\mathcal{X}}$, σ is not strongly dominated with respect to $S_R^{\mathcal{X}}$ by any $\sigma' \in \Sigma_S$. Likewise, $\forall \gamma \in S_R^{\mathcal{X}}$, γ is not strongly dominated with respect to $\Sigma_S^{\mathcal{X}}$ by any $\gamma' \in S_R$.*

Now, we can try to relate the result by Rabin to the more general concept of weakly $\Delta - rationalizability$. Remember that the following restriction for the

[16] The characterization of $\Delta - rationalizability$ has been investigated in Cappelletti (2004) [12] among others.

first order beliefs, $\Delta' = \Delta(S_R) \times \Delta_R$, holds:

$$\Delta'_R = \{(\mu_R(\cdot|\phi), \mu_R(\cdot|m)_{m\in M}) \in \Delta^*(\Sigma_S) : \text{marg}_\Theta \mu_R(\cdot|\phi) = p(\cdot)$$
$$\text{and } \mu_R\left(\theta|\left[\hat{\Theta}\right]\right) = \frac{p(\theta)}{\sum_{\theta'\in[\hat{\Theta}]} p(\theta')} \text{ if } \theta \in \left[\hat{\Theta}\right], \mu_R\left(\theta|\left[\hat{\Theta}\right]\right) = 0 \text{ otherwise for every } \left[\hat{\Theta}\right] \in M \}$$

Proposition 11 *If $\hat{\Sigma}_S \times \hat{S}_R$ is credible then $MPT(\mathcal{X}) = \Sigma_S^{\mathcal{X}} \times S_R^{\mathcal{X}}$ is (weakly) $\Delta' -$ rationalizable.*

Proof. Given the characterization of weakly $\Delta - rationalizability$ as the the set of payoff-strategy pairs belonging to $\mathcal{SW}^\infty(\Sigma_S \times S_R)$ (see among others Ben Porath (1997)). Hence Proposition 11 is equivalent to proving $\mathcal{S}^\infty\left(\hat{\Sigma}_S \times \hat{S}_R\right) \subseteq \mathcal{SW}^\infty(\Sigma_S \times S_R)$

On the basis of the classic equivalence between never best response and undominance we can deduce that:

$$\pi_R^{[\Theta]}(\mathcal{W}(\Sigma_S \times S_R)) = A^*([\Theta]) \text{ for every } [\Theta] \in M$$

where $\pi_R^{[\Theta]}$ is the natural projection on the set of the *R*eceriver's strategies given message $[\Theta]$.

Now, let us consider the *S*ender:

$$\pi_S^\theta(\mathcal{W}(\Sigma_S \times S_R)) \supseteq \hat{\Sigma}_S$$

given that messages are cheap. Therefore, we can conclude that:

$$\mathcal{S}[\mathcal{W}(\Sigma_S \times S_R) \times \mathcal{W}(\Sigma_S \times S_R)] \supseteq \mathcal{S}\left(\hat{\Sigma}_S \times \hat{S}_R\right). \tag{3.1}$$

The operator $\mathcal{S}(\cdot)$ is weakly decreasing therefore, we have proved the Proposition. ■

Given that Rabin's analysis does not explicitly rely on any kind of forward reasoning weak $\Delta - rationalizability$ is closer to his definition than strong $\Delta - rationalizabiliy$.

Now, we can qualify Proposition 11 remarking that the characterization proposed by Rabin is at least stronger than the one deduced by assuming rationality, common knowledge of rationality and that *R*eceiver's belief belongs to Δ'_R and this fact is common knowledge (at the beginning of the game).

In order to relate his definition to the stronger concept of strong $\Delta-$*rationalizability* we need to ensure that the set of strongly $\Delta'-$*rationalizable* strategy profiles is not empty. Note that previous existence results (see Battigalli and Siniscalchi (2003) [7]) do not apply to this setting.

Proposition 12 *If $\hat{\Sigma}_S \times \hat{S}_R$ is credible and $\bigcup_{\Theta_k \in \mathcal{X}} \Theta_k = \Theta$ then the set $\Delta'-$rationalizable payoff/message pairs and strategies is non empty.*

Proposition 13 *If $\hat{\Sigma}_S \times \hat{S}_R$ is credible and $\bigcup_{\Theta_k \in \mathcal{X}} \Theta_k = \Theta$ then $MPT(\mathcal{X}) = \Sigma_S^{\mathcal{X}} \times S_R^{\mathcal{X}}$ is $\Delta'-$rationalizable.*

Proof. See Appendix 3.C. ■

Without the additional assumption the two procedures could lead to dramatically different results. Let us show it by means of an example.

Suppose that $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ and $\mathcal{X} = \{[\theta_1], [\theta_2]\}$ is a credible message profile. Let us consider the iterative definition of $\Delta-$*rationalizability*:

(Step 1) $S_R(1, \Delta') = \left\{ s_R \in S_R : s_R\left(\left[\hat{\Theta}\right]\right) = \arg\max_{a' \in A} \int_{\hat{\Theta}} u_R(a', \theta) p(\theta) d\theta \text{ for every } \left[\hat{\Theta}\right] \in M \right\}$ where this set is equal to $\hat{S}_R$.
$\Sigma_S(1, \Delta') = \Sigma_S$ because the message are payoff irrelevant;

(Step 2) $S_R(2, \Delta') = S_R(1, \Delta')$ because for any message $\left[\hat{\Theta}\right]$ the set of types that could have sent the message is Θ, namely $\Theta\left(\left[\hat{\Theta}\right], k-1, \Delta\right) = \Theta$ for every message $\left[\hat{\Theta}\right]$.
$\Sigma_S(2, \Delta')$ contains the following payoff/message pairs $\{(\theta_1, [\theta_1]), (\theta_2, [\theta_2])\}$ because of condition 37.

(Step 3) $S_R(3, \Delta')$ could be empty. Let us consider the message $[\{\theta_1, \theta_2\}]$, $\Theta([\{\theta_1, \theta_2\}], 3, \Delta) \subseteq \{\theta_3, \theta_4\}$ because both θ_1 and θ_2 could guarantee their most preferred action choosing respectively $[\theta_1]$ and $[\theta_2]$.[17] But the requirements that $\mu \in \Delta'_R$ and $\mu(\Theta([\{\theta_1, \theta_2\}], 3, \Delta) \mid [\{\theta_1, \theta_2\}]) = 1$ are not compatible as long as $\Theta([\{\theta_1, \theta_2\}], 3, \Delta)$ is not empty. Therefore, $\Delta'-$*rationalizability* with the restriction used by Rabin could lead to an empty selection. Instead, Definition 10 does not. Therefore in non generic games the two definitions do not coincide.[18]

[17] Note that $u_S(\theta_1, \cdot) \neq u_S(\theta_2, \cdot)$ or $u_R(\theta_1, \cdot) \neq u_R(\theta_2, \cdot)$.
[18] In non generic games the two definitions do not intersect.

Given Proposition 11 and Proposition 13, we can conclude that our previous results allow to fully evaluate the procedure proposed by Rabin and to make precise Proposition 10. Moreover, the solution concept proposed fully reflects the usual epistemic assumptions without imposing implicit hypothesis and can be applied even if common prior assumption does not hold.

3.6 Conclusions

With this paper we have tried to characterize rational behavior in a communication game between two agents. Where one agent has some private information and direct communication by means of cheap talk is possible. Our results hint that neither common interests between agents nor incentive for the informed party to reveal information are sufficient for full disclosure to occur. As an interesting conclusion, our result is even stronger than the one found by Crawford and Sobel (1982). Perfect communication is not to be expected in general even if agents' interests almost coincide.

Moreover, given that we are using a non-equilibrium analysis we can investigate plainly lying, credibility and credulity which are essential features of strategic communication.

Given our framework, a worthwhile extension would be to analyze the effect of self-confidence in communication games. Namely, what if the informed party is convinced that he could fool his opponent. A first attempt in this direction has been pursued by Crawford (2003)[15].

As a matter of fact, this framework can also model active misrepresentation of intentions to opponents allowing for different epistemic assumption. For our results we have assumed the highest degree of sophistication but we can allow for a certain sort of bound rationality.

A different extension can be the analysis of situations where there is bilateral incomplete information and bilateral communication (see Krishna and Morgan (2003) [37]) or where all agents can take payoff relevant actions after the communication stage as in auction (see Benoit and Dubra (2004) [8]).

3.A Some preliminary results

Given that Θ is finite, we can index its element starting from the smallest element $\Theta = \{\theta_1, \theta_2, ..., \theta_K\}$. Given any message $\left[\hat{\Theta}\right]$ we define the set of conceivable best responses of the *R*eceiver:

$$A^*\left(\left[\hat{\Theta}\right]; k\right) := \left\{a \in A : \exists s_R \in S_R(k, \Delta) \text{ such that } a = s_R\left(\left[\hat{\Theta}\right]\right)\right\}$$

Given a simple or sharp message $[\theta]$, we define the following two "boundary" actions:

$$\begin{aligned} \underline{a}([\theta], k) &:= \arg\min_{a \in A^*([\theta], k)} u_S(a, \theta') \\ \overline{a}([\theta], k) &:= \arg\max_{a \in A^*([\theta], k)} u_S(a, \theta') \end{aligned}$$

$\underline{a}([\theta], k)$ represents the worst conceivable action for the *S*ender's type θ' when she chooses message $[\theta]$,while $\overline{a}([\theta], k)$ is the best conceivable action for θ' when he chooses message $[\theta]$.

Lemma 8 *There are no $\theta < \theta'$ such that*

$$a_R^*(\theta) < a_R^*(\theta') < a_S^*(\theta) \tag{3.2}$$

or

$$a_S^*(\theta) < a_S^*(\theta') < a_R^*(\theta) < a_R^*(\theta') \tag{3.3}$$

Proof. Suppose that (3.2) holds then

$$d(a_R^*(\theta), a_R^*(\theta')) < d(a_R^*(\theta), a_S^*(\theta))$$

but $\varepsilon < d(a_R^*(\theta), a_R^*(\theta'))$ and $d(a_R^*(\theta), a_S^*(\theta)) < \sigma$. Knowing that $\varepsilon > \sigma$ we obtain a contradiction. The same way of reasoning applies to (3.3). ■

Now, we impose a more stringent assumption on the beliefs of the *R*eceiver:

$$\text{if } \left|\hat{\Theta}\right| = 1 \text{ then } \mu_R(\theta | [\theta]) = 1 \tag{3.4}$$

This condition states that if the *R*eceiver sees a sharp message the he will trust completely it. This additional assumption simplifies the iterative procedure and implies that truthful disclosure is rational.

Proposition 14 *If the conditions (σ – alignment), (ε – sorting), (Trust), (Cautiousness) and (3.4) hold and $\mu_R(\bullet|[\Theta])$ is uniform on Θ then Δ–rationalizability implies truthful disclosure.*

Proof. Take the $\underline{\theta} := \theta_1 \in \Theta$ such that $a_R^*(\underline{\theta})$ or $a_S^*(\underline{\theta})$ is minimum among the set of optimal actions, $\{a_R^*(\theta_k), a_S^*(\theta_k)\}_{k=1}^K$ then, we can order the other optimal choices on the basis of their distance[19] from $a_R^*(\underline{\theta})$ (see Figure 4).

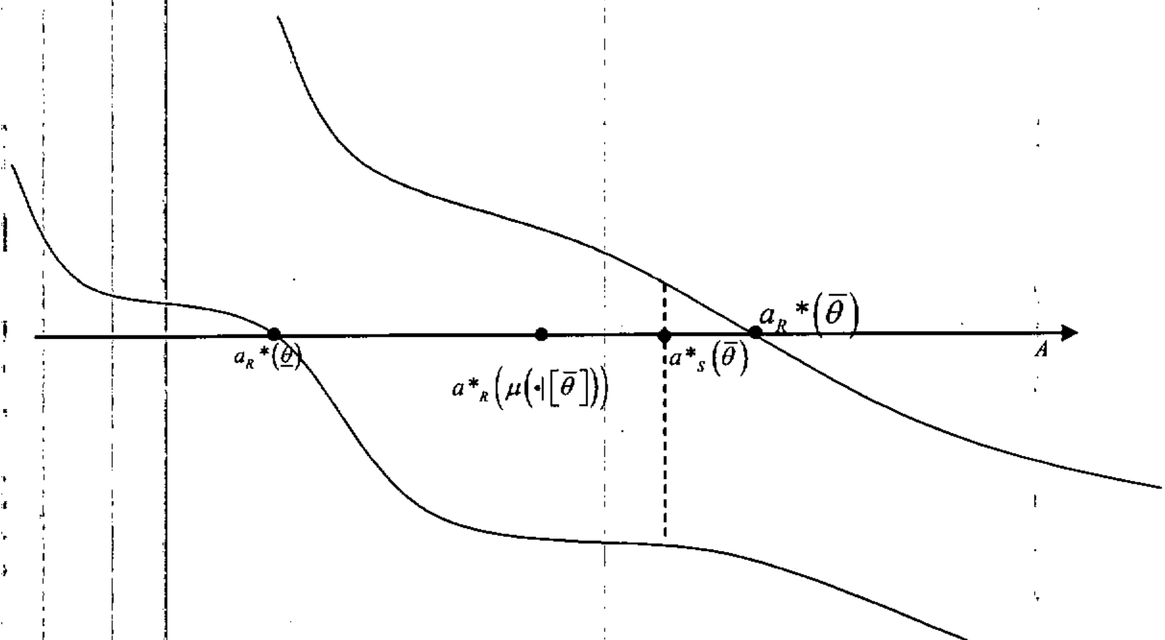


Figure 4

If we define $\Sigma_S(0,\Delta) = \Sigma_S$ and $S_R(0,\Delta) = S_R$, then for $k = 1, 2, ...$ let $\Theta(m, k-1, \Delta) := \{\theta : (\theta, m) \in \Sigma_S(k-1, \Delta)\}$. $\Theta(m, k-1, \Delta)$ is the set of payoff-types consistent with message m and a level of sophistication of the Sender at least equal to $k-1$ then

$$\begin{aligned}
\Sigma_S(k,\Delta) \quad &: = \{(\theta, m) \in \Sigma_S(k-1,\Delta) : \exists \mu_S \in \Delta_S, m \in BR_S(\theta, \mu_S) \text{ and } \mu_S(S_R(k-1,\Delta)) = 1\} \\
S_R(k,\Delta) \quad &: = \{s_R \in S_R(k-1,\Delta) : \exists \mu_R \in \Delta_R, \forall m, s_R(m) \in BR_R(m, \mu_R(\cdot|m)), \\
&\quad \text{and } \Theta(m, k-1, \Delta) \neq \emptyset \text{ implies } \mu_R(\Theta(m, k-1, \Delta)|m) = 1\}
\end{aligned}$$

[19] This follows from strict concavity

Therefore, at the first stage of the iterative procedure we obtain the following sets:

$$\begin{aligned}\Sigma_S(1,\Delta) &: = \{(\theta,m)\in\Sigma_S(0,\Delta):\exists\mu_S\in\Delta(S_R), m\in BR_S(\theta,\mu_S) \text{ and } \mu_S(S_R(0,\Delta))=1\}\\ S_R(1,\Delta) &: = \{s_R\in S_R(0,\Delta):\exists\mu_R\in\Delta_R,\forall m, s_R(m)\in BR_R(m,\mu_R(\cdot|m)),\\ &\quad \text{and given that } \Theta(m,0,\Delta)=\Theta \text{ for every } m\in M,\ \mu_R(\Theta(m,k-1,\Delta)|m)=1\}\end{aligned}$$

Take $\underline{\theta}$ and suppose that $a_S^*(\underline{\theta}) > a_R^*(\underline{\theta})$, now we consider all the possible messages for $\underline{\theta}$ and prove that there is a strictly dominant message namely, $[\underline{\theta}]$. If $\underline{\theta}$ send message $[\underline{\theta}]$ then the only $\Delta-$rational action for the Receiver is $a_R^*(\underline{\theta})$. Instead, if the message is $\left[\underline{\theta}\cup\hat{\Theta}\right]\neq[\Theta]$ we need to prove that $a_S^*(\underline{\theta}) < a_R^*\left(\mu\left(\cdot|\left[\underline{\theta}\cup\hat{\Theta}\right]\right)\right)$ for all $\mu\in\Delta_R$. As a matter of fact, $a_S^* < a_R^*\left(\mu\left(\cdot|\left[\underline{\theta}\cup\hat{\Theta}\right]\right)\right)$ is equivalent to:

$$\sum_{\theta'\in\Theta}\mu\left(\theta'|\left[\underline{\theta}\cup\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*(\underline{\theta}),\theta'\right)}{\partial a} > 0$$

or equivalently

$$\begin{aligned}\mu\left(\underline{\theta}|\left[\underline{\theta}\cup\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*(\underline{\theta}),\underline{\theta}\right)}{\partial a} &> -\sum_{\theta'\in\hat{\Theta}}\mu\left(\theta'|\left[\underline{\theta}\cup\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*(\underline{\theta}),\theta'\right)}{\partial a}+\\ &\quad -\sum_{\theta'\in\Theta\setminus\{\underline{\theta}\cup\hat{\Theta}\}}\mu\left(\theta'|\left[\underline{\theta}\cup\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*(\underline{\theta}),\theta'\right)}{\partial a}.\end{aligned}$$

Given $\frac{\partial u_S(a_S^*(\underline{\theta}),\underline{\theta})}{\partial a}=0$ and assumption ($\sigma-alignment'$) we can conclude that $\mu\left(\underline{\theta}|\left[\underline{\theta}\cup\hat{\Theta}\right]\right)\frac{\partial u_R(a_S^*(\underline{\theta}),\underline{\theta})}{\partial a} > -\sigma$.
Given assumptions (Cautiousness), ($\varepsilon-sorting$) and the supermodularity of the Receiver's utility function, we can infer that $\sum_{\theta'\in\hat{\Theta}}\mu\left(\theta'|\left[\underline{\theta}\cup\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*(\underline{\theta}),\theta'\right)}{\partial a}\geq k\varepsilon$ and the conclusion follows from the assumption that $\sigma<\kappa\varepsilon$.

Similarly, if the message sent is $\left[\hat{\Theta}\right]\neq[\Theta]$, where $\underline{\theta}\notin\hat{\Theta}$, then we can conclude that $a_S^* < a_R^*\left(\mu\left(\cdot|\left[\hat{\Theta}\right]\right)\right)$ for all $\mu\in\Delta_R$, because this is equivalent to:

$$\sum_{\theta'\in\Theta}\mu\left(\theta'|\left[\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*(\underline{\theta}),\theta'\right)}{\partial a} > 0$$

or equivalently

$$\mu\left(\underline{\theta}|\left[\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*\left(\underline{\theta}\right),\underline{\theta}\right)}{\partial a}>-\sum_{\theta'\in\hat{\Theta}}\mu\left(\theta'|\left[\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*\left(\underline{\theta}\right),\theta'\right)}{\partial a}-\sum_{\theta'\in\Theta\setminus\{\underline{\theta}\cup\hat{\Theta}\}}\mu\left(\theta'|\left[\underline{\theta}\cup\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*\left(\underline{\theta}\right),\theta'\right)}{\partial a}$$

which follows from the assumption $\sigma<\kappa\varepsilon$. For the message $[\Theta]$ it is sufficient that $\mu\left(\theta|\left[\Theta\right]\right)>\kappa$ for some $\theta\neq\underline{\theta}$.

Note that $\sum_{\theta'\in\Theta}\mu\left(\theta'|\left[\hat{\Theta}\right]\right)\left[\frac{\partial u_R\left(a_S^*(\underline{\theta}),\theta'\right)}{\partial a}-\frac{\partial u_R\left(a_R^*\left(\mu\left(\cdot|\left[\underline{\theta}\cup\hat{\Theta}\right]\right)\right),\theta'\right)}{\partial a}\right]>\sigma$ then by sublinearity of the Receiver's marginal utility we can conclude that $d\left(a_S^*\left(\underline{\theta}\right),a_R^*\left(\mu\left(\cdot|\left[\underline{\theta}\cup\hat{\Theta}\right]\right)\right)\right)>\sigma>d\left(a_S^*\left(\underline{\theta}\right),a_R^*\left(\underline{\theta}\right)\right)$and hence $a_R^*\left(\underline{\theta}\right)$ is preferred to $a_R^*\left(\mu\left(\cdot|\left[\underline{\theta}\cup\hat{\Theta}\right]\right)\right)$.

Suppose that $a_S^*\left(\underline{\theta}\right)<a_R^*\left(\underline{\theta}\right)$ then the message $[\underline{\theta}]$ is trivially dominant.

Therefore, for any message $\left[\hat{\Theta}\right]\neq[\underline{\theta}]$ we can conclude that $\underline{\theta}\notin\Theta\left(\left[\hat{\Theta}\right],2,\Delta\right)$.
Assume that $\mu\left(\cdot|\left[\Theta\right]\right)$ is uniform on Θ and that $\theta_l\notin\Theta\left(\left[\hat{\Theta}\right],l+1,\Delta\right)$ for every $\left[\hat{\Theta}\right]\neq[\theta_l]$ for $l=1,2,...k$. Take θ_{k+1} and $\Theta\left(\left[\hat{\Theta}\right],k+1,\Delta\right)$and assume that $a_S^*\left(\theta_{k+1}\right)>a_R^*\left(\theta_{k+1}\right)$. If θ_{k+1} send message $[\theta_{k+1}]$ then the only $\Delta-$rational action for the Receiver is $a_R^*\left(\theta_{k+1}\right)$.
Instead, if the message sent is $\left[\theta_{k+1}\cup\hat{\Theta}\right]\neq[\Theta]$ we need to prove that $a_S^*\left(\theta_{k+1}\right)<a_R^*\left(\mu\left(\cdot|\left[\theta_{k+1}\cup\hat{\Theta}\right]\right)\right)$ for all $\mu\in\Delta_R$. As a matter of fact, $a_S^*\left(\theta_{k+1}\right)<a_R^*\left(\mu\left(\cdot|\left[\theta_{k+1}\cup\hat{\Theta}\right]\right)\right)$ is equivalent to:

$$\sum_{\theta'\in\Theta}\mu\left(\theta'|\left[\theta_{k+1}\cup\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*\left(\theta_{k+1}\right),\theta'\right)}{\partial a}>0$$

or equivalently

$$\mu\left(\theta_{k+1}|\left[\theta_{k+1}\cup\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*\left(\theta_{k+1}\right),\theta_{k+1}\right)}{\partial a}>-\sum_{\theta'\in\hat{\Theta}}\mu\left(\theta'|\left[\theta_{k+1}\cup\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*\left(\theta_{k+1}\right),\theta'\right)}{\partial a}$$
$$-\sum_{\theta'\in\Theta\setminus\{\theta_{k+1}\cup\hat{\Theta}\}}\mu\left(\theta'|\left[\theta_{k+1}\cup\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*\left(\underline{\theta}\right),\theta'\right)}{\partial a}\tag{3.5}$$

Since $\{\theta_1,\theta_2,...,\theta_k\}\notin\Theta\left(\left[\theta_{k+1}\cup\hat{\Theta}\right],2,\Delta\right)$ then $\mu\left(\theta_l|\left[\theta_{k+1}\cup\hat{\Theta}\right]\right)=0$ for

$l = 1, 2, ..., k$ therefore the inequality (3.5) is equivalent to:

$$\mu\left(\theta_{k+1}|\left[\theta_{k+1}\cup\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*\left(\theta_{k+1}\right),\theta_{k+1}\right)}{\partial a} > -\sum_{\theta'\in\hat{\Theta}\setminus\{\theta_1,\theta_2,...,\theta_k\}}\mu\left(\theta'|\left[\theta_{k+1}\cup\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*\left(\theta_{k+1}\right),\theta'\right)}{\partial a}$$
$$-\sum_{\theta'\in\Theta\setminus\left\{\theta_{k+1}\cup\hat{\Theta}\right\}\cup\{\theta_1,\theta_2,...,\theta_k\}}\mu\left(\theta'|\left[\theta_{k+1}\cup\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*\left(\underline{\theta}\right),\theta'\right)}{\partial a}$$

Given $\frac{\partial u_S(a_S^*(\theta_{k+1}),\theta_{k+1})}{\partial a} = 0$ and assumption ($\sigma - alignment'$) we can conclude that $\mu\left(\theta_{k+1}|\left[\underline{\theta}\cup\hat{\Theta}\right]\right)\frac{\partial u_R(a_S^*(\theta_{k+1}),\theta_{k+1})}{\partial a} \geq -\sigma$. Given that $\mu\left(\theta_l|\left[\theta_{k+1}\cup\hat{\Theta}\right]\right) = 0$ for $l = 1, 2, ..., k$ and the supermodularity of the *R*eceiver's utility function:

$$-\sum_{\theta'\in\Theta\setminus\left\{\theta_{k+1}\cup\hat{\Theta}\right\}\cup\{\theta_1,\theta_2,...,\theta_k\}}\mu\left(\theta'|\left[\theta_{k+1}\cup\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*\left(\underline{\theta}\right),\theta'\right)}{\partial a} >$$
$$-\sum_{\theta'\in\hat{\Theta}\setminus\{\theta_1,\theta_2,...,\theta_k\}}\mu\left(\theta'|\left[\theta_{k+1}\cup\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*\left(\theta_{k+1}\right),\theta'\right)}{\partial a}$$

Given assumptions (Trust), ($\varepsilon-sorting$), we can infer that $\sum_{\theta'\in\theta\setminus\hat{\Theta}}\mu\left(\theta'|\left[\underline{\theta}\cup\hat{\Theta}\right]\right)\frac{\partial u_R(a_S^*(\underline{\theta}),\theta')}{\partial a} \geq k\varepsilon$, and the inequality follows from the assumption that $\sigma < \kappa\varepsilon$. If the message sent is $\left[\hat{\Theta}\right] \neq [\Theta]$, where $\underline{\theta} \notin \hat{\Theta}$, then we can conclude that $a_S^* < a_R^*\left(\mu\left(\cdot|\left[\hat{\Theta}\right]\right)\right)$ for all $\mu \in \Delta_R$, because this is equivalent to:

$$\sum_{\theta'\in\Theta}\mu\left(\theta'|\left[\hat{\Theta}\right]\right)\frac{\partial u_R\left(a_S^*\left(\underline{\theta}\right),\theta'\right)}{\partial a} > 0$$

Following a similar reasoning this inequality follows from $\sigma < \kappa\varepsilon$.

If $\mu\left(\cdot|[\Theta]\right)$ is uniform on Θ and $\{\theta_1, \theta_2, ..., \theta_k\} \notin \Theta\left(\left[\theta_{k+1}\cup\hat{\Theta}\right], 2, \Delta\right)$ then $a_R^*\left(\mu\left(\cdot|[\Theta]\right)\right)$ is such that $a_S^*\left(\theta_{k+1}\right) < a_R^*\left(\mu\left(\cdot|[\Theta]\right)\right)$. If $a_S^*\left(\underline{\theta}\right) < a_R^*\left(\underline{\theta}\right)$ then the message $[\underline{\theta}]$ is trivially dominant.
Therefore, for any message $\left[\hat{\Theta}\right] \neq [\underline{\theta}]$ we can conclude that $\underline{\theta} \notin \Theta\left(\left[\hat{\Theta}\right], 2, \Delta\right)$. This proves the last claim. ■

3.B Proof of Proposition 9

The proof of Proposition 9 is obtained by means of two lemmas and applying the previous result (Proposition 8).

Lemma 9 *Given a set of payoff types Θ there exists $K \in \mathbb{N}$ such that $\underline{\theta} \notin \Theta([\theta], K, \Delta)$ for $\theta \in \Theta \backslash \{\underline{\theta}\}$ and $\overline{\theta} \notin \Theta([\theta], K, \Delta)$ for $\theta \in \Theta \backslash \{\overline{\theta}\}$, where $\underline{\theta} := \arg\min_{\theta\in\Theta} \{a_R^*(\theta)\}$ and $\overline{\theta} := \arg\max_{\theta\in\Theta} \{a_R^*(\theta)\}$.*

Proof. The proof is by induction on the cardinality of the set Θ.

(Step 1) Take $\Theta = \{\theta_1, \theta_2\}$, where wlog $\theta_1 < \theta_2$. Given the assumption on the cross derivative of the utility function $\underline{\theta} = \theta_1$ and $\overline{\theta} = \theta_2$.
Let's take the Sender's type $\overline{\theta}$, there are just two "simple" messages available $[\underline{\theta}]$ and $[\overline{\theta}]$. Consider $\underline{a}([\underline{\theta}], 0)$ and $\overline{a}([\overline{\theta}], 0)$, briefly $\underline{a}$ and $\overline{a}$.
Since $\underline{a} \in A^*([\underline{\theta}], 0)$, $\underline{a}$ satisfies the following condition:

$$\mu(\underline{\theta}|[\overline{\theta}])\frac{\partial u_R(\underline{a},\underline{\theta})}{\partial a} + \mu(\overline{\theta}|[\overline{\theta}])\frac{\partial u_R(\underline{a},\overline{\theta})}{\partial a} = 0 \tag{3.6}$$

where $\mu \in \Delta_R$
Moreover, $\underline{a}$ maximizes $d(a_S^*(\overline{\theta}), a)$ then $\mu(\underline{\theta}|[\overline{\theta}])$ must be as high as possible. Given Assumption (Trust) $\mu_R(\overline{\theta}|[\overline{\theta}]) > \mu_R(\underline{\theta}|[\overline{\theta}])$ which implies $\frac{1}{2} > \mu_R(\underline{\theta}|[\overline{\theta}])$.

Similarly, $\overline{a}$ satisfies the following condition:

$$\mu(\underline{\theta}|[\underline{\theta}])\frac{\partial u_R(\overline{a},\underline{\theta})}{\partial a} + \mu(\overline{\theta}|[\underline{\theta}])\frac{\partial u_R(\overline{a},\overline{\theta})}{\partial a} = 0$$

where $\mu \in \Delta_R$
Given our assumption on the utility function $\overline{a}$ should minimize $d(a_S^*(\overline{\theta}), a)$ that is $\mu(\underline{\theta}|[\underline{\theta}])$ must be as lower as possible. Given Assumption (Trust) $\mu(\underline{\theta}|[\underline{\theta}]) > \mu_R(\overline{\theta}|[\underline{\theta}])$ which implies $\frac{1}{2} < \mu(\underline{\theta}|[\underline{\theta}])$. Therefore, when $\underline{a}$ and $\overline{a}$ are achieved the associated conditional beliefs are such that $\mu_R(\underline{\theta}|[\overline{\theta}]) < \mu(\underline{\theta}|[\underline{\theta}])$.
Equation (3.6) can be written as:

$$\mu(\underline{\theta}|[\overline{\theta}])\left[\frac{\partial u_R(\underline{a},\underline{\theta})}{\partial a} - \frac{\partial u_R(\underline{a},\overline{\theta})}{\partial a}\right] + \frac{\partial u_R(\underline{a},\overline{\theta})}{\partial a} = 0$$

First, we define $g(a;\mu(\underline{\theta}|[\cdot])) := \mu(\underline{\theta}|[\cdot])\left[\frac{\partial u_R(a,\underline{\theta})}{\partial a} - \frac{\partial u_R(a,\overline{\theta})}{\partial a}\right] + \frac{\partial u_R(a,\overline{\theta})}{\partial a}$ then, we can easily note that $g(a)$ is strictly decreasing in a. Given that

$\frac{\partial u_R(\overline{a},\underline{\theta})}{\partial a} - \frac{\partial u_R(\overline{a},\overline{\theta})}{\partial a} < 0$, if $\mu\left(\underline{\theta}\middle|\left\lceil\theta\right\rceil\right) < \mu\left(\underline{\theta}\middle|\left\lfloor\theta\right\rfloor\right)$ then $g\left(a;\mu\left(\underline{\theta}\middle|\left\lfloor\theta\right\rfloor\right)\right) < g\left(a;\mu\left(\underline{\theta}\middle|\left\lceil\theta\right\rceil\right)\right)$. Therefore, if $\underline{a}$ is such that $g\left(\underline{a};\mu\left(\underline{\theta}\middle|\left\lfloor\theta\right\rfloor\right)\right) = 0$ then $g\left(\underline{a};\mu\left(\underline{\theta}\middle|\left\lceil\theta\right\rceil\right)\right) > 0$ and, given that g is decreasing, $\overline{a} > \underline{a}$. Hence, $\overline{\theta}$ strictly prefers message $\left\lceil\theta\right\rceil$ to message $\left\lfloor\theta\right\rfloor$. The same way of reasoning applies if we consider type $\underline{\theta}$. Consequently $\overline{\theta} \notin \Theta\left(\left\lfloor\theta\right\rfloor, 1, \Delta\right)$ and $\underline{\theta} \notin \Theta\left(\left\lceil\theta\right\rceil, 1, \Delta\right)$.

Assume that given $\Theta = \{\theta_1, \theta_2 ..., \theta_k\}$, where $\theta_1 < \theta_2 < ... < \theta_k$, $\theta_1 \notin \Theta\left([\theta], K, \Delta\right)$ for $\theta \in \Theta \backslash \{\theta_1\}$ and $\theta_k \notin \Theta\left([\theta], K, \Delta\right)$ for $\theta \in \Theta \backslash \{\theta_k\}$.
(Step k+1) Take Θ and suppose to add a payoff type θ_{k+1} such that $\theta_{k+1} > \theta_k$. Applying the same reasoning as before we can prove that for $\underline{\theta} = \theta_1$ the message $\left\lceil\theta\right\rceil$, where $\overline{\theta} = \theta_{k+1}$ is strictly dominated by $\left\lfloor\theta\right\rfloor$ and symmetrically for $\overline{\theta} = \theta_{k+1}$ the message $\left\lfloor\theta\right\rfloor$, where $\overline{\theta} = \theta_{k+1}$ is strictly dominated by $\left\lceil\theta\right\rceil$.

If we consider the sets $\Theta\left(\left\lfloor\theta\right\rfloor, 1, \Delta\right)$ by the inductive hypothesis we can conclude that $\underline{\theta} \notin \Theta\left([\theta], K+1, \Delta\right)$ for $\theta \in \Theta\left(\left\lfloor\theta\right\rfloor, 1, \Delta\right) \backslash \{\underline{\theta}\}$, therefore $\underline{\theta} \notin \Theta\left([\theta], K+1, \Delta\right)$ for $\theta \in \Theta \backslash \{\underline{\theta}\}$. The same reasoning applies to $\overline{\theta}$ w.r.t. $\Theta\left(\left\lceil\theta\right\rceil, 1, \Delta\right)$. ■

Lemma 10 *Given a finite payoff type set Θ, there exists $K' \in \mathbb{N}$ such that $\Theta\left([\theta], K', \Delta\right) = \theta$ for every $\theta \in \Theta$.*

Proof. Take the set $\Theta = \{\theta_1, \theta_2, ..., \theta_k\}$ where $\theta_1 < \theta_2 < ... < \theta_k$. Given Lemma 9 $\theta_1 \notin \Theta\left([\theta], K, \Delta\right)$ for $\theta \in \Theta \backslash \{\theta_1\}$ and $\theta_k \notin \Theta\left([\theta], K, \Delta\right)$ for $\theta \in \Theta \backslash \{\theta_k\}$.

Take θ_2, the message $[\theta_2]$ excludes θ_1 and θ_k as possible *S*ender of the message (given the assumption that players are at least K order sophisticated). Then consider $\Theta \backslash \{\theta_1\}$ then Lemma 9 still holds hence $\theta_2 \notin \Theta\left([\theta], K_{(2)}, \Delta\right)$ for $\theta > \theta_2$, where $K_{(2)} := K+1$. If we consider θ_{k-1},by the same reasoning we can deduce that $\theta_{k-1} \notin \Theta\left([\theta], K^{(k-1)}, \Delta\right)$ for $\theta < \theta_{k-1}$ where $K^{(k-1)} := K+1$. Iterating the same reasoning, let the θ's vary across all Θ we can conclude that for each $\hat{\theta}_l \in \Theta$ $\hat{\theta} \notin \Theta\left([\theta], \hat{K}_l, \Delta\right)$ for all $\theta < \hat{\theta}_l$ and $\theta > \hat{\theta}_l$, where $\hat{K}_l = \max\left\{K^{(l)}, K_{(k-l)}\right\}$. If we consider $\hat{K} := \max\left\{\hat{K}_l\right\}$ we have proved the lemma. ■

Proposition 15 (9) *If the conditions ($\sigma -$ alignment), ($\varepsilon -$ sorting), (Trust) and (Cautiousness) hold $\Delta -$ rationalizability implies that truthful disclosure is rational.*

Proof. By means of Lemma 10 we can conclude that $\Theta\left([\theta], K', \Delta\right) = \theta$ for all $\theta \in \Theta$. Then for every $\mu \in \Delta_R$ the conditional belief after receiving a message $[\theta]$ is such that $\mu\left(\theta \middle| [\theta]\right) = 1$. Therefore, we can applies Proposition 14. ■

3.C Proof of Proposition 13

We say that a communication game has "*no relevant tie*" if the following holds: for all pair of outcomes (θ, a'), (θ, a'') if a', $a'' \in A$ such that $a' \neq a''$ then $u_R(\theta, a') \neq u_R(\theta, a'')$. Moreover, wlog we can assume that $u_R(\theta', \cdot) \neq u_R(\theta'', \cdot)$ if $\theta' \neq \theta''$.[20]

Proof. Suppose that $\left(\theta, \left[\hat{\Theta}\right], \sigma_R\right) \in MPT(\mathcal{X})$ but either $(\theta, m) \notin \Sigma_S(\infty, \Delta)$ and $s_R \notin S_R(\infty, \Delta)$.

$\left(\theta, \left[\hat{\Theta}\right]\right) \notin \Sigma_S(\infty, \Delta)$ only if there exists k such that $\left(\theta, \left[\hat{\Theta}\right]\right)$ is not $(k, \Delta) - rationalizable$ i.e., there is not μ_S such that $\left[\hat{\Theta}\right] \in BR_S(\theta, \mu_S)$ and $\mu_S(S_R(k-1, \Delta)) = 1$.
If $\left(\theta, \left[\hat{\Theta}\right], \sigma_R\right) \in MPT(\mathcal{X})$ then $\theta \in \hat{\Theta}$ and for any $s_R \in S_R(k-1, \Delta)$, $s_R\left(\left[\hat{\Theta}\right]\right) = \arg\max_{a' \in A} \int_{\hat{\Theta}} u_R(a', \theta) p(\theta) d\theta$ for all messages $\left[\hat{\Theta}\right] \in M$. By condition (37), $s_R\left(\left[\hat{\Theta}\right]\right) = A^*\left(\left[\hat{\Theta}\right]\right) = \arg\max_{a \in A^*} u_S(a, \theta)$ for all the $\theta \in \hat{\Theta}$. Therefore, for any probability measure on $S_R(k-1, \Delta)$ $\left[\hat{\Theta}\right]$ is a best response for θ.

$s_R \notin S_R(\infty, \Delta)$ only if there exists k such that s_R is not $(k, \Delta) - rationalizable$ i.e., there is not $\mu_R \in \Delta'_R$ such that $\forall \left[\hat{\Theta}\right]$ $s_R\left(\left[\hat{\Theta}\right]\right) \in BR_R\left(\left[\hat{\Theta}\right], \mu_R\left(\cdot \mid \left[\hat{\Theta}\right]\right)\right)$ and $\Theta\left(\left[\hat{\Theta}\right], k-1, \Delta\right) \neq \emptyset$ implies $\mu_R\left(\Theta\left(\left[\hat{\Theta}\right], k-1, \Delta\right) \mid \left[\hat{\Theta}\right]\right) = 1$.

Given condition (37), $\hat{\Theta} \subseteq \Theta\left(\left[\hat{\Theta}\right], k-1, \Delta\right)$ if $\left[\hat{\Theta}\right] \in M(\mathcal{X})$ and any $\mu \in \Delta'_R$ $\mu\left(\hat{\Theta}\right) = 1 = \mu\left(\Theta\left(\left[\hat{\Theta}\right], k-1, \Delta\right)\right)$. Let us consider $\left[\hat{\Theta}'\right] \notin M(\mathcal{X})$ θ belongs to $\Theta\left(\left[\hat{\Theta}'\right], k-1, \Delta\right)$ for $k > 1$ only if the expected payoff is at least to $\max_{a' \in A} \int_{\hat{\Theta}} u_R(a', \theta) p(\theta) d\theta$ where $\hat{\Theta} \in \mathcal{X}$ and $\theta \in \hat{\Theta}$. This implies that $\max_{a' \in A} \int_{\hat{\Theta}} u_R(a', \theta) p(\theta) d\theta = \max_{a' \in A} \int_{\hat{\Theta}'} u_R(a', \theta) p(\theta) d\theta$ but this is impossible given the assumptions of "no relevant tie" and of strictly positive prior $p(\cdot)$. ■

We should remark that Condition 37 does not play any role in the proof

[20] If there exists θ' and θ'' such that $u_S(\theta', \cdot) = u_S(\theta'', \cdot)$ we can just merge them or simply drop one of the two.

Bibliography

[1] K. Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, 4:328–346, 1950.

[2] R. J. Aumann. Interactive epistemology i: Knowledge. *International Journal of Game Theory*, 28(3):263–300, 1999.

[3] P. Battigalli. Rationalizability in infinite, dynamic games with incomplete information. *Research in Economics*, 57(1):1–38, 2003. TY - JOUR.

[4] P. Battigalli. Rationalization in signaling games: Theory and applications. *mimeo*, 2003.

[5] P. Battigalli and M. Siniscalchi. Hierarchies of conditional beliefs and interactive epistemology in dynamic games. *Journal of Economic Theory*, 88(1):188–230, 1999.

[6] P. Battigalli and M. Siniscalchi. Strong belief and forward induction reasoning. *Journal of Economic Theory*, 106(2):356–391, 2002. TY - JOUR.

[7] P. Battigalli and M. Siniscalchi. Rationalization and incomplete information. *Advances in Theoretical Economics*, 3, 2003.

[8] J. P. Benoit and J. Dubra. Information revelation in auctions. *mimeo*, 2004.

[9] A. Brandenburger and E. Dekel. Rationalizability and correlated equilibria. *Econometrica*, 55:1391–1402, 1987.

[10] J. Broome. Utilitarianism and expected utility. *The journal of Philosophy*, 84:405–422, 1987.

[11] G. Cappelletti. On the equivalence between semantic and syntactic approach. 2003.

[12] G. G. L. Cappelletti. A note on the characterization of strong d-rationalizability in static and extensive form games. 2004.

[13] C. Chang and H. J. Keisler. *Model Theory*. North Holland, 1998.

[14] W. B. Charles Blackorby and D. Donaldson. HarsanyiŠs social aggregation theorem: A multi-profile approach with variable-population extensions. *mimeo*, 2003.

[15] V. Crawford. Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Economic Review*, 93(1):133–149, 2003.

[16] V. P. Crawford and J. Sobel. Strategic information transmission. *Econometrica: Journal of the Econometric Society*, 50(6):1431–1451, 1982.

[17] C. D'Aspremont and L. Gevers. Equity and the infromational basis of collective choice. *Review of Economic Studies*, 44:199–209, 1977.

[18] R. M. Dudley. *Real Analysis and Probability*. Cambridge Press, 2002.

[19] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about knowledge*. MIT Press, 1995.

[20] J. Farrell. Meaning and credibility in cheap-talk games. 5, 1993.

[21] C. Genest and J. V. Zidek. Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 1:114–135, 1986.

[22] P. R. Halmos. *Measure Theory*. Springer Verlag, 1974.

[23] P. J. Hammond. Equity in two person situations: Some consequences. *Econometrica*, 47:1127–1136, 1979.

[24] P. J. Hammond. Ex-ante and ex-post welfare optimality under uncertainty. *Economica*, 48:235–250, 1981.

[25] P. J. Hammond. Equity, arrow's conditions, and rawls' difference principle. *Econometrica*, 44:793–804, 76.

[26] J. Harsanyi. Cardinal welfare, individualistic ethics, and interpersonal comparison of utility. *Journal of Political Economics*, 63:309–321, 1955.

[27] A. Heifetz and D. Samet. Knowledge spaces with arbitrarily high rank. *Games and Economic Behavior*, 22(2):260–273, 1998. TY - JOUR.

[28] A. Heifetz and D. Samet. Topology-free typology of beliefs. *Journal of Economic Theory*, 82(2):324–341, 1998. TY - JOUR.

[29] A. Heifetz and D. Samet. Coherent beliefs are not always types. *Journal of Mathematical Economics*, 32(4):475–488, 1999. TY - JOUR.

[30] A. Heifetz and D. Samet. Hierarchies of knowledge: An unbounded stairway. *Mathematical Social Sciences*, 38(2):157–170, 1999. TY - JOUR.

[31] T.-W. Hu. General rationalizability and its robustness for strategic form games with incomplete information. 2004. Mimeo.

[32] A. Hylland and R. Zeckhauser. The impossibility of bayesian group decision making with separate aggregation of beliefs and values. *Econometrica*, 47:1321–1336, 1979.

[33] D. S. I. Gilboa and D. Schmeidler. Utilitarian aggregation of beliefs and tastes. *mimeo*, March 2001.

[34] M. e. A. Kaneko. A map of common knowledge logics. *Studia Logica*, 71:57–86, 2002.

[35] H. J. Keisler. *Model Theory for Infinitary Logic*. North Holland, 1991.

[36] M. C. Kemp and Y. Ng. On the existence of social welfare functions, social orderings and social decision functions. *Economica*, 43:59–66, 1976.

[37] V. Krishna and J. Morgan. The art of conversation: Eliciting information from experts through multi-stage communication. *mimeo*, 2003.

[38] D. Marker. *Model Theory: An Introduction*. Springer, 2002.

[39] J. Mertens and S. Zamir. Formulation of bayesian analysus for games with incomplete information. *International Journal of Game Theory*, 14:1–29, 1985.

[40] B. D. Meyer and P. Mongin. A note on affine aggregation. *Economics Letters*, 47:177–183, 1995.

[41] P. Mongin. Harsanyi's aggregation theorem: Multiple-profile version and unsettled questions. *Social Choice and Welfare*, 11:331–354, 1994.

[42] P. Mongin. Consistent bayesian aggregation. *Journal of Economic Theory*, 66:313–351, 1995.

[43] J. Nash. Non-cooperative game. *The Annals of Mathematics*, 54:286–295, 1951.

[44] M. Okuno-Fujiwara, A. Postlewaite, and K. Suzumura. Strategic information revelation. *The Review of Economic Studies*, 57(1):25–47, 1990. FLA 00346527 Society for Economic Analysis Limited Copyright 1990 The Review of Economic Studies Ltd.

[45] R. P. Parks. An impossiblility theorem for fixed preference: A dictatorial bergson-samuelson welfare function. *The Review of Economic Studies*, 43:447–4450, 1976.

[46] P. Pollak. Bergson - samuelson social welfare functions and the theory of social choice. *The Quarterly journal of Economics*, 93:73–90, 1979.

[47] M. Rabin. Communication between rational agents. *Journal of Economic Theory*, 51(1):144–170, 1990.

[48] H. Raiffa. *Decision Analysis*. Addion-Wesley, 1970.

[49] A. Rubinstein. The single profile analogues to multi profile theorems: Mathematical logic's approach. *International Economic Review*, 25:719–730, 1984.

[50] L. Savage. *The Foundation of Statistics*. Dover Publications, 1954.

[51] A. Sen. *Collective Choice and Social Welfare*. Holden Day, 1974.

[52] M. Shimoji and J. Watson. Conditional dominance, rationalizability, and game forms. *Journal of Economic Theory*, 83:161 195, 1998.

[53] J. B. K. T. Seidenfeld and M. Schervish. On the shared preferences of two bayesian decision makers. *The Journal of Philosophy*, 86:225–244, 1989.

[54] T. C. Tan and S. R. D. C. Werlang. The bayesian foundations of solution concepts of games. *Journal of Economic Theory*, (45), 1988.

Chapter 4

A Note on the Characterization of Δ–rationalizability

A Note on the Characterization of strong Δ–rationalizability in static and Extensive Form Games

Giuseppe G. L. Cappelletti[1]
Bocconi University

January 2^{nd}, 2005

Abstract

We characterize strong Δ – *rationalizability* using the notion of iterated dominance.
Keywords: Rationalizability, Dominance
JEL Classifications: C72, D82

[1] I wish to thank Pierpaolo Battigalli, Sandro Brusco and Marco Mariotti for very useful comments.

4.1 Introduction

The concept of rationalizability, which has introduced by Bernheim and Pearce, has been widely accepted in the study of normal form games. For dynamic games, Pearce defined the concept of extensive form rationalizability. It incorporates sequential rationality and implies form of backward and forward induction. There are a situation where is natural to introduce assumptions on players' belief. Consider a situation where players' interests are perfectly alligned, there it is natural assume that each agent is certain of his opponents' action. In this settings it is natural to consider the implications of rationality, common certainty of rationality and common certainty of some restrictions on players beliefs. This leads to an extension of the rationalizability solution concept to static and dynamic games of incomplete information, which takes as given some exogenous restrictions on players' beliefs. The solution concept is called $\Delta - \textit{rationalizability}$ (see for example Battigalli (2003) [3]). In this note we want to characterize the set of rationalizable actions in terms of dominance ralations.

4.2 Static Form Games

A static form games of perfect information is a structure:

$$G = \langle N, \{S_i, u_i\}_{i \in N} \rangle$$

where:

- for each $i \in N = \{1, 2, ..., n\}$ and S_i is a finite set of possible actions for player i. The opponent of player i is denoted as $-i$.
- for each $i \in N$, $u_i : \prod_{i \in N} S_i \to \mathbb{R}$ is the payoff function for player i ($\mathbb{R}$ denotes the set of real numbers).

Paramenter θ_i represents player i's private information about the rules of the game. Players' first order beliefs is a probability measure on strategies of their opponents, $\mu_i \in B_i^0 := \Delta(S_{-i})$, where B_i^0 is the set of his first order beliefs and S_{-i} is the first order uncertainty. The higher order beliefs are define recursively as $B_i^k := \Delta(X_i^k)$ where $X_i^k := X_i^{k-1} \times \prod_{j \in N/\{i\}} B_j^{k-1}$. A player's beliefs may be assumed to satisfy some restrictions which are not implied by assumptions concerning belief in rationality, or beliefs about such beliefs. Such restrictions may be justified or related to some structural properties of the game model. Let us consider the correspondence $\Delta_i : N \to \Delta(S_{-i})$ and $\Delta := (\Delta_i)_{i \in N}$. The set Δ_i is interpreted as the exogenous restriction on player i's first order beliefs and $\Delta = (\Delta_1, \Delta_2, ..., \Delta_n)$

A strategy is rational for a player i with first order beliefs μ_i if it maximizes the expected utility. Namely, given a belief μ_i and an action s_i let

$$U_i(s_i, \mu_i) := \sum_{s_{-i} \in S_{-i}} u_i(s_i, s_{-i}) \mu_i(s_{-i})$$

denote the expected payoff for player i from playing s_i.

Definition 39 *A strategy s_i is rational player i with respect to $\mu_i \in \Delta_i$, denoted as $s_i \in \rho_i(\mu_i)$ if for every $s_i' \in S_i$ the following inequality is well-defined and satisfied:*

$$U_i(s_i, \mu_i) \geq U_i(s_i', \mu_i)$$

or equivalently

$$s_i \in \arg\max_{s_i' \in S_i} U_i(s_i', \mu_i)$$

Therefore, the set of pure best response to μ_i for type θ_i is denoted as $\rho_i(\mu_i)$. Now, we can introduce the solution concept $\Delta-$*rationalizability*.[2] $\Delta-$*rationalizability* characterizes the set of actions realized when all the following conditions are true:

(0) every player i has first-order conditional beliefs in Δ_i and is sequentially rational;

(W1) every player i is certain of (0);

(WA2) every player i is certain of (W1);

....

(WAk+1) every player i is certain of (Wk);

....

Let $W_i(0, \Delta) = S_i$ for $i \in N$. Assume that subsets $W_i(k, \Delta)$, $i \in N$ have been defined for $k = 1, 2, ...$ Then for each $i \in N$ $W_i(k+1, \Delta)$ is the set

[2] In static games the distintion betweeen strong and weak $\Delta-$*rationalizability* is vacous.

of feasible s_i such that $s_i \in \rho_i(\mu_i)$ with respect to some $\mu_i \in \Delta_i$ such that $\mu_i(W_{-i}(k,\Delta)) = 1$. An action $s_i \in W_i(k,\Delta)$ is called $(k,\Delta) - rationalizable$. An action is $\Delta - rationalizable$ if it is $(k,\Delta) - rationalizable$ for all $k = 1, 2, \ldots$. The set of $\Delta - rationalizable$ actions for player i is denoted by $W_i(\infty, \Delta)$.

Battigalli and Siniscalchi (2002) [6] formally express these assumptions (by means of "complete extensive-form type spaces") and show that the set of actions consistent with assumptions (0)-(WAk) are those and only those which belong, respectively, to the subsets $\{S_i(k,\Delta)\}_{i\in N}$ defined by the following procedure that iteratively deletes actions for each player i:

- first, define $S_i(0,\Delta) = S_i$;
- for $k = 1, 2, \ldots$ let

$$S_i(k,\Delta) := \{s_R \in S_R(k-1,\Delta) : \exists \mu_i \in \Delta_i, s_i \in \rho_i(\mu_i) \text{ and } \mu_i(S_{-i}(k-1,\Delta)) = 1\}.$$

Definition 40 *Fix a pair of subsets of beliefs* $\Delta = \prod_{i\in N}(\Delta_i)$, *where* $\emptyset \neq \Delta_i \subseteq \Delta(S_i)$ *for every* $i \in N$. *The action* s_i *is* $(k,\Delta) - rationalizable$ *if and only if* $s_i \in S_i(k,\Delta)$.

Definition 41 *The action* s_i *is* $\Delta - rationalizable$ *if and only if* $s_i \in S_R(\infty,\Delta) := \bigcap_{k\geq 1} S_i(k,\Delta)$.

Furthermore, we focus on games without relevant ties. A game has no relevant tie if the following holds: for each player i and all pairs s' and s'' such that $s'_i \neq s''_i$ then $u_i(s') \neq u_i(s'')$. This means that if player i is certain of its opponents' actions then different actions imply different payoff. He cannot be indifferent between any two feasible actions.

The set of $\Delta - rationalizable$ actions should be further characterized for generic finite games in terms of dominance relations. A strategy $s_i \in S_i$ is strictly dominated by a mixed strategy $\sigma_i \in \Delta(S_i)$[3] for type player i on $B_{-i} \subseteq S_{-i}$ if

$$\forall \sigma_{-i} \in B_{-i}, U_i(s_i, \sigma_{-i}) < \sum_{s'_i \in S_i} \sigma_i(s'_i) u_i(s'_i, \sigma_{-i})$$

We let $UD_i \subseteq S_i$ denote the set of pure actions not dominated by any mixed action.

[3] We are assuming that the set of feasible actions does not depend on the type. It's straightforward the extension to the more general case.

For a given rectangular $B \subseteq S$ let $\mathcal{S}(B)$ denote the set of $(s_i)_{i\in N} \in S$ such that, for each i, s_i is not strictly dominated on B_{-i}.

As the following example shows, the concepts of Δ - *rationalizability* and strictly undominance do not coincide. Take the following static game with perfect information, there is a unique Nash equilibrium, characterized by an obvious restriction on beliefs while $\mathcal{S}(S)$ coincides with in S.

	A	B
a	2,2	2,1
b	1,2	3,3

Figure 1 :

Therefore, we need to extend this definition in order to take into account the exogenous restrictions on players' beliefs. We need some preliminary definitions.

Let us define $p(s_i, s_{-i}; \sigma_i)$ as $\inf_{\{\mu_i \in \Delta(S_{-i}) : U_i(s_i,\mu_i) \geq U_i(\sigma_i,\mu_i)\}} \mu_i(s_{-i})$ and $\bar{p}(s_i, s_{-i})$ as $\inf_{\{\mu_i \in \Delta(S_{-i}) : U_i(s_i,\mu_i) \geq U_i(\sigma_i,\mu_i) \text{ for every } \sigma_i \in \Delta(S_i)\}} \mu_i(s_{-i})$. $\bar{p}(s_i, s_{-i})$ is the smallest probability attributed to s_{-i} such that s_i is the best action for player i.

Definition 42 *A strategy $s_i \in S_i$ is $\Delta-$ not strictly dominated by any mixed action for player i if for every σ_i there exists $s_{-i}(\sigma_i) \in S_{-i}$ such that $u_i(s_i, s_{-i}(\sigma_i)) \geq u_i(\sigma_i, s_{-i}(\sigma_i))$ and there exists $\mu_i \in \Delta_i$ such that $\mu_i(s_{-i}(\sigma_i)) \geq p(s_i, s_{-i}(\sigma_i); \sigma_i)$.*[4]

We let $UD_i(\Delta) \subseteq S_i$ denote the set of pure actions $\Delta-$ *not strictly dominated* by any mixed action. For a given rectangular $B \subseteq S$ let $\mathcal{S}(B, \Delta)$ denote the set of $(s_i)_{i\in N} \in S$ such that, for each i, s_i is $\Delta-$ *not strictly dominated* on B_{-i}.

Lemma 11 *Let G be a finite static game an action profile s_i is $\Delta-$ not strictly dominated where Δ is a convex subset of $\Delta(S)$ if and only if there exists $\mu_i \in \Delta_i$ such that $s_i \in \rho_i(\mu_i)$.*

Proof. If $s_i \in r_i(\mu_i)$ for some μ_i belonging to Δ_i then s_i is $\Delta-$ *not strictly dominated.* To establish the converse, suppose there exists no $\mu_i \in \Delta_i$ such that $s_i \in \rho_i(\mu_i)$. Then there exists a function $b_i : \Delta_i \to S_i$ such that for

[4] Alternatively we can define when a strategy is Δ - *strictly dominated* by σ_i.

Definition 43 *σ_i Δ - strictly dominates s_i if and only if either for every s_{-i} $u_i(s_i, s_{-i}) < u_i(\sigma_i, s_{-i})$ or $u_i(s_i, s_{-i}) \geq u_i(\sigma_i, s_{-i})$ implies $\max_{\mu_i \in \Delta_i} \mu_i(s_{-i}) < p(s_i, s_{-i}(\sigma_i); \sigma_i)$.*

every μ_i $b_i(\mu_i)$ is such that $U_i(b_i(\mu_i),\mu_i) > U_i(s_i,\mu_i)$. Consider the zero-sum game $\bar{G} = \langle N, \{S_i, \bar{u}_i\}_{i\in N}\rangle$ where $\bar{u}_i(s_i', s_{-i}') := u_i(s_i, s_{-i}') - u_i(s_i', s_{-i}')$ and $\bar{u}_j(s_i', s_{-i}') = -u_j(s_i', s_{-i}')$. Let $(\sigma_i^*, \sigma_{-i}^*)$ be the Nash Equilibrium of the game $\bar{G}$. For any $\sigma_{-i} \in \Delta_i \subseteq \Delta(S_{-i})$:

$$\bar{U}_i(\sigma_i^*, \sigma_{-i}) \geq \bar{U}_i(\sigma_i^*, \sigma_{-i}^*) \geq \bar{U}_i(b(\sigma_{-i}^*), \sigma_{-i}^*) > \bar{U}_i(s_i, \sigma_{-i}^*) = 0.$$

Hence, for every $\sigma_{-i} \in \Delta_i$ $U_i(\sigma_i, \sigma_{-i}) > U_i(s_i, \sigma_{-i})$. This proves the claim since is an action is $\Delta-$ *not strictly dominated* then it is not stricly dominated. ■

As a matter of fact a stronger result holds. The convexity assumption is not needed in order to prove the characterization result, even if convexity is a tight condition for existence to hold (refer to Theorem 6).

Lemma 12 *Let G be a finite static game an action profile s_i is $\Delta-$ not strictly dominated if and only if there exists $\mu_i \in \Delta_i$ such that $s_i \in \rho_i(\mu_i)$.*

Proof. If $s_i \in r_i(\mu_i)$ for some μ_i belonging to Δ_i then s_i is $\Delta-$ *not strictly dominated.* Suppose that for every $\mu_i \in \Delta(S_{-i})$ $s_i \notin \rho_i(\mu_i)$ then first Pearce's lemma applies. Otherwise $s_i \in \rho_i(\mu_i)$ implies that $\mu_i \notin \Delta_i$. Let us define the set of player i's which are justifiable by some belief in Δ_i, formally denoted as $D_i^{\Delta_i} := \{s_i \in S_i : \exists \mu_i \in \Delta(S_{-i}) \text{ such that } s_i \in \rho_i(\mu_i)\}$. Consider any mixed strategy σ_i with support equal to $D_i^{\Delta_i}$. By Construction σ_i $\Delta-$ *dominates* s_i. ■

If the set of restrictions define a non empty convex and closed subset of $\Delta(S)$ there exists a a set of actions which is nonempty.

Proposition 16 *For every $k = 1, 2, ...$ the following equivalence holds: $S_i(k, \Delta) = S^k(B, \Delta)$.*

Proof. By iterative application of Lemma 11. ■

Theorem 6 *If Δ is a nonempty convex and closed subset of $\Delta(S)$ and for every $i \in N$ $\bigcap_{j\neq i} \pi_{S_i}(\mu_j)$ is non empty convex sub set of $\Delta(S_i)$ then set of $\Delta-$ rationalizable strategies is nonempty*

Proof. By application of Theorem 1 (Nash 1951) [43]. ■

4.3 Static Form Games with incomplete information

A static form games of incomplete information is a structure:

$$\Gamma = \langle N, \{\Theta_i, S_i, u_i\}_{i\in N}\rangle$$

where:

- for each $i \in N = \{1, 2, ..., n\}$, Θ_i is a finite set of possible payoff-types for player i, and S_i is a finite set of possible actions for player i. The opponent of player i is denoted as $-i$.
- for each $i \in N$, $u_i : \prod_{i\in N} \Theta_i \times \prod_{i\in N} S_i \to \mathbb{R}$ is the payoff function for player i ($\mathbb{R}$ denotes the set of real numbers).

Paramenter θ_i represents player i's private information about the rules of the game.

Players' beliefs can be represented as probability measures about types and strategies of their opponents ($\mu_i \in \prod_{i\in N} \Delta(\Theta_i \times S_i)$). The state of nature $\theta = (\theta_1, \theta_2, ..., \theta_n)$ specifies the unknown paramenters of the game and the players' interactive knowledge about them. Note that Γ does not specify players' beliefs about the state of Nature θ. In order to provide a general representation of players' beliefs, we explicitly introduce epistemic types; a "Harsanyi type" t_i for Player i is the modeled as a pair consisting as a payoff type θ_i and an epistemic type e_i.

A strategy is rational for a player of type θ_i with beliefs $\mu_i$if it maximizes the expected utility of θ_i. Given a belief μ_i and an action s_i let

$$U_i(\theta_i, s_i, \mu_i) = \sum_{\theta_{-i}\in\Theta_{-i}, s_{-i}\in S_{-i}} U(\theta_i, s_i, \theta_{-i}, s_{-i})\, \mu_i(\theta_{-i}, \sigma_{-i})$$

denote the expected payoff for type θ_i from playing s_i.

Definition 44 *A strategy s_i is rational for type θ_i with respect to $\mu_i \in \Delta_i$, denoted as $(\theta_i, s_i) \in \rho_i(\mu_i)$ or equivalently $s_i \in r_i(\theta_i, \mu_i)$ if for every $s_i' \in S_i$ the following inequality is well-defined and satisfied:*

$$U_i(\theta_i, s_i, \mu_i) \geq U_i(\theta_i, s_i', \mu_i)$$

or equivalently

$$s_i \in \arg\max_{\bar{s}_i \in S_i} U_i\left(\theta_i, \bar{s}_i, \mu_i\right)$$

Lemma 11 and 12 could be extended to this setting focusing each payoff type of each player. For existence result we can rely on previous results by Hu (2004) [31].

4.4 Discussion

In order to extend Lemma 11 for dynamic games we should proceed along the analysis by the Shimonji and Watson (1998) [52] and Siniscalchi and Battigalli (1999) [5]. Characterization should be stated in term of *conditional* $\Delta-$*dominance*. It still remains an issue to generalize Theorem 6.

Bibliography

[1] K. Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, 4:328–346, 1950.

[2] R. J. Aumann. Interactive epistemology i: Knowledge. *International Journal of Game Theory*, 28(3):263–300, 1999.

[3] P. Battigalli. Rationalizability in infinite, dynamic games with incomplete information. *Research in Economics*, 57(1):1–38, 2003. TY - JOUR.

[4] P. Battigalli. Rationalization in signaling games: Theory and applications. *mimeo*, 2003.

[5] P. Battigalli and M. Siniscalchi. Hierarchies of conditional beliefs and interactive epistemology in dynamic games. *Journal of Economic Theory*, 88(1):188–230, 1999.

[6] P. Battigalli and M. Siniscalchi. Strong belief and forward induction reasoning. *Journal of Economic Theory*, 106(2):356–391, 2002. TY - JOUR.

[7] P. Battigalli and M. Siniscalchi. Rationalization and incomplete information. *Advances in Theoretical Economics*, 3, 2003.

[8] J. P. Benoit and J. Dubra. Information revelation in auctions. *mimeo*, 2004.

[9] A. Brandenburger and E. Dekel. Rationalizability and correlated equilibria. *Econometrica*, 55:1391–1402, 1987.

[10] J. Broome. Utilitarianism and expected utility. *The journal of Philosophy*, 84:405–422, 1987.

[11] G. Cappelletti. On the equivalence between semantic and syntactic approach. 2003.

[12] G. G. L. Cappelletti. A note on the characterization of strong d-rationalizability in static and extensive form games. 2004.

[13] C. Chang and H. J. Keisler. *Model Theory*. North Holland, 1998.

[14] W. B. Charles Blackorby and D. Donaldson. HarsanyiŠs social aggregation theorem: A multi-profile approach with variable-population extensions. *mimeo*, 2003.

[15] V. Crawford. Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Economic Review*, 93(1):133–149, 2003.

[16] V. P. Crawford and J. Sobel. Strategic information transmission. *Econometrica: Journal of the Econometric Society*, 50(6):1431–1451, 1982.

[17] C. D'Aspremont and L. Gevers. Equity and the infromational basis of collective choice. *Review of Economic Studies*, 44:199–209, 1977.

[18] R. M. Dudley. *Real Analysis and Probability*. Cambridge Press, 2002.

[19] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about knowledge*. MIT Press, 1995.

[20] J. Farrell. Meaning and credibility in cheap-talk games. 5, 1993.

[21] C. Genest and J. V. Zidek. Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 1:114–135, 1986.

[22] P. R. Halmos. *Measure Theory*. Springer Verlag, 1974.

[23] P. J. Hammond. Equity in two person situations: Some consequences. *Econometrica*, 47:1127–1136, 1979.

[24] P. J. Hammond. Ex-ante and ex-post welfare optimality under uncertainty. *Economica*, 48:235–250, 1981.

[25] P. J. Hammond. Equity, arrow's conditions, and rawls' difference principle. *Econometrica*, 44:793–804, 76.

[26] J. Harsanyi. Cardinal welfare, individualistic ethics, and interpersonal comparison of utility. *Journal of Political Economics*, 63:309–321, 1955.

[27] A. Heifetz and D. Samet. Knowledge spaces with arbitrarily high rank. *Games and Economic Behavior*, 22(2):260–273, 1998. TY - JOUR.

[28] A. Heifetz and D. Samet. Topology-free typology of beliefs. *Journal of Economic Theory*, 82(2):324–341, 1998. TY - JOUR.

[29] A. Heifetz and D. Samet. Coherent beliefs are not always types. *Journal of Mathematical Economics*, 32(4):475–488, 1999. TY - JOUR.

[30] A. Heifetz and D. Samet. Hierarchies of knowledge: An unbounded stairway. *Mathematical Social Sciences*, 38(2):157–170, 1999. TY - JOUR.

[31] T.-W. Hu. General rationalizability and its robustness for strategic form games with incomplete information. 2004. Mimeo.

[32] A. Hylland and R. Zeckhauser. The impossibility of bayesian group decision making with separate aggregation of beliefs and values. *Econometrica*, 47:1321–1336, 1979.

[33] D. S. I. Gilboa and D. Schmeidler. Utilitarian aggregation of beliefs and tastes. *mimeo*, March 2001.

[34] M. e. A. Kaneko. A map of common knowledge logics. *Studia Logica*, 71:57–86, 2002.

[35] H. J. Keisler. *Model Theory for Infinitary Logic.* North Holland, 1991.

[36] M. C. Kemp and Y. Ng. On the existence of social welfare functions, social orderings and social decision functions. *Economica*, 43:59–66, 1976.

[37] V. Krishna and J. Morgan. The art of conversation: Eliciting information from experts through multi-stage communication. *mimeo*, 2003.

[38] D. Marker. *Model Theory: An Introduction.* Springer, 2002.

[39] J. Mertens and S. Zamir. Formulation of bayesian analysus for games with incomplete information. *International Journal of Game Theory*, 14:1–29, 1985.

[40] B. D. Meyer and P. Mongin. A note on affine aggregation. *Economics Letters*, 47:177–183, 1995.

[41] P. Mongin. Harsanyi's aggregation theorem: Multiple-profile version and unsettled questions. *Social Choice and Welfare*, 11:331–354, 1994.

[42] P. Mongin. Consistent bayesian aggregation. *Journal of Economic Theory*, 66:313–351, 1995.

[43] J. Nash. Non-cooperative game. *The Annals of Mathematics*, 54:286–295, 1951.

[44] M. Okuno-Fujiwara, A. Postlewaite, and K. Suzumura. Strategic information revelation. *The Review of Economic Studies*, 57(1):25–47, 1990. FLA 00346527 Society for Economic Analysis Limited Copyright 1990 The Review of Economic Studies Ltd.

[45] R. P. Parks. An impossiblility theorem for fixed preference: A dictatorial bergson-samuelson welfare function. *The Review of Economic Studies*, 43:447–4450, 1976.

[46] P. Pollak. Bergson - samuelson social welfare functions and the theory of social choice. *The Quarterly journal of Economics*, 93:73–90, 1979.

[47] M. Rabin. Communication between rational agents. *Journal of Economic Theory*, 51(1):144–170, 1990.

[48] H. Raiffa. *Decision Analysis.* Addion-Wesley, 1970.

[49] A. Rubinstein. The single profile analogues to multi profile theorems: Mathematical logic's approach. *International Economic Review*, 25:719–730, 1984.

[50] L. Savage. *The Foundation of Statistics.* Dover Publications, 1954.

[51] A. Sen. *Collective Choice and Social Welfare.* Holden Day, 1974.

[52] M. Shimoji and J. Watson. Conditional dominance, rationalizability, and game forms. *Journal of Economic Theory*, 83:161 195, 1998.

[53] J. B. K. T. Seidenfeld and M. Schervish. On the shared preferences of two bayesian decision makers. *The Journal of Philosophy*, 86:225–244, 1989.

[54] T. C. Tan and S. R. D. C. Werlang. The bayesian foundations of solution concepts of games. *Journal of Economic Theory*, (45), 1988.

Part III

Social Choice

Chapter 5

Il Problema dell'aggregazione delle preferenze in Economia

Il Problema dell'aggregazione delle preferenze in Economia

Giuseppe G. L. Cappelletti[1]
Bocconi University

January 2^{nd}, 2005

Abstract

Il problema della aggregazione delle preferenze individuali è stato al centro della ricerca scientifica in differenti campi, ad esempio in economia per la teoria delle scelte sociali (Social Choice Theory), in statistica per la teoria delle decisioni, in filosofia per quanto riguarda la teoria dell'utilitarismo. Con questa ricerca illustriamo sinteticamente i diversi contributi cercando di individuare gli elementi comuni e rilevanti dei diversi contributi

[1] I wish to thank Pierpaolo Battigalli, Sandro Brusco and Marco Mariotti for very useful comments.

Keywords: Aggregation, Social Welfare Function
JEL Classifications: D71, D81.

5.1 Introduzione

Il problema della aggregazione delle preferenze individuali è stato al centro della ricerca scientifica in differenti campi, ad esempio in economia per la teoria delle scelte sociali (Social Choice Theory), in statistica per la teoria delle decisioni, in filosofia per quanto riguarda la teoria dell'utilitarismo. Con questa ricerca illustriamo sinteticamente i diversi contributi cercando di individuare gli elementi comuni e rilevanti dei diversi contributi.

Iniziando dalla Social Choice (solo per una certa prossimità con la mia formazione) gli articoli di [1] e di [26] furono interpretati come una caratterizzazione della funzione di benessere sociale e dell'utilitarismo.

5.2 Social choice theory

5.2.1 Harsanyi (1955) e Arrow (1960). Single e Multi profile

Una relazione di preferenza è una relazione binaria, che indicheremo con $\succ$ sull'insieme delle scelte $\mathcal{X}$.[2] Indicheremo con $\succ_0$ la relazione di preferenza sociale e con $\succ_i$ la relazione di preferenza dell'individuo i. Elenchiamo sinteticamente alcune ipotesi fatte sulla relazione di preferenza sociale:

1. (Asimmetria della relazione di preferenza sociale) Se $X \succ_0 Y$ allora $Y \not\succ_0 X$.

2. (Transitività della relazione di preferenza sociale) Se $X \succ_0 Y$ e $Y \succ_0 Z$ allora $X \succ_0 Z$.

3. (Transitività (debole) della relazione di preferenza sociale) Se $X \sim_0 Y$ e $Y \sim_0 Z$ allora $X \sim_0 Z$.

4. Se per ogni coppia di alternative $X, Y \in \mathcal{X}$ una ed una sole delle seguenti relazione vale $X \succ_0 Y$ oppure $Y \succ_0 X$ oppure $X \sim_0 Y$ allora la relazione di preferenza sociale induce un *ordinamento completo* nello spazio delle scelte.

5. ("Positività" tra preferenza sociale e preferenza individuale) Se $\exists i \in N$ $X \succ_i Y$ e $X \succsim_j Y \; \forall j \neq i$ allora $X \succ_0 Y$.

[2] Le relazioni di preferenza debole ed indifferenza possono essere definite nel modo usuale.

6. $\{X\,\tilde{}_i X'$ e $Y\,\tilde{}_i Y'$ ma $X \succ_i Y$ e $X' \succ_i Y'\}$ e $\{X\,\tilde{}_j X'$ e $Y\,\tilde{}_j Y'$ ma $Y \succ_j X$ e $Y' \succ_j X'\}$ e $\{X\,\tilde{}_k Y$ e $X'\,\tilde{}_k Y' \forall k \neq i, j\}$ allora $\{X \succ_0 Y$ sse $X' \succ_0 Y'\}$ e $\{Y \succ_0 X$ implica $Y' \succ_0 X'\}$ e $\{Y\,\tilde{}_0 X$ implica $Y'\,\tilde{}_0 X'\}$.

Axiom 1 *La relazione di preferenza sociale è un ordinamento completo che soddisfa la condizione di continuità (Se $X \succsim_0 Z \succsim_0 Y$ allora esiste $\alpha \in [0,1]$ t.c $.\alpha X + (1-\alpha)\,\tilde{}_0 Z$) e unanimità ("individualistic value judgment", ref. Fleming) (Se $X\,\tilde{}_i Y\ \forall i \in N$ allora $X\,\tilde{}_0 Y$).*

Axiom 2 *Le relazioni di preferenza individuali sono ordinamenti completi, continui e soddisfano la condizione di equivalenza (Se $X\,\tilde{}_i Y$ allora $\forall \alpha \in [0,1]$ $\alpha X + (1-\alpha) Y\,\tilde{}_i X$)*

Dopo aver elencato la pletorea di assiomi attribuibili ad una relazione di preferenza possiamo enunciare due noti teoremi di rappresentazione delle preferenze.

Theorem 7 *Se vale l'Assioma (1) esiste una funzione di benessere sociale che rappresenta la relazione di preferenza sociale. Questa è unica a meno di trasformazioni lineari.*

Theorem 8 *Se vale l'Assioma (2) esiste una funzione (a' la von Neumann-Morgenstern) di utilità individuale che rappresenta la relazione di preferenza individuale $\succsim_i$ per ciascun individuo.*

Dati i precedenti teoremi possiamo, immediatamente, individuare una prima proprietà della funzione di benessere sociale.

Theorem 9 *La funzione di benessere sociale W è una funzione delle funzioni di utilità individuali $\{U_i\}_{i \in N}$ ed è omogenea di grado uno.*

Proof. $W := W(U_1, ..., U_n)$ e $W(k(U_1, ..., U_n)) = kW(U_1, ..., U_n)$

Se $k \in [0,1]$ consideriamo le due lotterie con esito certo O, per cui $U_i(O) = 0$ $\forall i \in N$, (quindi $W = 0$) e Q in cui $U_i = u_i$ $\forall i \in N$ (quindi $W = w$). Data una qualsiasi mistura P di O e Q allora $U_i = pu_i$ $\forall i \in N$. e $W = pw_i$. Se consideriamo $p = k$ otteniamo la tesi.

Se $k \in (-\infty, 0)$ basta considerare R t.c. 0 è equivalente ad una mistura $(p, 1-p)$ di R e Q; quindi R è tale che che $U_i = \left(1 - \frac{1}{p}\right) u_i$ $\forall i \in N$ e $W = kw$; prendendo $k = 1 - \frac{1}{p}$ si ha che:

$$
\begin{aligned}
W\left[\left(p\left(1-\frac{1}{p}\right)u_i(R)\right)_{i\in N}\right] &= pW\left[\left(\left(1-\frac{1}{p}\right)u_i(R)\right)_{i\in N}\right] \\
W\left[((1-p)\,u_i(Q))_{i\in N}\right] &= (1-p)\,W\left[(u_i(Q))_{i\in N}\right] \\
W\left[\left((1-p)\,u_i(Q)+p\left(1-\frac{1}{p}\right)u_i(R)\right)_{i\in N}\right] &= (1-p)\,W\left[u_i(Q)\right]_{i\in N}+pW\left[\left(1-\frac{1}{p}\right)u_i(R)\right]_{i\in N}=0
\end{aligned}
$$

da cui per additività della utilità attesa:

$$pW\left[\left(\left(1-\frac{1}{p}\right)u_i(R)\right)_{i\in N}\right]+(1-p)\,W\left[(u_i(Q))_{i\in N}\right]=0$$

da cui la tesi.

Se $k\in(1,\infty)$ dobbiamo considerare S t.c. Q è equivalente ad ottenere S con probabilità p e O con probabilità $(1-p)$; da cui si ricava $k=\frac{1}{p}$. ■

Theorem 10 *W è la media ponderata delle funzioni d'utilità individuali* $W=\sum_{i\in N}a_iu_i$.

Remark 4 *La funzione di benessere sociale dovrebbe avere connotati etici ed ogni individuo dovrebbe avere una funzione di utilità "individuale" e una funzione di utilità "sociale" od etica.*

Remark 5 *L'utilità di un individuo può essere valutata in base alle sue scelte e alle sue dichiarazioni tuttavia Harsanyi individuò due problematiche relative alle due misure dell'utilità:*

- *una che si basa sul "Principle of Unwarranted Differentiation" secondo il quale due individui che hanno preferenze identiche e che compiono le stesse scelte possono trarre utilità differenti dalle stesse situazioni;*
- *e una di natura psicologica riguardante la presenza di diverse determinanti (culturali, psicologiche, ambientali) del livello di soddisfazione che un individuo ricava da un azione.*

Remark 6 *Tuttavia nel caso in cui queste difficoltà potessero essere superate anche grazie ad una maggiore conoscenza dei meccanismi psicologici Harsanyi sostiene che le funzioni di preferenza sociali (di ciascun individuo) dovrebbero convergere alla somma non ponderata delle utilità individuali sulla base di una etica puramente individualistica.*

Arrow dimostrò che l'unica funzione di preferenza sociale, definita per ogni profilo di preferenze individuali, che soddisfi il *principio di Pareto*[3] e sia *indipendente dalle alternative irrilevanti*[4] è *dittatoriale*[5].

Mentre Arrow considera alternative con esito certo Harsanyi considera atti/conseguenze incerte ma contestualmente assume che gli individui concordino nell'attribuzione di una misura di probabilità sull'insieme degli stati del mondo; inoltre Harsanyi adotta un approccio riconducibile a Bergson e Samuelson[6], secondo il quale le preferenze sono prefissate (modello "single profile") mentre Arrow considera ogni profilo di preferenze individuali possibile (modello "multi profile") Una serie di articoli [36] [46] [45] dimostrarono che teoremi di impossibilità a' la Arrow erano derivabili anche in approcci "multi-profile", infatti basta ampliare a sufficienza lo spazio delle scelte[7].

5.2.2 Pollak (1979)

Prendiamo come esempio l'articolo di [46] che è particolarmente rappresentativo dell'importanza delle ipotesi che si fanno sulla relazione di preferenza sociale (SR), a seconda che si assuma che questa sia un ordinamento completo oppure una semplice relazione binaria riflessiva e completa.

Notazione

Indichiamo il profilo delle preferenze degli individui in una società $R = \succsim := \{\succsim_1, \ldots, \succsim_n\}$ e quindi $R(\Sigma)$ rappresenta le preferenze degli individui sulle alternative $x \in \Sigma \subseteq S$ dove S è l'insieme delle alternative; gli assiomi utilizzati sono:

Indipendenza e Neutralità (IN) Se per qualche coppia di alternative $\{x, y\}$ si ha che $x \succsim_0 y$ allora per tutte le alternative $\{z, w\}$ se $\succsim (z, w) = \succsim (x, y)$ allora $z \succsim_0 w$.

[3] La condizione di Pareto utilizzata è:
se $x \succ_i y$ per $\forall i$ allora $x \succ_0 y$.

[4] L'indipendenza dalle alternative irrilevanti richiede che la relazione di preferenza sociale tra x e y non muti se il profilo delle preferenze individuali sia cambiate lasciando però le preferenze relative a x e y invariate.

[5] Una relazione di preferenza sociale è dittatoriale se esiste un individuo j t.c. $x \succ_j y$ implica $x \succ_0 y$ per ogni x, y e ogni profilo di preferenze individuali.

[6] Esisterebbe un ulteriore differenza tra i due approcci mentre Samuelson parla di una funzione di preferenza (welfare) sociale Arrow parla di un ordinamento sociale.

[7] Per una dimostrazione delle analogie tra un approccio a' la Bergson Samuelson (o *Single Profile*) e uno a' la Harsanyi (o *Multiple Profile*) si faccia riferimento all'articolo di Rubinstein (84) [49].

Questo assioma richiede che la scelta tra due alternative (x e y) dipenda solo dalle preferenze su x, y e se il sotto-profilo di preferenze su x, y è uguale a quello delle preferenze su z, w allora la scelta sociale deve essere la stessa nei due casi. Questo assioma corrisponde all'assioma di indipendenza dalle alternative irrilevanti nella modellizzazione a'la Arrow (o multiple profile); come quest'ultimo può suscitare qualche perplessità basti pensare ad uno stesso comitato che compie scelte differenti in contesti differenti.

Questa ipotesi insieme alla richiesta che la relazione di preferenza sociale sia un ordinamento impica che se esiste un gruppo di individui che decide tra due alternative questo dovrà decidere tra qualsiasi coppia di alternative (v. dimostrazione Teorema 11)

Principio di Pareto (P) Se per ogni agente i $x \succ_i y$ allora $x \succ_0 y$.

Nel caso di un modello single profile l'ipotesi che la funzione di preferenza sociale sia definita sull'insieme di tutti i profili di preferenze individuali possibili non ha senso, visto che per costruzione si considera un profilo di preferenze dato; ma una condizione equivalente dato l'assioma (IN) è:

(U^3) Un profilo di preferenze individuali $\succsim$ soddisfa la condizione U^3 se per ogni possibile profilo di preferenze definito su tre alternative x, y, z (non necessariamente distinte) devono esistere allora tre alternative a, b, c t.c. $\succsim(a,b,c) = \overline{\succsim}(a,b,c)$.

In altre parole devono esistere almeno un insieme di alternative almeno pari a 3^{2n}, dove n è il numero degli individui.

Definition 45 *Un sottoinsieme di individui, V, appartenenti alla società si dice "quasi decisivo nella scelta di x contro y" se per* $\forall i \in V$ $x \succ_i y$, $\forall i \notin V := N \backslash V$ $y \succ_i x$ *e* $x \succ_0 y$

Lemma 13 *Date quattro alternative (x, y, w, z) e la relazione di preferenza sociale $\succsim_0$ completa, quasi-transitiva e che soddisfa gli assiomi (U^3), (P) e (In). Se $V \subsetneq N$ è "quasi decisivo per la scelta di x contro y" e $z \succ_i w$ per $\forall i \in V$ allora V è decisivo per la scelta di z contro w.*

Proof. Consideriamo 3 alternative (a, b, c) t.c. $\succsim(a,b) = \succsim(x,y)$ e $\succsim(a,c) = \succsim(z,w)$, $b \succ_i c$ per $\forall i \in V$ e $b \succ_i a$ e $b \succ_i c$ per $\forall \notin V$[8]. Dall'ipotesi (U^3)

[8] Le relazioni di preferenze tra b e c non sono vincolate dall'enunciato del Lemma e per il principio di indipendenza e neutralità dovrebbero essere irrilevanti, perciò possiamo considerare questo caso particolare.

sappiamo che esiste un insieme di alternative (a, b, c) che soddisfa le precedenti ipotesi; quindi possiamo dedurre che $a \succ_0 b$ per l'ipotesi (IN) e $b \succ_0 c$ per l'ipotesi (P) quindi per la condizione di quasi transitività della relazione di preferenza sociale possiamo concludere che $a \succ_0 c$, ma per ipotesi $\succsim (a, c) = \succsim (z, w)$ quindi ancora per l'ipotesi (IN) possiamo concludere che $z \succ_0 w$. ■

Theorem 11 *Sotto le ipotesi (U^3) l'unica relazione d'ordine sociale che soddisfa (P) e (IN) è dittatoriale.*

Proof. Consideriamo il più piccolo sottoinsieme di individui quasi decisivo nella scelta tra a e b (che questo esista deriva dall'ipotesi U^3), se V è costituito da un solo individuo il lemma precedente implica che questo è un dittatore, cioè $x \succ_i y$ implica $x \succ_0 y$.

Supponiamo che $|V| \geqslant 2$ allora consideriamo $j \in V$ e consideriamo la seguente partizione della società $V \backslash \{j\}, \{j\}, V^c := N \backslash V$; grazie all'ipotesi (U^3) sappiamo che esistono tre alternative (a, b, c) t.c.:

1. $a \succ_j b \succ_j c$
2. $c \succ_k a \succ_k b$ per $\forall k \in V \backslash \{j\}$
3. $b \succ_k c \succ_k a$ per $\forall k \in V^c$

Da questo profilo di preferenze individuali possiamo dedurre che $a \succ_0 b$ (per costruzione V è quasi-decisivo per a, b), $b \succsim_0 c$ (altrimenti $V \backslash \{j\}$ sarebbe quasi-decisivo per c, b contraddicendo il fatto V è l'insieme quasi-decisivo più piccolo) data l'ipotesi di transitività della relazione d'ordinamento sociale $a \succ_0 c$; ma allora j è quasi-decisivo per a, c. ■

L'ipotesi (U^3) non è una assunzione sulla relazione di preferenza sociale quanto un ipotesi sulla struttura dell'insieme di scelta e può essere soddisfatta o meno a seconda della situazione.

Per comprendere appieno il legame tra l'ipotesi di non restrizione delle preferenze degli individui nella modelizzazione "multi profile" e l'ipotesi (U^3) nel modello "single profile" si faccia riferimento all'articolo di Rubinstein (1985)[49].

L'approccio "single profile" sembra più coerente con le osservazioni di Harsanyi di una componente etica nell'aggregazione delle preferenze d'altra parte proprio in questo approccio l'ipotesi di indipendenza e neutralità sembra irragionevole tanto più per il fatto che lo spazio delle alternative per soddisfare l'ipotesi (U^3) deve essere molto ampio.

A questo punto però dobbiamo ricordare l'osservazione fatta da Sen (1986): il Teorema di Harsanyi è costruito e dimostrato in un modello "single profile" cioè che prende come dato profilo di preferenze mentre tutta la teoria della "Social Choice" a partire da Arrow ha tra gli assiomi costitutivi l'universalità dell'insieme delle preferenze individuali. Molti articoli tra cui quelli di [41] e quello di [49] colmano questa differenza[9] in modo da poter far riferimento al risultato di Harsanyi in un modello multi profile e poter dare anche una base assiomatica all'utilitarismo.

5.3 Preferenze individuali a'la Savage

5.3.1 Meyer and Mongin (1995)[40] and Mongin (1994)[41]

Presa una funzione vettoriale definita su un generico spazio X: $F := (f_0, ..., f_n) : X \to \mathbb{R}^{n+1}$ si può parafrasare le condizioni di unanimità o di Pareto come vincoli sul codominio di questa funzione:

(P_0) Se $f_i(x) = f_i(y)$, $i = 1, ..., n$ allora $f_0(x) = f_0(y)$

(P_1) Se $f_i(x) \geq f_i(y)$, $i = 1, ..., n$ allora $f_0(x) \geq f_0(y)$

($P2$) Se $f_i(x) > f_i(y)$, $i = 1, ..., n$ allora $f_0(x) > f_0(y)$

(P_3) Se $f_i(x) \geq f_i(y)$ per $i = 1, ..., n$ ed esiste un j t.c. $f_j(x) > f_j(y)$ allora $f_0(x) > f_0(y)$.

(P_4) Se $f_i(x) \geq f_i(y)$, $i = 1, ..., n$ e $f_0(x) \leq f_0(y)$ allora $f_i(x) = f_i(y)$ $i = 0, 1, ..., n$.

Inoltre si può introdurre la seguente condizione di minimo accordo:

(C) $\exists x^*, y^* \in X | \forall i = 1, ..., n$ $f_i(x^*) > f_i(y^*)$

Lemma 14 *La condizione (P_4) implica (P_3) che a sua volta implica (P_2).*
La condizione (P_4) implica (P_1) che a sua volta implica (P_0).
La condizione (P_4) è equivalente alla congiuzione delle condizioni (P_0) e (P_3).
Se $F(X)$ è un insieme convesso allora (C) congiuntamente a (P_3) implica (P_4).

[9] Per comprendere le difficoltà insite in un approccio "single profile" si può considerare la condizione di simmetria che pur essendo definibile facilmente in modello "multi profile" non lo è in modello "single profile", oltre al fatto che nel primo caso la funzione di preferenza sociale è derivata mentre nel secondo caso viene assunta e poi caratterizzata.

Proof. Per le prime tre serie di implicazioni basta far riferimento alla prossima sezione mentre per l'ultima:

Supponiamo che valgano le condizioni (C) e (P_3) e $f_i(x) \geq f_i(y)$, $i = 1, ..., n$ e $f_0(x) \leq f_0(y)$. Valendo la condizione (P_3), e sapendo che esiste almeno una coppia di alternative, tra cui gli agenti non sono indifferenti sappiamo, se almeno una delle disuguaglianze vale in senso stretto allora $f_0(x) > f_0(y)$ quindi $f_i(x) = f_i(y)\ \forall i = 1, ..., n$.

Supponiamo ora che $f_0(x) < f_0(y)$ allora possiamo considerare il vettore u:

$$\begin{aligned} u &= (1-\varepsilon)[F(x) - F(y)] + \varepsilon[F(x^*) - F(y^*)] \\ \text{con } \varepsilon &\in (0,1) \end{aligned}$$

Allora esiste un ε t.c. $u_i > 0\ \forall i = 1, ..., n$ e $u_0 < 0$ ma allora data la convessità delll'insieme $F(X) - F(X) = \{F(x) - F(y) \,|\, x, y \in X\}$ esistono ξ_1, ξ_2 t.c. $u = F(\xi_1) - F(\xi_2)$ contraddicendo quindi la condizione (P_3). ■

Remark 7 *Questa lemma permette, sotto l'ipotesi che la funzione vettoriale costituita dalle funzioni di utilità individuali e quella sociale abbia un'immagine convessa*[10]*, di legare le differenti condizioni di Pareto alla forma funziona della funzione di utilità sociale, in particolare impongono che questa sia una funzione affine delle funzioni di utilità individuali.*

Proposition 17 *Se $K := F(X)$ è convesso allora la condizione (P_0) vale sse esistono $\lambda_1, ..., \lambda_n, \mu \in \mathbb{R}$ t.c.:*

$$f_0(x) = \sum_{i=1}^{n} \lambda_i f_i(x) + \mu \tag{5.1}$$

La condizione (P_1) vale sse la precedente uguaglianza vale e $\lambda_i \geq 0\ \forall i = 1, ..., n$.

Proof. (Si considerano le differenze tra le utilità associate a due alternative e si danno delle caratterizzazioni per (P_0) e (P_1) in termini di insiemi in $\mathbb{R}^{n+1}$ e poi si applica il classico teorema di separazione per insiemi convessi).

Si definiscono due insiemi:

[10] Questa ipotesi sembra essere implicata dalla condizione di non atomicità utilizzata da Mongin nell'articolo del 1983.

$$\begin{aligned} R_0 &: = \left\{z \in \mathbb{R}^{n+1} : z_0 < 0 \text{ e } z_i = 0 \text{ per } i = 1, ..., n\right\} \\ R_1 &: = \left\{z \in \mathbb{R}^{n+1} : z_0 < 0 \text{ e } z_i \geq 0 \text{ per } i = 1, ..., n\right\} \end{aligned}$$

Si può osservare[11] che definiti $K := F(X)$ e $K^- := K - K := \left\{z \in \mathbb{R}^{n+1} : \exists x, y \in K \text{ and } z = x - y\right\}$:

$$\begin{aligned} (P_0) &\iff K^- \cap R_0 = \varnothing \\ (P_1) &\iff K^- \cap R_1 = \varnothing \end{aligned}$$

Dato che, per ipotesi, K è convesso K^- è convesso e simmetrico rispetto a 0; se definiamo ancora $V(K^-)$ lo spazio vettoriale generato da K^-, cioè $V(K^-) := \left\{z \in \mathbb{R}^{n+1} : \exists \{\alpha_1, ...\alpha_m\} \subseteq \mathbb{R} \text{ e } \{x_1, ..., xm\} \subseteq K^- \text{ t.c. } z = \sum_{i=1}^{m} \alpha_i x_i\right\}$.

Si può ora scrivere che[12]:

$$(P_s) \iff V(K^-) \cap R_s = \varnothing \text{ per } s = 0, 1$$

ma dato che $\overline{R}_s - e_0 \subset R_s$[13] dove $e_0 = (1, 0, ..., 0)$ allora:

$$(P_s) \Rightarrow V(K^-) \cap \left\{\overline{R}_s - e_0\right\} = \varnothing \text{ per } s = 0, 1$$

Quindi abbiamo come implicazione delle condizioni di Pareto delle condizioni di separazione tra due insiemi convessi e chiusi (o più specificatamente dei poliedri) applicando un teorema standard di separazione, quindi esiste un $v = (v_0, v_1, ..., v_n)$ t.c.:

---

[11] Gli insiemi R_0 e R_1 sono la negazione dell'enunciato delle condizioni (P_0) e (P_1).

[12] Un implicazione $((P_s) \iff V(K^-) \cap R_s = \varnothing$ implica $(P_s) \iff K^- \cap R_s = \varnothing)$ è banale visto che $K^- \subseteq V(K^-)$ ma supponiamo per assurdo $((P_s) \iff K^- \cap R_s = \varnothing$ implica $(P_s) \iff V(K^-) \cap R_s = \varnothing$ per assurdo assumiamo che $(P_s) \iff K^- \cap R_s = \varnothing$ e $(P_s) \wedge V(K^-) \cap R_s \neq \varnothing)$ che valga la condizione (P_0) e che esista un $z \in V(K^-) \cap R_s$ quindi $z = \sum_{i=1}^{m} \alpha_i x_i$ con $\{x_1, ..., x_m\} \subseteq K^-$ ma data la simmetria di K^- rispetto all'origine si possono scegliere una combinazione di coefficienti e vettori tali che $\alpha_i \geq 0$ per $i = 1, ..., m$ ma allora data la convessità di K^-, $z\prime := \frac{1}{\left(\sum_{i=1}^{m} \alpha_i\right)} \sum_{i=1}^{m} \alpha_i x_i \in K^-$ ma allora $K^- \cap R_0 \neq \varnothing$.

[13] Ad esempio nel caso di R_0 questo insieme non è altro che la semiretta di un asse privata dell'origine la sua chiusura non è altro che la retta con l'origine perciò se a R_0 aggiungo l'origine mua la traslo verso $-\infty$ ottengo un sotto insieme di R_0; una untuizione simile vale per R_1.

$$\langle v, z - e_0 \rangle > \langle v, k \rangle \text{ per } \forall z \in \overline{R}_s, k \in V\left(K^-\right) \tag{5.2}$$

poichè $V(K^-)$ è uno spazio vettoriale allora:

$$\langle v, k \rangle = 0 \ \forall k \in V\left(K^-\right) \tag{5.3}$$

Quindi considerando $z = 0$ possiamo concludere che $v_0 < 0$, e riscrivere la precedente uguaglianza come:

$$\sum_{i=0}^{n} v_i k_i = 0$$
$$k_0 = \frac{1}{v_0} \sum_{i=1}^{n} v_i k_i$$
$$f_0(x) - f_0(y) = \frac{1}{v_0} \sum_{i=1}^{n} v_i \left[f_i(x) - f_i(y)\right]$$

Perciò una volta fissato y abbiamo determinato i parametri dell'espressione $f_0(x) = \sum_{i=1}^{n} \lambda_i f_i(x) + \mu$.

Se in particolare vale la condizione (P_1) si può notare che $\alpha e_i \in \overline{R}_1$ per $\forall i = 1, ..., n$ e $\alpha > 0$, quindi, utilizzando la disuguaglianza (5.2):

$$\langle v, \alpha e_i \rangle > \langle v, e_0 \rangle$$

quindi $v_i > \frac{1}{\alpha} v_0$ dove sappiamo che $v_0 < 0$.

Per la seconda parte dell'enunciato si procede in maniera analoga salvo il fatto che il teorema di separazione vale in senso forte. ■

Proposition 18 *Se si assume che $K := F(X)$ sia convesso e che valga la condizione (P_2) allora esistono i coefficienti $\lambda_1, ..., \lambda_n, \mu$ t.c. $\lambda_i \geq 0 \ \forall i = 1, ..., n$ ed $\exists j \in \{1, 2, ..., n\}$ t.c. $\lambda_j > 0$ ($\lambda_i > 0 \ \forall i = 1, ..., n$) e un numero $k \geq 0$ t.c.:*

$$\forall x \in X, \ k f_0(x) = \sum_{i=1}^{n} \lambda_i f_i(x) + \mu$$

Se valgono le condizioni (P_2) e (C) [(P_3) e (C)] allora esistono dei coefficienti $\lambda_1, ..., \lambda_n, \mu$ t.c. $\lambda_i \geq 0 \ \forall i = 1, ..., n$ ed $\exists j \in \{1, 2, ..., n\}$ t.c. $\lambda_j > 0$

$$[\lambda_i > 0 \forall i = 1, ..., n] \; t.c. \; f_0(x) = \sum_{i=1}^{n} \lambda_i f_i(x) + \mu.$$

Proof. La dimostrazione è simile alla precedente dove si considera l'insieme:

$$R_2 = \{z \in \mathbb{R}^{n+1} : z_0 \leq 0 \text{ e } z_i > 0, \; i = 1, ..., n\}$$

e si osserva che gli insiemi convessi e chiusi $R_2' := \overline{R}_2 + \sum_{i=1}^{n} e_i$ e $V(K^-)$ sono disgiunti e quindi separabili e similmente si può fare per la condizione (P_3) dato che la seconda parte della proposizione deriva dal fatto che $(C) \wedge (P_3) \Rightarrow (P_4) \Rightarrow (P_1) \Rightarrow (P_0)$. ■

Remark 8 *Queste proposizioni implicano direttamente il Teorema di Harsanyi per un modello multiple-profile.*

Si deve sottolineare l'importanza della condizione (C) che può essere interpretata sia come una condizione di non-"trivialità" sia come una condizione dell'esistenza di un minimo accordo.

E' da sottolineare l'interesse per gli studiosi di social science per il teorema di aggregazione di Harsanyi come possibile fondazione assiomatica dell'utilitarismo, ma come ha osservato Sen (1974) ed in seguito Mongin ([41]) in un modello single profile non è assolutamente garantito che i pesi assegnati alle diverse utilità degli individui siano indipendenti dal profilo delle preferenze individuali; quindi diversi autori ([51], [25], [23], [24],

[17]) hanno dimostrato in un contesto "multiple profile" il teorema di aggregazione.

In particolare d'Aspremont-Gevers hanno caratterizzato l'utilitarismo a partire dall'ipotesi che le preferenze individuali siano cardinali e parzialmente comparabili (la condizione IOU); questo fa nascere la domanda di come le ipotesi relative alla relazione di preferenza sociale implichino o almeno (più generalmente) sottintendano la comparabilità delle preferenze individuali. In questa ottica la Proposizione 1 ([41]) chiaramente lega l'ipotesi che la relazione di preferenza sociale (o più precisamente il funzionale di welfare sociale) soddisfi le condizioni di von Neumann-Morgenstern di continuità e indipendenza con il fatto che le preferenze individuali siano cardinali comparabili. Proprio questo tipo di riflessione ha portato Mongin ad investigare altre ipotesi relative alla rappresentazioni delle preferenze tra cui l'impostazione soggettivista a'la Savage.

5.3.2 Consistent Bayesian Aggregation - P. Mongin [42]

Mongin distingue due casi: nel primo gli individui sono chiamati ad esprimere direttamente le proprie credenze mentre nel secondo caso invece gli individui esprimono delle preferenze che si suppone rispettino le ipotesi fatte da [50].

Individui esprimono direttamente le proprie probabilità soggettive

Consideriamo uno spazio misurabile $(\Omega, \mathcal{A})$ dove $\mathcal{A}$ è una $\sigma-$ algebra e $n+1$ misure di probabilità $(P_0, P_1, ..., P_n)$ su questo spazio misurabile.

Adesso definiamo differenti condizioni di unanimità (Pareto)

Definition 46 *(C) $(=)$ $\forall A, B \in \mathcal{A}$ $P_i(A) = P_i(B)$ for $\forall i = 1, .., n$ then $P_0(A) = P_0(B)$.*

Definition 47 *(C_1) $(\geqslant)$ $\forall A, B \in \mathcal{A}$ $P_i(A) \geqslant P_i(B)$ for $\forall i = 1, .., n$ then $P_0(A) \geqslant P_0(B)$.*

Definition 48 *(C_2) $(>)$ $\forall A, B \in \mathcal{A}$ $P_i(A) > P_i(B)$ for $\forall i = 1, .., n$ then $P_0(A) > P_0(B)$.*

Definition 49 *(C_3) $(\geqslant + >)$ $\forall A, B \in \mathcal{A}$ $P_i(A) \geqslant P_i(B)$ for $\forall i = 1, .., n$ and $\exists k$ s.t. $P_i(A) > P_i(B)$ then $P_0(A) > P_0(B)$.*

Definition 50 *$(C^+) := (C) \,\&\, (C_3)$*

Remark 9 *$(C_1) \Longrightarrow (C)$ poichè l'uguaglianza è un caso particolare di $\geqslant$; $(C_3) \Longrightarrow (C_2)$ poichè se tutte le disuguaglianze valgono "in modo stretto" la validità di (C_3) implica la conclusione; $(C^+) \Longrightarrow (C_1)$*

Se vogliamo determinare P_0 in modo da rispettare la condizione (C^+) basta prendere $P_0(A) = \sum_{i=1}^{n} \alpha_i P_i(A)$ per ogni $A \in \mathcal{A}$ t.c. $\sum_{i=1}^{n} \alpha_i = 1$.

Definition 51 *Dato uno spazio di probabilità (Ω, A, P) si dice che P è nonatomica se $\forall A \in \mathcal{A}$, $P(A) > 0$ allora $\exists B \in \mathcal{A}$, $B \subseteq A$ e $P(B) \leqslant P(A)$.*

Theorem 12 *(Liapunov) Il codominio di una misura non atomica finita su un insieme convesso è convessa.*

Lemma 15 *Un misura (oppure più in generale una misura "vector valued") finita $P = (P_1, ..., P_n)$ su uno spazio misurabile $(\Omega, \mathcal{A})$ è nonatomica sse $\forall A \in \mathcal{A}$ $\forall \alpha \in [0, 1]$, $\exists B \in \mathcal{A}$ t.c. $B \subseteq A$ e $P(B) = \alpha P(A)$.*

Proof. La dimostrazione si basa sull'osservazione che la misura condizionata $P_i(\cdot|A)$ $\forall i \in \{1,...,n\}$ è nonatomica ed applicare il Teorema di Liapunov.

Consideriamo $P_i'(\cdot) := \begin{cases} 0 \text{ if } P_i(A) = 0 \\ P_i(\cdot|A) \text{ if } P_i(A) \neq 0 \end{cases}$ then $P' := (P_1, P_2, ..., P_n)$ è una misura nonatomica e quindi grazie al teorema di Liapunov sappiamo che ha "range" convesso cioè $[0,1]^n$ quindi $\forall \alpha \in [0,1]$ esiste $B \in \mathcal{A}$ t.c. $P'(B) = \alpha$ e quindi $\frac{P(B \cap A)}{P(A)} = \alpha$ da cui la tesi. ■

Lemma 16 *Consideriamo uno spazio di probabilità nonatomico $(\Omega, \mathcal{A}, P_i)$ con $i \in I := \{1, 2, ..., n\}$ allora $(C_2) \Longrightarrow (C_1)$. Quindi $(C^+) \Longleftrightarrow (C_3)$.*

Proof. Assumiamo che valga (C_2) e consideriamo due insiemi $A, B \in \mathcal{A}$ t.c. $A \cap B = \varnothing$[14] and $P_i(A) \geqslant P_i(B)$ per $i = 1, ..., n$. Partizioniamo I in I', I'', I''' definiti come:

- $I' := \{i \in I | P_i(A) = P_i(B) = 0\}$
- $I'' := \{i \in I | P_i(A) = P_i(B) > 0\}$
- $I''' := \{i \in I | P_i(A) > P_i(B)\}$

Consideriamo $C := (A \cup B)^c$ e dal Lemma precedente e per ciascun $k \in \{1, ..., n\}$ possiamo trovare $C_k \subseteq C$ e $B_k \subseteq B$ t.c.:

$$\frac{1}{p}\begin{pmatrix} P_0(C) \\ P_1(C) \\ ... \\ P_n(C) \end{pmatrix} = \begin{pmatrix} P_0(C_p) \\ P_1(C_p) \\ ... \\ P_n(C_p) \end{pmatrix} \text{ e } \frac{1}{p}\begin{pmatrix} P_0(B) \\ P_1(B) \\ ... \\ P_n(B) \end{pmatrix} = \begin{pmatrix} P_0(B_p) \\ P_1(B_p) \\ ... \\ P_n(B_p) \end{pmatrix}$$

Si noti che $P_i(A \cup C_p) > P_i(B \backslash B_p)$ infatti $C_p \cap A = \varnothing$ quindi $P_i(A \cup C_p) = P_i(A) + P_i(C_p) = P_i(A) + \frac{1}{p} P_i(C)$ ma $P_i(B \backslash B_p) = P_i(B) - P_i(B_p) = \left(1 - \frac{1}{p}\right) P_i(B) < P_i(A) \leqslant P_i(A) + \frac{1}{p} P_i(C)$.

Se $P_i(A) = P_i(B) = 0$ allora $P_i(C) = 1$, quindi la disuguaglianza vale ancora.

Se P_0 soddisfa la condizione $(C_2)(>)$ allora $P_0(A \cup C_p) > P_0(B \backslash B_p)$ e quindi $P_0(A) + \frac{1}{p} P_0(C) > \left(1 - \frac{1}{p}\right) P_0(B)$. Facendo tendere $p \to \infty$ ricaviamo che $P_0(A) \geqslant P_0(B)$. ■

[14] Dati due qualsiasi insiemi A, B possiamo sepre considerare gli insiemi disgiunti A e $B \cap A^c$.

Quindi $(C^+) \Longrightarrow (C_3)$ ma anche $\underbrace{(C_3) \Longrightarrow (C_1)}_{\text{Sotto l'ipotesi di nonatomicità}}$, $\overbrace{(C_1) \Longrightarrow (C)}^{\text{Remark Precedente}} \Longrightarrow$ (C^+).

Esistenza di P_0

Proposition 19 *Consideriamo $(\Omega, \mathcal{A}, P_i)_{i=\{1,...,n\}}$ spazi di probabilità nonatomici e P_0 è una misura di probabilità su $(\Omega, \mathcal{A})$ che soddisfa la condizione (C). Allora esistono $\{\alpha_i\}_{i=\{1,...,n\}} \in \mathbb{R}$ e t.c. $\sum_{i=1}^{n} \alpha_i = 1$ e $P_0(\cdot) = \sum_{i=1}^{n} \alpha_i P_i(\cdot)$.*

Remark 10 *Come osservato in [21] il "linear pooling" non ripspetta la proprietà "independence preservation".*

Proof. (P_0 è nonatomica)

Se $P_0(A) > 0$ grazie al Lemma 1 sappiamo che $\exists A_1(i) \in \mathcal{A}$ t.c. $A_1(i) \subseteq A$ e $P_i(A_1(i)) = \frac{1}{2}P(A)$ quindi A_1 e $A_2 := A \cap A_1^c$ sono tali che $P_i(A_1) = P_i(A_2) = \frac{1}{2}P_i(A)$ questo vale per ogni $i \in \{1, ..., n\}$ ma allora valendo la condizione (C) $P_0(A_1) = \frac{1}{2}P_0(A)$.

Se consideriamo l'insieme M delle funzioni semplici $\Omega \to \mathbb{R}$ consideriamo il funzionale lineare:

$$\forall f \in M \ \varphi_i(f) = \int_\Omega f(\omega)\, dP_i(\omega)$$

(Se $\varphi_i(f) = \varphi_i(g)$ $i \in \{1, ..., n\}$ allora $\varphi_0(f) = \varphi_0(g)$)

$$\text{Se } \varphi_i(f) = \varphi_i(g)\, i \in \{1, ..., n\} \text{ allora } \varphi_0(f) = \varphi_0(g) \qquad ((D))$$

Se $f(\omega) = I_A(\omega)$ e $g = I_B(\omega)$ la condizione (C) ci implica la conclusione, infatti:

$\int_\Omega f(\omega)\, dP_i(\omega)$ si riduce a $\int_A dP_i(\omega) = \int_B dP_i(\omega) \Longleftrightarrow P_i(A) = P_i(B)$

Se $f(\omega) = \sum_{i=1}^{L} \lambda_i I_{L_i}(\omega)$ e $g(\omega) = \sum_{k=1}^{M} \gamma_k I_{G_k}(\omega)$ e supponiamo $\sum_{i=1}^{L} \lambda_i > \sum_{k=1}^{M} \gamma_k > 0$

La misura $\left(\sum_{i=1}^{L}\lambda_i\right)^{-1}(\varphi_0(f),\varphi_1(f),...,\varphi_n(f))$ è una combinazione convessa di misure di probabilità quindi dal Teorema di Liapunov sappiamo che esiste un insieme $A \in \mathcal{A}$ t.c.:

$$\left(\sum_{i=1}^{L}\lambda_i\right)^{-1}(\varphi_0(f),\varphi_1(f),...,\varphi_n(f)) = (P_0(A),P_1(A),...,P_n(A))$$

Lo stesso vale per il vettore:

$$\left(\sum_{i=1}^{L}\gamma_i\right)^{-1}(\varphi_0(g),\varphi_1(g),...,\varphi_n(g)) = (P_0(B),P_1(B),...,P_n(B))$$

Grazie al Lemma 1 sappiamo che esiste $B' \subseteq B$ t.c.:

$$\begin{aligned}\left(\sum_{i=1}^{L}\lambda_i\right)^{-1}(\varphi_0(g),\varphi_1(g),...,\varphi_n(g)) &= \left(\sum_{i=1}^{L}\lambda_i\right)\left(\sum_{i=1}^{L}\lambda_i\right)^{-1}\left(\sum_{i=1}^{L}\lambda_i\right)^{-1}(\varphi_0(g),\varphi_1(g),...,\varphi_n(g)) = \\ &= (P_0(B'),P_1(B'),...,P_n(B'))\end{aligned}$$

Da cui la (D) si riduce ancora alla condizione (C)

Se consideriamo $f,g \in M$ allora per ricondurci al caso precedente basta considerare $f' := f + K$ $g' := g + K$ t.c. le due funzioni siano positive.

Se vale la condizione (D) allora per un ricaviamo la tesi. ■

Questa proposizione è la controparte nel framework costruito da Mongin del Teorema di Harsanyi.

Unicità di P_0 Intuitivamente se due misure di probabilità sono uguali oppure "proporzionali" esistono un'infinità di pesi che danno luogo alla stessa P_0, proprio per questo per avere la possibilità di un unica rappresentazione dobbiamo imporre una certa forma di indipendenza lineare.

Definition 52 *Un famiglia di funzioni reali $\{f_1, f_2, ..., f_n\}$ si dice linearmente indipendente sse $\sum_{i=1}^{n}\alpha_i f_i + \beta = 0$ implica che $\alpha_i = 0$ per $i = 1, 2, ..., n$.*

In altre parole non vogliamo che nessuna funzione sia ricavabile come combinazione lineare delle altre; adesso enunciamo una condizione equivalente alla indipendenza lineare/affine.

Corollary 1 *Data una famiglia di spazi di probabilità* $(\Omega, \mathcal{A}, P_i)$ *per* $i = \{1, 2, ..., n\}$ P_i *è linearmente indipendente da* $\{P_j\}_{j \neq i}$ *sse* $\forall i, j \in \{1, 2, ..., n\}$ $\exists A(i), B(i) \in \mathcal{A}$ *t.c.:*

$$P_i(A(i)) > P_i(B(i)) \text{ and } P_j(A(i)) = P_j(B(i)) \ \forall j \neq i \qquad ((I))$$

Data una famiglia di misure di probabilità nonatomiche indipendenti sappiamo dalla Proposizione 1 che:

$$P_0(A) = \sum_{i=1}^{n} \alpha_i P_i(A)$$

Per ricavare α_i prendiamo $A(i)$ e $B(i)$ ottenuti grazie al precedente corollario e:

$$\begin{aligned} P_0(A(i)) &= \sum_{i=1}^{n} \alpha_i P_i(A(i)) \\ P_0(B(i)) &= \sum_{i=1}^{n} \alpha_i P_i(B(i)) \end{aligned}$$

Ma $P_j(A(i)) = P_j(B(i))$ per ogni $j \neq i$ e $P_i(A(i)) > P_i(B(i))$ quindi possiamo ricavare che:

$$\begin{aligned} P_0(A(i)) - P_0(B(i)) &= \sum_{i=1}^{n} \alpha_i P_i(A(i)) - \sum_{i=1}^{n} \alpha_i P_i(B(i)) \\ \alpha_i &= \frac{P_0(A(i)) - P_0(B(i))}{P_i(A(i)) - P_i(B(i))} \text{ per } i \in \{1, ..., n\} \end{aligned}$$

Vincoli sul segno degli coefficienti

Proposition 20 *Data una famiglia di spazi di probabilità nonatomici e una misura di probabilità* P_0 *su* $(\Omega, \mathcal{A})$ *che soddisfa uno dei criteri di unanimità* (C_1), (C_2) *oppure* (C_3)*. Allora* P_0 *è una combinazione convessa delle misure* $\{P_i\}_{i \in \{1,...,n\}}$.

Proof. Assumiamo che valga l'ipotesi (C_1) $(\geqslant + >)$ grazie lla Proposizione precedente sappiamo che P_0 è nonatomica, utilizzando il teorema di Liapunov sappiamo che $K := Range(P_0, P_1, ...P_n)$ è convesso.

Definiamo $C = \{(x_0, x_1, ...x_n) \in R^{n+1} | x_0 < 0, x_i \geq 0 \ i = 0, 1, ..., n\}$ e $K' = \{x - y | x, y \in K\}$; allora possiamo definire (C_1) in modo equivalente come $C \cap K' = \emptyset$

Basta definire $x_0 := P_0(A) - P_0(B)$ e $x_i := P_i(A) - P_i(B)$ quindi non vogliamo che nel range delle differenze tra la probabilità di due eventi ci sia la possibilità che (C_1) non valga.

Sia C che K' sono due insiemi convessi nello spazio R^{n+1}, se vale la condizione (C_1) i due insiemi sono disgiunti ed applicando il teorema di separazione sappiamo esistere un iperpiano separatore. Possiamo, quindi, concludere che: $\exists \lambda \in R^{n+1}$ e $\alpha \in R$ t.c.:

$$\langle \lambda, x \rangle \geq \alpha \text{ se } x \in C \text{ e}$$
$$\langle \lambda, x \rangle \text{ se } x \in K'$$

Si osserva che l'origine (0) appartiene all'insieme K' mentre $(\varepsilon, 0, ..., 0) \in C$ $\forall \varepsilon > 0$, perciò $\alpha = 0$. Inoltre K' è un insieme "simmetrico" rispetto all'origine, cioè se $x \in K'$ allora $-x \in K'$.

Perciò dalla proprietà del prodotto interno secondo cui $-\langle x, y \rangle = \langle -x, y \rangle$ segue che per ogni elemento x di K' $\langle \lambda, x \rangle = 0$.

Se consideriamo il punto in K' $(\varepsilon, 0, ..., 0)$ sappiamo che $\langle \lambda, (\varepsilon, 0, ..., 0) \rangle \geq 0$ per ogni $\varepsilon > 0$, quindi facendo tendere ε a zero ricaviamo che $\lambda_1 \geq 0$ (il discorso si ripete per i diversi λ_i) mentre se consideriamo $(1, 1, ..., 1) \in K'$ ricaviamo (dalla precedente osservazione $\langle \lambda, x \rangle = 0 \forall x \in K'$) che $\sum_{i=1}^{n} \lambda_i = \lambda_0$. ■

Remark 11 *Si potrebbe cercare di trarre delle implicazioni in termini di prezzi determinati in un mercato che deve aggregare valutazioni di un "asset".*

Remark 12 (C_3) *non implica nessun vincolo per λ_i nel caso di misure di probabilità linearmente dipendenti od affini.*

Corollary 2 *Consideriamo la famiglia di spazi di probabilità $(\Omega, A, P_i)_{i \in \{1,2,...,n\}}$ non atomiche e linearmente indipendenti. e sia P_0 una misura di probabilità sociale che soddisfa la condizione (C_3), allora $P_0 = \sum_{i=1}^{n} \alpha_i P_i$ t.c. $\sum_{i=1}^{n} \alpha_i = 1$ e $\alpha_i > 0$ per ogni i. Inoltre la rappresentazione è unica.*

Basta mettere insieme i precedenti risultati per poter applicare a questo risultato tutte le critiche mosse nella letteratura statistica al "linear pooling rule" [21].

Definition 53 *Per un $x \in [0,1]$ diciamo che P_0 soddisfa la condizione (C_x) sse $\forall A \in \mathcal{A}\ P_i(A) = x$ implica $P_0(A) = x$.*

Definition 54 *P_0 soddisfa la condizione (C') sse P_0 soddisfa la condizione (C_x) per $\forall x \in [0,1]$.*

Proposition 21 *Data una famiglia di spazi di probabilità $(\Omega, \mathcal{A}, P_i)_{i\in\{1,2,\ldots,n\}}$ nonatomiche allora le seguenti condizioni sono equivalenti:*

$$C_{1/2} \Longleftrightarrow C' \Longleftrightarrow C$$

Questo risultato non è centrale per la dimostrazione del teorema ma potrebbe essere una possibile strada per lo studio (non puramente logico Bonanno - Nerhig, 1999) sul problema dell' "agreement" delle credenze.

Corollary 3 *Data una famiglia di spazi di probabilità $(\Omega, \mathcal{A}, P_i)_{i\in\{1,2,\ldots,n\}}$ nonatomiche allora $P_0 = P_1$ sse P_0 e P_1 danno probabilità $\frac{1}{2}$ agli stessi insiemi.*

Aggregazione di preferenze a' la Savage

Enunciamo rapidamente gli assiomi alla base dell'analisi di [50]:

(P1) Gli individui hanno un ordinamento completo sull'insieme degli atti C;

(P2) Dati gli atti a, b, a', b' t.c.:
Per $\omega \in E^c$ $a(\omega) = b(\omega)$ e $a'(\omega) = b'(\omega)$; Per $\omega \in E$ $a(\omega) = a'(\omega)$ e $b(\omega) = b'(\omega)$ e $a \preccurlyeq b$ allora $a' \preccurlyeq b'$.

(P3) Se due atti sono costanti[15] $a = g$ e $a' = g'$ e $E \subseteq \Omega$ un evento non nullo[16] allora $a \preccurlyeq a'$ sse $g \leqslant g'$.

(P4) Dati $f > f', g > g' \in C$ e $E, E' \subseteq \Omega$ e $a_s := \begin{cases} f \text{ se } \omega \in s \\ f' \text{ se } \omega \in s^c \end{cases}$, $b_s := \begin{cases} g \text{ se } \omega \in s \\ g' \text{ se } \omega \in s^c \end{cases}$ per $s = \{E, E'\}$, se $a_E \preccurlyeq b_{E'}$ allora $a_E \preccurlyeq b_{E'}$.

(P5) Devono esistere almeno due conseguenze f, f' t.c. $f' < f$.

[15] Un atto f è costante sse $f(\omega) = f$ per $\forall\omega \in \Omega$.

[16] Un evento $E \subseteq \Omega$ si dice nullo se condizionatamente a E per ogni coppia di atti a, b accade $a \sim b$.

(P6) Presi due atti $a \prec b$ e $f \in C$ allora esiste una partizione $\{P_1, P_2, ...P_n\}$ di Ω t.c. $a'(\omega) = \begin{cases} f \text{ se } \omega \in P_i \\ a(\omega) \text{ se } \omega \in P_i^c \end{cases}$ e $a' \prec b$.

(P7) Se $f \geq g(s)$ dato B per ogni $s \in B$, allora $f \geq g$.

Ma a differenza della costruzione di Savage, Mongin vuole considerare misure di probabilità $\sigma - additive$ concentrandosi però solo sul l'insieme degli atti semplici[17] (Ac) per fare ciò introduce un nuovo postulato:

(P8) Per ogni coppia di eventi $A, B \in \mathcal{A}$ e una successione decrescente di eventi $\{A_p\}_{p\in\mathbb{N}}$ t.c. $\lim_{p\to\infty} A_p = A$ allora:

$$A_p \succcurlyeq_i B \text{ per } \forall p \in \mathbb{N} \Longrightarrow A \succcurlyeq_i B$$

Applicando un ragionamento simile a quello di Savage si può arrivare ad una rappresentazione delle preferenze attraverso un utilità attesa secondo una misura di probabilità $\sigma - additiva$.

Theorem 13 *Date $\succsim_i$ $i = \{1, 2, ..., n\}$ relazioni di preferenza che soddisfano gli assiomi (P1)-(P8). Allora $\succsim_i$ è rappresentabile da una funzione $V : Ac \to \mathbb{R}$ t.c. $V_i(a) = \int_\Omega U_i \circ a dP_i$ dove P_i è una unica misura di probabilità ($\sigma - additiva$) su $(\Omega, \mathcal{A})$ e $U_i : C \to \mathbb{R}$. Ogni V_i' che rappresenta $\succsim_i$ è t.c. U_i' è una trasformazione positiva affine di U_i.*

Si può osservare che (P6) implica la nonatomicità di P_i.

Si può osservare che per ogni $F_i \in \Delta(\Omega)$ attraverso un atto $a : \Omega \to C$ possiamo definire la misura di propabilità immagine su C, definiamo la famiglia di misure di probabilità semplici su C con l'espressione $\mathcal{L}(C)$. Questo suggerisce di poter derivare una funzione di utilità von Neumann - Morgenstern definita su lotterie semplici.

Lemma 17 *Assumendo che le preferenze per gli n individui $\succsim_i$ per $i = \{1, 2, ..., n\}$ soddisfino gli assiomi (P1)-(P8) allora per ogni i $\{P_i \circ a | a \in Ac^s\} = \mathcal{L}(C)$. Definendo $W_i(x) = V_i(a)$ per ogni $x \in \mathcal{L}(C)$ dove $a \in Ac^s$ e $P_i \circ a = x$, abbiamo ricavato una funzione da $\mathcal{L}(C)$ a $\mathbb{R}$. W_i è "mixture preserving".*

Lemma 18 *Data una famiglia di misure $P = (P_1, P_2, ..., P_n)$ nonatomiche e finite su $(\Omega, \mathcal{A})$. Per $\forall A \in \mathcal{A}$ t.c. $P_i(A) > 0$ per qualche $i \in \{1, 2, ..., n\}$ e per*

[17] Atti il cui range è un insieme finito

$\forall m \in \mathbb{N}$ *e* $\forall (p_1, ..., p_m) \in \mathbb{R}^m$ *t.c.* $p_k > 0$ *per* $\forall k \in \{1, 2, ..., m\}$ *e* $\sum_{k=1}^{m} p_k = 1$,

$\exists (A_1, ..., A_m) \subseteq \mathcal{A}$ *t.c.* $\bigcup_{k=1}^{m} A_k = A$ *e* $A_j \cap A_i = \varnothing$ *per* $\forall j \neq i$ *e:*

$$P(A_k) = p_k P(A) \text{ per } \forall k = 1, ..., m$$

Definition 55 *Definiamo diverse condizioni di unanimità nel caso gli individui abbiano solo delle relazioni di preferenza:*

(Pareto Indifference)	Se $\forall a, b \in Ac$ e $a \sim_i b$ per $\forall i = 1, 2, ..., n$ allora $a \sim_0 b$. $(C)(=)$
(Pareto-weak preference)	Se $\forall a, b \in Ac$ e $a \succsim_i b$ per $\forall i = 1, 2, ..., n$ allora $a \succsim_0 b$. $(C_1)(\geqslant)$
(Weak Pareto)	Se $\forall a, b \in Ac$ e $a \succ_i b$ per $\forall i = 1, 2, ..., n$ allora $a \succ_0 b$. $(C_2)(>)$
(Strong Pareto)	Se $\forall a, b \in Ac$ e $a \succsim_i b$ per $\forall i = 1, 2, ..., n$ ed $\exists k \in \{1, 2, ..., n\}$ t.c. $a \succ_k b$ allora $a \succ_0 b$. $(C_3)(\geqslant + >)$

$(C^+) := (C_3) + (C)$

Remark 13 *Le diverse conclusioni sono raggiunte considerando solo atti semplici* (Ac^s)[18].

Per avere un legame tra le diverse definizione Mongin propone di rafforzare l'assioma (P5):

(MAC) $\exists c, c' \in C$ s.t. $i = 0, 1, ..., n$ $c \succ_i c'$ (Minimum agreement on the consequences)

Lemma 19 *Se per* $i = 1, ..., n$ *assumiamo che valgano qui assiomi (P1)-(P8) e (MAC) allora* $(C_2^s) \Longrightarrow (C_1^s)$, *quindi* $(C^{+s}) \Longleftrightarrow (C_3^s)$

Proof. Consideriamo due atti semplici $a, a' \in Ac^s$ t.c. $V_i(a) \geqslant V_i(a')$ per ogni $i = 1, 2, ..., n$ e la partizione ottenuta dall'intersezione di delle partizioni $P := \{\omega \in \Omega | a(\omega) = x\}_{x \in Range(a)}$ e $P' := \{\omega \in \Omega | a(\omega) = x\}_{x \in Range(a')}$, perciò $B := \{B \subseteq \Omega | \exists Pi \in P \cup \{\varnothing\}$ e $P'_j \in P' \cup \{\varnothing\}$ t.c. $B = P'_j \cap P'_i\}$, quindi sapendo che sia a che a' sono costanti sui diversi B possiamo scrivere:

$$\sum_{j=1}^{b} P_i(B_j)\left[U_i(a(B_j)) - U_i(a'(B_j))\right] \geqslant 0 \text{ dove } b := |B|$$

[18] Le precedenti definizioni particolarizzate al caso di atti semplici vengono indicate aggiungendo ad apice ^{c}.

Per ogni numero reale $\varepsilon \in (0,1)$ possiamo definire una B_j^ε e $B_j^{1-\varepsilon} = B_j^{\varepsilon c}|_{B_j}$ per $j = 1, 2, ..., m$ come:

$$B_j^\varepsilon \; : \; = \begin{cases} B_j \text{ se } P_i(B_j) = 0 \text{ per ogni individuo } i \\ B_j^\varepsilon \text{ se } P_i(B_j) > 0 \text{ per almeno un individuo } i \in 1,..,n \end{cases}$$

$$B_j^{1-\varepsilon} \; : \; = \begin{cases} \varnothing \text{ se } P_i(B_j) = 0 \text{ per ogni individuo } i \\ B_j^{\varepsilon c} \text{ se } P_i(B_j) > 0 \text{ per almeno un individuo } i \in 1,..,n \end{cases}$$

B_j^ε nel secondo caso è ricavato grazie al lemma precedente che ci dice che esiste una partizione $\mathcal{A}-misurabile$ $(B_j^\varepsilon, B_j^{\varepsilon c})$di B_j t.c.:

$$P_i(B_j^\varepsilon) = \varepsilon P_i(B_j)$$

Definiamo anche gli atti semplici $a_\varepsilon, a'_\varepsilon$ definiti come:

$$a_\varepsilon(\omega) \;=\; \begin{cases} c \text{ se } \omega \in B_j^\varepsilon \text{ per } j = 1, ..., m \\ a(\omega) \text{ se } \omega \in B_j^{1-\varepsilon} \text{ per } j = 1, ..., m \end{cases}$$

$$a_\varepsilon(\omega) \;=\; \begin{cases} c' \text{ se } \omega \in B_j^\varepsilon \text{ per } j = 1, ..., m \\ a(\omega) \text{ se } \omega \in B_j^{1-\varepsilon} \text{ per } j = 1, ..., m \end{cases}$$

dove c, c' sono le due conseguenze t.c. per ogni agente i $c \succ_i c'$, quindi possiamo concludere che:

$$V_i(a_\varepsilon) - V_i(a'_\varepsilon) \;=\; \underbrace{(1-\varepsilon)\sum_{j=1}^{m} P_i(B_j)\left[U_i(a) - U_i(a')\right]}_{\text{Non negativo (ipotesi inziale)}} + \overbrace{(\varepsilon)\sum_{j=1}^{m} P_i(B_j)\left[U_i(c) - U_i(c')\right]}^{\text{strettamente positivo (MAC)}} > 0$$

per ogni individuo $i \;=\; 1, 2, ..., n$

Quindi valendo la condizione (C_2) $(>)$ possiamo concludere che la precedente disuguaglianza vale anche per $i = 0$ (la relazione di preferenza sociale):

$$V_0(a_\varepsilon) - V_0(a'_\varepsilon) > 0$$

Facendo tendere $\varepsilon \to 0$ e sfruttando il postulato (P8) ricaviamo che $V_0(a_\varepsilon) \geqslant V_0(a'_\varepsilon)$. ■

Remark 14 ***Si deve sottolineare che l'importanza di avere una conseguenza su cui i pareri sono concordi e il ruolo parallelo a quello dell'atto certo. ([40]).***

Remark 15 *Grazie a c, c' possiamo normalizzare la funzione di probabilità i modo tale che $U_i(c) = 1$ e $U_i(c') = 0$ e quindi definendo* $a_A(\omega) := \begin{cases} c \text{ se } \omega \in A \\ c' \text{ se } \omega \in A^c \end{cases}$ *ricaviamo che:*

$$V_i(a_A) = P_i(A) \text{ per } \forall A \in \mathcal{A} \text{ e } \forall i = 1, ..., n$$

Remark 16 *"Ramsey pair"*[19] *per l'elicitazione delle probabilità a partire dalle scelte osservabili; consideriamo due eventi $A, B \in \mathcal{A}$ e definiamo due atti:*

$$a_A(\omega) := \begin{cases} c \text{ se } \omega \in A \\ c' \text{ se } \omega \in B \\ d \text{ se } \omega \in (A \cup B)^c \end{cases}$$

$$a_B(\omega) := \begin{cases} c' \text{ se } \omega \in A \\ c \text{ se } \omega \in B \\ d \text{ se } \omega \in (A \cup B)^c \end{cases}$$

dove $d \in \mathcal{C}$, allora $P_i(A) \geqslant P_i(B)$ sse $V_i(a_A) \geqslant V_i(a_B)$

Definition 56 *Dato un profilo di preferenze $\succsim_0, \succsim_1, \succsim_2, ..., \succsim_n$ si dice che un individuo i è:*

1. "utility dictator" *se $U_0(c) = U_i(c)\ \forall c \in \mathcal{C}$;*
2. "inverse utility dictator" *se $U_0(c) = -U_i(c)\ \forall c \in \mathcal{C}$;*
3. "probability dictator" *se $P_0(\omega) = P_i(\omega)\ \forall \omega \in \Omega$;*
4. "overall dictator" *se è dittatore in termini di utilità e probabilità.*

Definition 57 *Dato un profilo di preferenze $\succsim_0, \succsim_1, \succsim_2, ..., \succsim_n$ si dice che esiste:*

1. *un* "probability agreement" *se $P_i = P_k$ per $\forall i, k \in \{1, ..., n\}$;*

[19] Continuno a valere gli assiomi proposti da Savage ed in particolare l'indipendenza della funzione d'uilità rispetto a Ω.

2. *un* "pairwise utility dependence"(pud) $U_i = \pm U_k$ *per* $\forall i, k \in \{1, .., n\}$ *per una data rappresentazione delle preferenze.*

Remark 17 *Se assumiamo una* $U_i = U_k$ $\forall i, k$ *allora le preferenze degli individui coincideremo per gli atti costanti.*

Remark 18 *Ovviamente* "overall dictatorship" *è una soluzione per l'aggregazione delle preferenze in modo consistente in termini Bayesiani sotto la condizione* (C), (C_1) *e* (C_2)*; anche se anche la* "inverse overall dictatorship" *è una soluzioner se assumiamo solo la condizione di "Pareto indifference"*(C)*. Mentre "utility dictatorship" può risolvere il problema se assumiamo che le preferenze degli individui soddisfino la proprietà di "probability agreement"; e "probability dictatorship" nel caso di "pariwise utility dependence".*

Remark 19 *Se imponiamo restrizioni sulle preferenze possiamo trovare meccanismi di aggregazione (rimandiamo al libro di Austen-Smith e Banks).*

Difficoltà nell'aggregazione delle preferenze se assumiamo (C_3)

Example 1 *Consideriamo due individui le cui preferenze rispettano gli assiomi (P1)-(P8) t.c.* $P_1 \neq P_2$ *e* $c, c', d, d' \in \mathcal{C}$ *t.c.:*

$$\begin{cases} c \succ_i c' \text{ per } i = 1, 2 \\ d \succ_1 d' \text{ e } d' \succ_2 d \end{cases}$$

allora $\succ_0$ *soddisfa la proprietà* (C_2) *sse* $\exists i \in N$ *"probability dictator".*

Proof. Dobbiamo dimostrare che (C_2) implica che $\exists i \in N$ t.c.:

$$\forall A, B \in \mathcal{A}, (A \cap B = \varnothing), P_0(A) = P_0(B) \Longrightarrow P_i(A) = P_i(B) \tag{5.4}$$

Se non vale la condizione (5.4) allora devono esistere due coppie di insiemi disgiunti A_1, B_1, A_2, B_2 t.c.:

$$\begin{cases} [P_0(A_1) - P_0(B_1)][U_0(c) - U_0(c')] = 0 \\ [P_1(A_1) - P_1(B_1)][U_1(c) - U_1(c')] > 0 \end{cases} \tag{5.5}$$

e

$$\begin{cases} [P_0(A_2) - P_0(B_2)][U_0(c) - U_0(c')] = 0 \\ [P_2(A_2) - P_2(B_2)][U_2(c) - U_2(c')] > 0 \end{cases} \tag{5.6}$$

Quindi se definiamo $a_{A_1,c}(\omega) := \begin{cases} c \text{ se } \omega \in A_1 \\ c' \text{ se } \omega \in B_1 \\ d \text{ se } \omega \in (A_1 \cup B_1)^c \end{cases}$ $a_{B_1,c}(\omega) := \begin{cases} c \text{ se } \omega \in B_1 \\ c' \text{ se } \omega \in A_1 \\ d \text{ se } \omega \in (A_1 \cup B_1)^c \end{cases}$ $a_{A_2,c}(\omega) := \begin{cases} c \text{ se } \omega \in A_2 \\ c' \text{ se } \omega \in B_2 \\ d \text{ se } \omega \in (A_2 \cup B_2)^c \end{cases}$, $a_{B_2,c}(\omega) := \begin{cases} c \text{ se } \omega \in B_2 \\ c' \text{ se } \omega \in A_2 \\ d \text{ se } \omega \in (A_2 \cup B_2)^c \end{cases}$ allora l'equazione (5.5) e (5.6) implicano che $a_{A_1} \succ_1 a_{B_1}, a_{A_2} \succ_2 a_{B_2}, a_{A_1} \tilde{}_0 a_{B_1}, a_{A_2} \tilde{}_0 a_{B_2}$ quindi valendo (C_2) deve accadere che $a_{A_1} \preccurlyeq_2 a_{B_1}, a_{A_2} \preccurlyeq_1 a_{B_2}$ cioè[20]:

$$\begin{aligned} P_1(A_2) - P_1(B_2) &\leqslant 0 & (5.7) \\ P_2(A_1) - P_2(B_1) &\leqslant 0. & (5.8) \end{aligned}$$

Se supponiamo che $P_2(A_1) - P_2(B_1) < 0$ allora per ipotesi $(d' \succ_2 d)$:

$$\begin{aligned} [P_2(A_1) - P_2(B_1)][U_2(d) - U_2(d')] &> 0 \\ [P_0(A_1) - P_0(B_1)][U_0(d) - U_0(d')] &= 0 \end{aligned}$$

definendo delle nuove "Ramsey pair" $a_{A_1,d}(\omega) := \begin{cases} d \text{ se } \omega \in A_1 \\ d' \text{ se } \omega \in B_1 \\ e \text{ se } \omega \in (A_1 \cup B_1)^c \end{cases}$ $a_{B_1,d}(\omega) := \begin{cases} d \text{ se } \omega \in B_1 \\ d' \text{ se } \omega \in A_1 \\ e \text{ se } \omega \in (A_1 \cup B_1)^c \end{cases}$ dove $e \in \mathcal{C}$, quindi la precedente disuguaglianza può essere riletta equivalentemente come $V_2(a_{A_1,d}) > V_2(a_{B_1,d})$, cioè $a_{A_1,d} \succ_2 a_{B_1,d}$ da cui data la condizione (C_2) possiamo concludere che:

$$[P_1(A_1) - P_1(B_1)][U_2(d) - U_2(d')] \leqslant 0$$

dato che $d \succ_1 d'$ allora $P_1(A_1) - P_1(B_1) \leqslant 0$ contraddicendo l'ipotesi iniziale (5.5).

Quindi (5.8) e (5.8) devono valere come uguaglianze quindi possiamo riassumere quando dimostrato fino ad ora come:

---

[20] $[P_2(A_1) - P_2(B_1)][U_2(c) - U_2(c')] \leqslant 0$
$[P_1(A_2) - P_1(B_2)][U_1(c) - U_1(c')] \leqslant 0$

$$\begin{pmatrix} P_0(A_1) - P_0(B_1) \\ P_1(A_1) - P_1(B_1) \\ P_2(A_1) - P_2(B_1) \end{pmatrix} = \begin{pmatrix} 0 \\ k \\ 0 \end{pmatrix} \text{ e } \begin{pmatrix} P_0(A_2) - P_0(B_2) \\ P_1(A_2) - P_1(B_2) \\ P_2(A_2) - P_2(B_2) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ k' \end{pmatrix}$$

Grazie al Teorema di Liapunov sappiamo che esistono due eventi disgiunti $C, D \in \mathcal{A}$ t.c. $C \subseteq A_1 \cup A_2$ e $D \subseteq B_1 \cup B_2$ e $P_1(C) - P_1(D) > 0, P_2(C) - P_2(D) > 0$ quindi possiamo costruire un altra coppia di atti di Ramsey dove viene contraddetta la condizione (C_2):

$$a_{C,c}(\omega) := \begin{cases} c \text{ se } \omega \in C \\ c' \text{ se } \omega \in D \\ e \text{ se } \omega \in (C \cup D)^c \end{cases} \quad a_{D,c}(\omega) := \begin{cases} c \text{ se } \omega \in D \\ c' \text{ se } \omega \in C \\ e \text{ se } \omega \in (C \cup D)^c \end{cases}$$

$a_{C,c} \succ_i a_{D,c}$ per $i = 1, 2$ ma $a_{C,c} \sim_0 a_{D,c}$. ■

La dimostrazione ruota sulle ipotesi (MAC) e (C_2) e quello che appare strano è che i due individui hanno la stessa valutazione su un atto perché sono in disaccordo sia relativamente alla probabilità da assegnare a due eventi che sulla valutazione di due conseguenze (ref. Esempio [?]).

Example 2 *Consideriamo due individui le cui preferenze rispettano gli assiomi (P1)-(P8) t.c.* $P_1 \neq P_2$ *e* $c, c', d, d' \in C$ *t.c.:*

$$\begin{cases} c \succ_i c' \text{ per } i = 1, 2 \\ d \succ_1 d' \text{ e } d' \succ_2 d \end{cases}$$

allora $\nexists \succ_0$ *soddisfa la proprietà* (C_3)

Proof. Poichè $(C_3) \implies (C_2)$ allora possiamo concludere che $\succsim_0$ deve essere dittatoriale ma $P_1 \neq P_2$ quindi sono indipendenti perciò per il Corollario "1.1" devono esistere due coppie di eventi disgiunti A_1, B_1, A_2, B_2 t.c.:

$$\begin{aligned} P_1(A_1) &> P_1(B_1) \wedge P_2(A_1) = P_2(B_1) \\ P_2(A_2) &> P_2(B_2) \wedge P_1(A_2) = P_1(B_2) \end{aligned}$$

Quindi considerando le coppie di atti di Ramsey $a_{A_1,c}(\omega) := \begin{cases} c \text{ se } \omega \in A_1 \\ c' \text{ se } \omega \in B_1 \\ e \text{ se } \omega \in (A_1 \cup B_1)^c \end{cases}$,

$a_{B_1,c}(\omega) := \begin{cases} c \text{ se } \omega \in B_1 \\ c' \text{ se } \omega \in A_1 \\ e \text{ se } \omega \in (A_1 \cup B_1)^c \end{cases}$ e $a_{A_2,c}(\omega) := \begin{cases} c \text{ se } \omega \in A_2 \\ c' \text{ se } \omega \in B_2 \\ e \text{ se } \omega \in (A_2 \cup B_2)^c \end{cases}$,

$a_{B_2,c}(\omega) := \begin{cases} c \text{ se } \omega \in B_2 \\ c' \text{ se } \omega \in A_2 \\ e \text{ se } \omega \in (A_2 \cup B_2)^c \end{cases}$ possiamo concludere che:

$$a_{A_1,c} \succ_1 a_{B_1,c} \wedge a_{A_1,c} \sim_2 a_{B_1,c}$$
$$a_{A_2,c} \sim_1 a_{B_2,c} \wedge a_{A_2,c} \succ_2 a_{B_2,c}$$

che data l'ipotesi che $\succ_0$ soddisfi la condizione (C_3) implica che:

$$a_{A_1,c} \succ_0 a_{B_1,c} \wedge a_{A_2,c} \succ_2 a_{B_2,c}$$

e quindi $P_0(A_1) > P_0(B_1)$ e $P_0(A_2) > P_0(B_2)$ contraddicendo la conclusione che $\succsim_0$ sia dittatoriale. ■

Aggregazione di preferenze a' la Savage

Proposition 22 *Consideriamo $n+1$ individui con preferenze che soddisfano gli assiomi (P1)-(P8) ($i = \{0, 1, ..., n\}$) ed assumiamo che $\succsim_0$ soddisfi la condizione (C^s). Se V_i, U_i rappresentano $\succsim_i$ e $W_i : \mathcal{L}(C) \to \mathbb{R}$ ($\mathcal{L}(C)$ è l'insieme delle probabilità semplici su C) associata a V_i, U_i allora:*

(1) esistono $\lambda_1, ..., \lambda_n \in \mathbb{R}$ t.c. $P_0 = \sum_{i=1}^{n} \lambda_i P_i$ e $\sum_{i=1}^{n} \lambda_i = 1$

(2) esistono $\alpha_1, ..., \alpha_n, \beta \in \mathbb{R}$ t.c. $\forall a \in Ac$, $V_0(a) = \sum_{i=1}^{n} a_i V_i(a) + \beta$

(3) gli stessi coefficienti sono t.c. $U_0 = \sum_{i=1}^{n} a_i U_i(a) + \beta$ e $W_0 = \sum_{i=1}^{n} a_i W_i(a) + \beta$

Inoltre esistono $c, c' \in C$ che soddisfano la condizione (MAC) si può scegliere V_i, cioè U_i e W_i in modo tale che $\lambda_i = \alpha_i$ e $\beta = 0$.

Example 3 *Senza imporre condizioni particolari su* $\{P_i\}_{i\in N}$ *e* C *ma mantenendo la condizione (C) per* $\succsim_0$ *si può verificare che al precedente proposizione non vale; consideriamo* $\Omega = \{\omega_1, \omega_2, \omega_3\}$, $N = \{1,2\}$ *e* $C = \{c', c''\}$, *supposto che le preferenze siano rappresentabili come funzioni di utilità attesa* ($\sum_{c\in C} U_i(c) P_i(a^{-1}(c))$) *t.c.* $U_i(c') > U_i(c'')$ *per* $n = 1, 2$, *in questo semplice esempio possiamo normalizzare le utilità degli agenti in modo tale che* $U_i(c') = 1, U_i(c'') = 0$ *e la condizione si riduce alla condizione di unanimità del caso in cui gli agenti esprimano direttamente le probabilità ma se esistessero dei pesi* α_1, α_2 *compatibili con la condizione (2) allora dovrebbero esistere anche i pesi* λ_1, λ_2 *ma questo dal precedente esempio sappiamo non essere possibile.*

Corollary 4 *Se valgono gli assiomi (P1)-(P8) per gli individui allora* $(C^s) \Longleftrightarrow (C)$

Corollary 5 *Date le preferenze degli individui* $i = \{1, ..., n\}$ *che soddisfino gli assiomi (P1)-(P8) e quindi* V_i *siano le loro rappresentazioni in termini di funzione di utilità attesa, le* V_i *sono linearmente indipendenti sse* $\forall i, k \in \{1, ..., n\}, \exists a_i, b_i \in Ac^s$ *t.c.* $V_i(a_i) \neq V_i(b_i) \wedge V_k(a_j) = V_k(b_j)$ *per* $k \neq j$.

Si utilizza la condizione (2) in modo tale da utilizzare i pesi α_i come gli indici della combinazione lineare delle funzione di utilità di attesa.

Corollary 6 *Date le preferenze degli individui* $i = \{1, ..., n\}$ *che soddisfino gli assiomi (P1)-(P8) e quindi* V_i *siano le loro rappresentazioni in termini di funzione di utilità attesa, se le* V_i *sono linearmente indipendenti allora esistono* $a, b \in Ac^s$ *t.c.* $V_i(a) > V_i(b)$ *per* $i = 1, ..., n$.

Proof. Per i diversi individui prendiamo gli atti $a_j, b_j \in Ac$ definiti dal precedente corollario e una n-pla di coefficienti $\lambda_j > 0$ e $\sum_{j=1}^{n} \lambda_j = 1$. Consideriamo $\overline{w} = \sum_{j=1}^{n} \lambda_j V_k(a_j)$ e $\overline{w}' = \sum_{j=1}^{n} \lambda_j V_k(b_j)$ che possono essere espresse come $\overline{w} = V_k(a)$ e $\overline{w} = V_k(b)$ per qualche $a, b \in Ac^s$, visto che il range delle diverse V_k è un insieme convesso perciò visto che $\overline{w} > \overline{w}'$ possiamo concludere di aver trovato $a, b \in Ac^s$ t.c. $V_k(a) > V_k(b)$ per $k = 1, ..., n$. ∎

Per avere una caratterizzazione del metodo di aggregazione Mongin fa riferimento alle funzioni di probabilità ed utilità piuttosto che agli ordinamenti, possiamo normalizzare le funzioni di utilità degli individui imponendo che $V_i(c) =$

$U_i(c)$ per un arbitrario $c \in C$; si può osservare che la normalizzazione non ha effetti sulla indipendenza lineare delle preferenze degli individui.

Possiamo indicare con I e J l'insieme degli indici rispettivamente delle funzioni di utilità e delle funzioni di probabilità linearmente indipendenti.

Corollary 7 *Dati n individui ($i \in N := (1, ..., n)$) le cui preferenze possono essere rappresentate tramite le funzioni di utilità attesa normalizzate:*

$$V_i(a) = \int U_i \circ a dP_i$$

dove U_i non è costante su C e P_i è nonatomica su $(\Omega, \mathcal{A})$. Prese U_0 funzione di utilità non costante e P_0 funzione di probabilità nonatomica e costante allora le seguenti condizioni sono necessarie e sufficienti perché V_0 sia la rappresentazione in termini di utilità attesa relativamente a P_0 di $\succcurlyeq_0$ t.c. soddisfa (C) rispetto alle preferenze $\succcurlyeq_i$ indotte dalle V_i:

1. *Esiste una n-pla di coefficienti $(\alpha_i)_{i \in N} \in \mathbb{R}^n$ e per ogni insieme di probabilità linearmente indipendente $I \subseteq N$ esistono $(\lambda_i)_{i \in N}$ t.c.:*

$$\begin{aligned} V_0 &= \sum_{i \in N} \alpha_i V_i \\ P_0 &= \sum_{i \in I} \lambda_i P_i \\ \text{e per } \forall i &\in I \ \lambda_i U_0 = \alpha_i U_i + \sum_{k \in N \setminus I} \alpha_k \lambda_k U_k \end{aligned}$$

2. *Esiste una n-plu di coefficienti $(\alpha_i)_{i \in N}$ e per ogni insieme di utilità linearmenre indipendenti $J \subseteq I$ esistono $(\mu_j)_{j \in J}$ t.c.:*

$$\begin{aligned} V_0 &= \sum_{i \in N} \alpha_i V_i \\ U_0 &= \sum_{j \in J} \lambda_i U_j \\ \text{e per } \forall j &\in J: \ \mu_j P_0 = \alpha_j P_j + \sum_{k \in N \setminus J} \alpha \mu_{kj} P_k \end{aligned}$$

dove per ogni k (μ_{kj}) sono le cordinate di U_k sulla base di J.

Proof. (Condizione Necessaria) dalla Proposizione (22) sappiamo che esistono dei coefficienti $(\alpha_i)_{i\in N}$ e $(\lambda_i)_{i\in I}$ per le $|I|$ probabilità linearmente indipendenti t.c. $V_0 = \sum_{i\in N}\alpha_i V_i$ e $P_0 = \sum_{i\in I}\lambda_i P_i$ quindi:

$$\begin{aligned} P_0U_0 &= \sum_{i\in I}\lambda_i P_i U_0 \\ P_0U_0 &= \sum_{i\in N}\alpha_i P_i U_i = \sum_{i\in I}\alpha_i P_i U_i + \sum_{k\in N\setminus I}\alpha_k\left(\sum_{i\in I}\lambda_{ik}P_i\right)U_i \end{aligned}$$

uguagliando le due espressioni possiamo identificare i coefficienti delle funzioni di probabilità P_i, $i \in I$.

(Condizione sufficiente) L'esistenza dei coefficienti $(\alpha_i)_{i\in N}$ t.c $V_0 = \sum_{i\in N}\alpha_i V_i$ implica che la V_0 soddisfa il principio di unanimità (C) per verificare che V_0 è la funzione di utilità attesa associata a P_0 basta considerare un atto $a \in Ac$ scriviamo la funzione $\lambda_i U_0 = \alpha_i U_i + \sum_{k\in N\setminus I}\alpha_k\lambda_{ik}U_k$ composta per a:

$$\lambda_i U_0 \circ a = \alpha_i U_i \circ a + \sum_{k\in N\setminus I}\alpha_k\lambda_{ik}U_k \circ a$$

ed integrare rispetto a P_i:

$$\lambda_i \int U_0 \circ a dP_i = \alpha_i \int U_i \circ a dP_i + \sum_{k\in N\setminus I}\alpha_k\lambda_{ik}\int U_k \circ a dP_i \ \forall i \in I$$

sommando per i:

$$\begin{aligned} \sum_{i\in I}\lambda_i \int U_0 \circ a dP_i &= \sum_{i\in I}\alpha_i \int U_i \circ a dP_i + \sum_{i\in I}\sum_{k\in N\setminus I}\alpha_k\lambda_{ik}\int U_k \circ a dP_i \\ \text{dato che} \sum_{i\in I}\lambda_i P_i &= P_0 \text{ e } \sum_{i\in I}\lambda_{ik}P_i = P_k \\ \int U_0 \circ a dP_0 &= \sum_{i\in N}\int \alpha_i U_i \circ a dP_i \\ V_0(a) &= \sum_{i\in N}\alpha_i V_i(a) \end{aligned}$$

La seconda perte della dimostrazione ricalca la prima. ■

Proposition 23 *Dato un insieme di individui $N = \{0, 1, ..., n\}$ con preferenze, che soddisfano gli assiomi (P1)-(P8), $i = 0, 1, ..., n$, supponiamo che le preferenze sociali soddisfano il principio di unanimità (C), e prendiamo una qualunque loro rappresentazione funzionale (P_i, U_i) $i = 0, 1, ..., n$.*

Se le funzioni di probabilità $\{P_1, P_2, ..., P_n\}$ sono indipendenti allora esiste un "utility or inverse utility dictator" i^*. Se inoltre le $U_i \neq \alpha U_j + \beta$ per $\forall i \neq j$ allora i^* è anche un "probability dictator".

Se le funzioni di utilità $\{U_1, ..., U_n\}$ sono linearmente indipendenti allora esiste un "probability dictator" j^*. Se inoltre le $P_i \neq P_j$ $\forall i \neq j \in N$ allora j^* è anche un "utility dictator" o un "inverse utility dictator".

Proof. Consideriamo le funzioni di utilità e utilità attesa normalizzate, supponiamo che $\{P_1, ..., P_n\}$ siano indipendenti, allora P_0, U_0, V_0 devono soddisfare le condizione del Corollario precedente nel caso in cui $I = N$ e quindi $\lambda_i U_0 = \alpha_i U_i$ per $\forall i \in N$.

La dittatorialità deriva dal fatto che almeno un λ_{i^*} è strettamente maggiore di zero come anche α_{i^*} data l'ipotesi che U_0 non sia identicamente uguale a zero. D'altra parte se esistesse un altro $\lambda_j \neq 0$ allora U_j sarebbe linearmente dipendente; quindi U_0 è dittatoriale. ■

Ovviamente la compresenza si indipendenza per le funzioni di utilità e probabilità comporta che l'unica relazione di preferenza è dittatoriale "overall dictatorial".

Se vale la condizione (MAC) esclude la possibilità che "inverse utility dictatorship" e "overall inverse dictatorship" non possono essere preferenze sociali compatibili con la proprietà (C).

Proposition 24 *Se* $\dim\{P_1, ..., P_n\} = \dim\{V_1, ..., V_n\}$ *allora le uniche relazioni di preferenza compatibili con il pincipio (C) sono "Utility dictatorship" e "Inverse utility dictatorship".*

Proposition 25 *Se* $\dim\{U_1, ..., U_n\} = \dim\{V_1, ..., V_n\}$ *allora l'unica relazione di preferenza compatibile con il pincipio (C) è "Probability dictatorship".*

Example 4 $N = \{1, 2, 3\}$ $\dim\{P_1, P_2, P_3\} = \dim\{U_1, U_2, U_3\} = 2$, $\dim\{V_1, V_2, V_3\} = 3$ *quindi $I = J = 2$. Se assumiamo che:*

$$P_3 = \frac{1}{2}P_1 + \frac{1}{2}P_2 \text{ e } U_3 = 4U_1 + 4U_2$$

Allora la funzione $V_0 = -\frac{1}{3}V_1 + \frac{2}{3}V_2 + \frac{1}{3}V_3$*, cioè la funzione d'utilità attesa può essere scritta come:*

$$\begin{aligned} V_0(a) &= -\frac{1}{3}\int U_1 \circ adP_1 + \frac{2}{3}\int U_2 \circ adP_2 + \frac{4}{3}\int U_1 + U_2 d\left(\frac{1}{2}P_1 + \frac{1}{2}P_2\right) = \\ &= -\frac{1}{3}\int U_1 \circ adP_1 + \frac{2}{3}\int U_2 \circ adP_2 + \frac{4}{3}\int U_1 d\left(\frac{1}{2}P_1 + \frac{1}{2}P_2\right) + \frac{4}{3}\int U_2 d\left(\frac{1}{2}P_1 + \frac{1}{2}P_2\right) = \\ &= \frac{1}{3}\int U_1 \circ adP_1 + \frac{2}{3}\int U_1 d(P_2) + \frac{4}{3}\int U_2 \circ adP_2 + \frac{2}{3}\int U_2 d(P_1) = \\ &= \int U_1 \circ ad\left(\frac{1}{3}P_1 + \frac{2}{3}P_2\right) + \int 2U_2 \circ ad\left(\frac{1}{3}P_1 + \frac{2}{3}P_2\right) = \int (U_1 + 2U_2) \circ ad\left(\frac{1}{3}P_1 + \frac{2}{3}P_2\right) \end{aligned}$$

Risultati simili possono essere ottenuti se imponiamo la condizione di unanimità (C_3) avendo assunto che le preferenze soddisfino la proprietà (MAC).

5.4 Utilitarian Aggregation of Beliefs and Tastes - I. Gilboa, D. Samet and D. Schmeidler

Questo working paper vuole dimostrare come non imponendo la proprietà di Pareto o unanimità per la funzione di decisione sociale i risultati d'impossibilità vengano meno.

Prima di tutto gli autori sottolineano l'importanza di utilizzare un'impostazione a la Ramsey, De Finetti, Savage che tenga conto delle probabilità (soggettive) e della funzione di utilità (cardinale) come due aspetti distinti caratterizzanti un individuo; in particolare gli autori ritengono di fondamentale importanza situazioni dove le preferenze sono identiche ma le credenze no (l'esempio utilizzato è quello del duello).

Per coprendere le implicazioni della condizione di Pareto si può far riferimento agli esempi proposti rispettivamente da [10] e [32]:

1. [10][21]

[21] Broome parte dalla sua analisi di bene ("good" e "good relation") e dall'osservazione che la funzione di preferenza sociale è solo un meccanismo che deve essere valutato in base alle scelte che comporta.

Il suo ragionamento si basa sulla distinzione tra "A è preferito a B" e "A è meglio di B", parte questa distinzione è simile alla richiesta avanzata da Hammond di valutare le conseguienze degli atti e non gli atti stessi.

Questo perchè una meccanismo di scelta sociale dovrebbe, da un punto di vista etico pensare il bene derivante a ciascun individuo in modo uguale.

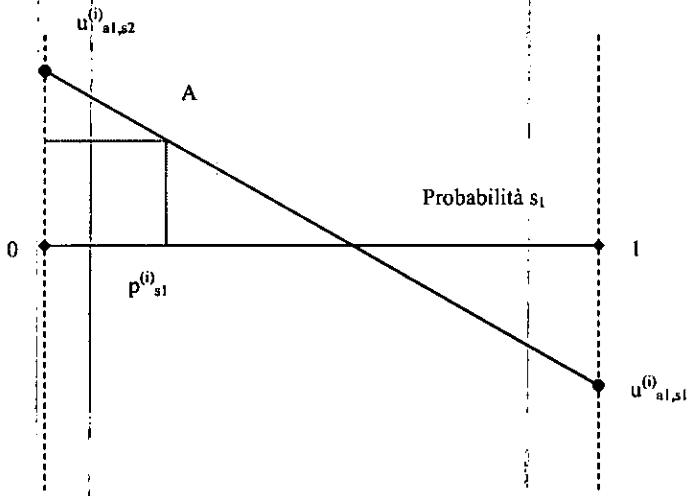
P_i (Evento 1) = 0.7 P_j (Evento 1) = 0.3		
	Evento 1	Evento 2
Atto/Prospet A	$(2,2)$	$(2,2)$
Atto/Prospect B	$(3,0)$	$(0,3)$

Supposto che le preferenze degli individui siano rappresentabili tramite una funzione d'utilità attesa, sia entrambi gli agenti $\{i,j\}$ entrambi gli agenti preferiscono B a A; qundi se imponiamo il principio di Pareto dovremmo concludere che "B è socialmente preferibile a A". Ma proprio da questo esempio Broome inizia la sua analisi di cosa possa definirsi "Bene" e le implicazioni del Teorema di Harsanyi di cosa sia un "Bene Sociale".

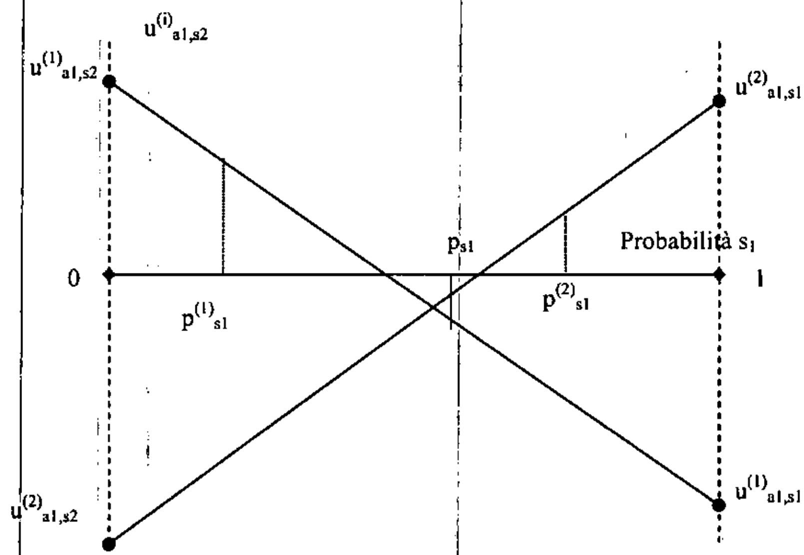
2. [32]

Consideriamo due individui $\{1,2\}$, che devono scegliere due azioni/atti $\{a_1,a_2\}$ nel caso gli stati del mondo siano solo due $\{s_1,s_2\}$; in questo particolare caso le probabilità assegnate dai due individui sono identificare dalla probabilità che l'individuo i assegna alllo stato del mondo

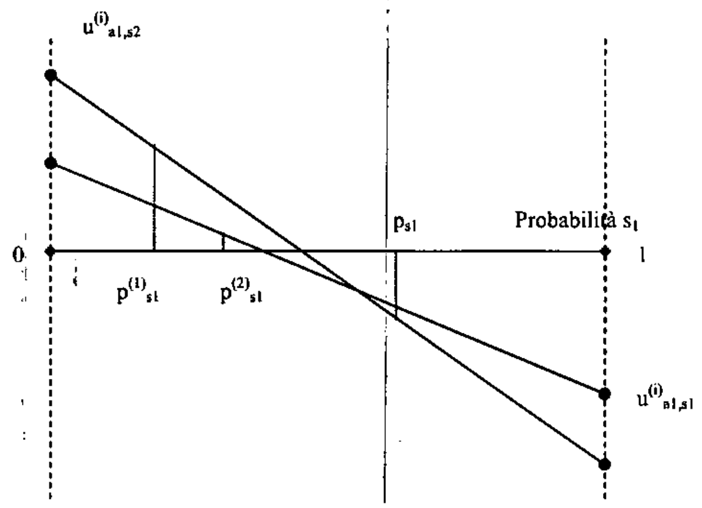
s_1, quindi le preferenze di un individuo sono caratterizzate da un vettore di probabilità $\left(p_{s_1}^{(i)}, p_{s_1}^{(i)}\right)$ e una matrice di utilità $\begin{bmatrix} u_{s_1,a_1}^{(i)} & u_{s_1,a_2}^{(i)} \\ u_{s_2,a_1}^{(i)} & u_{s_2,a_2}^{(i)} \end{bmatrix}$. Per semplicità Hylland e Zeckhauser assumono che $u_{s_j,a_2}^{(i)} = 0$ per $j = 1,2$ e coneguentemente l'utilità attesa dell'azione a_1 è rappresentata da un punto lungo il segmento che unisce i punti $\left(p_{s_1}^{(i)} = 0, u_{s_1,a_1}^{(i)}\right)$ e $\left(p_{s_1}^{(i)} = 1, u_{s_2,a_1}^{(i)}\right)$ (Figura 1).



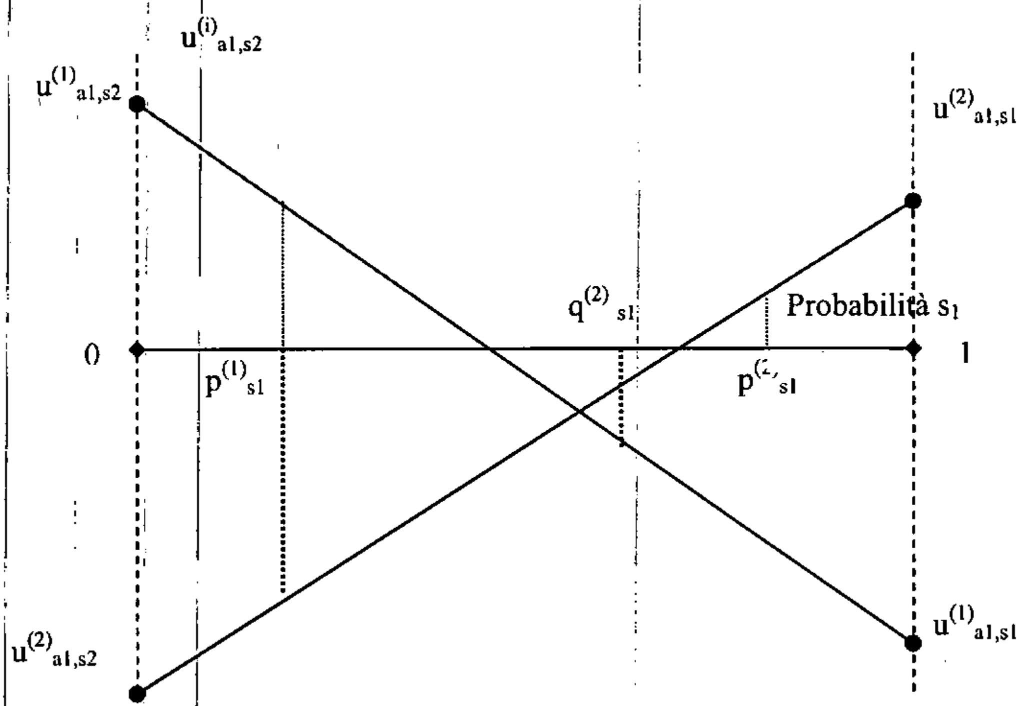
Ad esempio la precedente figura indica che l'individuo i preferisce l'azione a_1. Adesso consideriamo la situazione illustrata dalla Figura 2; entrambi gli individui preferiscono l'azione a_1 quindi secondo il principio di Pareto la a_1 dovrebbe essere scelto d'altra parte se consideriamo come misura di probabilità sociale quella ricavata come combinazione lineare convessa di quelle individuali (ad esempio p_{s_1}) allora entrambi gli individui preferiscono a_2 quindi dovrebbe essere scelta a_2. Quindi a parità di misura di probabilità sociale abbiamo che il principio di Pareto ci indica due scelte differenti.



Se $p_{s_1} \notin \left[\min\left\{p_{s_1}^{(1)}, p_{s_1}^{(2)}\right\}, \max\left\{p_{s_1}^{(1)}, p_{s_1}^{(2)}\right\}\right]$ per raggiungere lo stesso tipo di risultato basta considerare la situazione rappresentato dalla figura 3.



L'unica alternativa, che non sia dittatoriale $(p_{s_1} = p_{s_1}^{(i)})$ deve essere tale che esistano almeno due profili di probabilità $\left(p_{s_1}^{(1)}, p_{s_1}^{(2)}\right)$ e $\left(q_{s_1}^{(1)}, q_{s_1}^{(2)}\right)$ t.c. nel primo caso $p_{s_1} = p_{s_1}^{(1)}$ e nel secondo $p_{s_1} = q_{s_1}^{(2)}$ consideriamo i due seguenti profili $\left(p_{s_1}^{(1)}, p_{s_1}^{(2)}, u_1, u_2\right)$ e $\left(q_{s_1}^{(2)}, p_{s_1}^{(1)}, u_1, u_2\right)$ figura 4, nel primo caso la scelta del gruppo sarà a_1 mentre nel secondo caso a_2 (sempre valendo la condizione di Pareto) quindi per avere una contraddizione[22] deve accadere che $p_{s_1}\left[q_{s_1}^{(2)}, p_{s_1}^{(1)}\right] \neq p_{s_1}\left[p_{s_1}^{(1)}, p_{s_1}^{(2)}\right] = p_{s_1}^{(1)}$ con un ragionamento analogo[23] si può concludere che $p_{s_1}\left[q_{s_1}^{(2)}, p_{s_1}^{(1)}\right] \neq p_{s_1}\left[q_{s_1}^{(1)}, q_{s_1}^{(2)}\right] = q_{s_1}^{(2)}$ ma allora $p_{s_1}\left[q_{s_1}^{(2)}, p_{s_1}^{(1)}\right] \neq \left\{p_{s_1}^{(1)}, q_{s_1}^{(2)}\right\}$ (contraddizione).



Alla luce del precedente esempio Raiffa propose di emendare il principio di Pareto allorchè questo violi la massimizzazione dell'utilità attesa del gruppo; propio questa proposta suggeriscre un indebolimento della condizione di unanimità imponendola solo nei casi in cui gli agenti abbiamo delle credenze uguali ma preferenze differenti.

Definition 58 *Una alternativa (atto) è definita lotteria sse tutti gli individui*

[22] Dato che il profilo delle utilità è lo stesso nei due casi.
[23] Basta considerare la coppia $\left(q_{s_1}^{(2)}, p_{s_1}^{(1)}, u_1, u_2\right)$ e $\left(q_{s_1}^{(1)}, q_{s_1}^{(2)}, u_1, u_2\right)$

sono d'accordo sulla distribuzione di probabilità indotta sullo spazio delle conseguenze dalla stessa alternativa.

Definition 59 *Una funzione di scelta sociale rispetta la condizione di Pareto in senso debole sse se ogni individuo è indifferente tra due lotterie allora anche la società è indifferente tra le due.*

5.4.1 Notazione

(S, Σ) è uno spazio misurabile dove S è l'insieme degli stati del mondo e Σ è un $\sigma-$algebra di S. (X, Ξ) è lo spazio misurabile delle conseguenze. $A := \{a|a : S \to X,$ a è una funzione $\Sigma - misurabile\}$ è lo spazio degli atti o alternative. La nostra società è composta da un insieme di individui $N := \{1, ..., n\}$

Assumiamo che esista per ogni individuo i una relazione di (pre)ordine completa sugli atti $\succeq_i$ed assumiamo che sia rappresentabile ([50]) tramite un'utilità attesa:

$$a \succeq_i b \Leftrightarrow \int_S u(a(s))\, d\mu_i \geq \int_S u(b(s))\, d\mu_i$$

"μ_i è $\sigma-additiva$, non-atomistica, e u_i non è costante per ogni conseguenza": ref Savage Villegas Arrrow

Definiamo l'insieme di tutti gli eventi a cui tutti gli individui delle popolazione assegnano la stessa probabilità come:

$$\Lambda := \{E \in \Sigma | \text{ per ongi } 1 \leq i, j \leq n \ \mu_i(E) = \mu_j(E)\}$$

quindi un atto a è una lotteria sse per ogni conseguenza $y \in Y$ $a^{-1}(y) \in \Lambda$.

Obs per ora una lotteria è una funzione misurabile mentre in vNM è una distribuzione su (X, Ξ).

Definition 60 *Per ogni coppia di lotterie a,b se $a\tilde{}_i b$ per ogni $i \in N$ allora $a\tilde{}_o b$*

Theorem 14 *La* condizione di Pareto (ristretta) *è soddisfatta sse μ_0 è una combinazione affine di $\{\mu_i\}_{i\in N}$ e u_0 è una combinazione lineare di $\{u_i\}_{i\in N}$.*

Remark 20 *Il Teorema è una diretta conseguenza del Teorema visto in [40].*

Proof. Funzione di probabilità di gruppo z_0.

Consideriamo le misure prodotto definite come $\mu := (\mu_1, \mu_2, ..., \mu_n)$ e $\widehat{\mu} := (\mu_0, \mu_1, \mu_2, ..., \mu_n)$ dove μ_0 è la misura di probabilità sociale; si osserva che μ è la proiezione di $\widehat{\mu}$ su S^n e deifiniti Z e $\widehat{Z}$ il range delle due misure si può osservare che se $\widehat{z} \in \widehat{Z}$ allora $\widehat{\mu} - \widehat{z} \in \widehat{Z}$.

Dal Teorema di Liapunov possiamo concludere che Z e $\widehat{Z}$ sono convessi ([18]).

Se S è convesso allora $\Delta(S)$ è convesso.

Step 1[24]: Se $\left(z_0, \frac{1}{2}\mu(S)\right) \in \widehat{Z}$ allora $z_0 = \frac{1}{2}\mu_0(S)$

Infatti supponiamo che $z_0 > \frac{1}{2}$ allora $\exists E \in \Sigma$ t.c. $\mu(E) = \frac{1}{2}\mu(S)$ e $\mu_0(E) > \frac{1}{2}$ quindi se consideriamo le lotterie (v. definizione sopra) a, b così definite:

$$a(s) = \begin{cases} x \text{ se } s \in E \\ y \text{ se } s \in E^c \end{cases}$$

$$b(s) = \begin{cases} y \text{ se } s \in E \\ x \text{ se } s \in E^c \end{cases}$$

dove $x > y$ abbiamo che $a \sim_i b$ per $\forall i \in N$ ma $a \succ_0 b$ contraddicendo la *condizione di Pareto (ristretta)*

Step 2: (Unicità) $\forall z \in Z$, $\exists! z_0$ t.c. $(z_0, z) \in \widehat{Z}$

Infatti supponiamo che $\widehat{z} := (z_0, z)$ e $\widehat{w} := (w_0, z)$ appartengano a $\widehat{Z}$ allora per quanto detto prima $\widehat{\mu}(S) - \widehat{w} \in \widehat{Z}$ ma data la convessità di $\widehat{Z}$ allora:

$$\frac{1}{2}\widehat{z} + \frac{1}{2}\left(\widehat{\mu}(S) - \widehat{w}\right) \in \widehat{Z}$$

cioè

$$\left(\frac{1}{2}z_0 + \frac{1}{2}\left(\mu_0(S) - w_0\right), \frac{1}{2}\mu(S)\right) \in \widehat{Z}$$

Dal precedente passo sappiamo che questa inclusione implica che $\frac{1}{2}z_0 + \frac{1}{2}\left(\mu_0(S) - w_0\right) = \frac{1}{2}\mu_0(S)$ da cui l'assurdo.

OSS: Non si richiede che μ e μ_0 siano misure di probabilità

Step 3: (La funzione di probabilità di gruppo è omogenea di grado uno) $\forall z, w \in Z$ ed ogni $\beta \in [0,1]$ $z_0(\beta z + (1-\beta)w) = \beta z_0(z) + (1-\beta) z_0(w)$

Se $(z_0(z), z)$ e $(z_0(w), w) \in \widehat{Z}$ e data la convessità di $\widehat{Z}$ sappiamo che anche $(\beta z_0(z) + (1-\beta) z_0(w), \beta z + (1-\beta) w) \in \widehat{Z}$ ma data l'unicità dell'immagine tramite z_0, segue la tesi.

Quindi z_0 è una funzione affine e quindi rappresentabile come $z_0(z)(E) =$

[24] Questa implicazione può essere vista come un caso particolare del principio di unanimità per le misure di probabilità ($\left(z_0, \frac{1}{2}\right) \in \widehat{Z}$ implica che $z_0 = \frac{1}{2}$) v. Mongin.

$\sum_{i\in N}\lambda_i z_i(E)$ e quindi $\sum_{i\in N}\lambda_i = 1$

Funzione di utilità di gruppo $u_0 = u_0(u_1, ..., u_n)$

Step 4: $\forall\{p_1, ...p_m\} \in R_+$ t.c. $\sum_{k=1}^{m} p_k = 1$, $\exists\{E_1, ..., E_m\}$ t.c. $\mu(E_k) = p_k\mu(S)$.

In questo modo si è costruita una relazione biunivoca tra le lotterie (semplici o equivalentemente a valori finiti) sullo spazio delle conseguenze con gli atti che inducono la stessa distribuzione su X.

OSS. gli atti che inducono la stessa distribuzione su X sono indifferenti per ogni membro dela popolazione.

Se su questo nuovo spazio degli atti applichiamo il teorema di Harsanyi possiamo concludere che u_0 è una combinazione lineare delle $\{u_i\}_{i\in N}$. ■

5.5 Conclusioni

In questa breve rassegna ho cercato di chiarire anche alla luce della letteratura precedente i contributi di [42] e [33] [GSS], in particolar modo ho illustrato come Mongin partendo dall'osservazione che l'ipotesi di una relazione di preferenza sociale a'la Von Neumann - Morgenstern sottointenda o almeno implichi la comparabilità tra le le preferenze individuali vuole esplorare altre possibili ipotesi relative alle preferenze sociali ed individuali, mentre [GSS] portano avanti l'osservazione prima di [48] , poi di [32] e [10] sulla possibilità che gli individui concordino nel preferire una alternativa ad un altra solo perché hanno preferenze e credenze discordanti; tradendo in parte quella che può essere una giustificazione intuitiva del principio di Pareto.

Devo aggiungere che in questa ricerca viene completamente ignorata la possibilità di preferenze "state dependent" e viene sotto - rappresentato il contributo proveniente dal campo della statistica anche se si può far riferimento all'articolo di [21] e [53] per osservare l'attenzione rivolta all'argomento.

Remark 21 *Se si assume che le preferenze siano determinabili in maniera indipendente dalle probabilità si potrebbe auspicare la possibilità di separare i due problemi di aggregazione in un uno più prettamente statistico informativo e in un secondo più vicino alla Social-Choise e alla Teoria dei giochi, più specificatamente al "mechanism design".*

Remark 22 *L'articolo di [21] sottolinea come alla base dei problemi decisionali di gruppi e di "pooling " di opinioni ci sia il problema di definire delle condizioni* *che la funzione di aggregazione deve rispettare e il problema di una caratterizzazione del consenso*

Per concludere voglio citare il recente contributo di Charles Blackorby, Walter Bossert e David Donaldson (2003) [14] che semplifica in modo sensibile la trattazione del problema della aggregazione delle preferenze tramite l'utilizzo il concetto dei "prospects".

Bibliography

[1] K. Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, 4:328–346, 1950.

[2] R. J. Aumann. Interactive epistemology i: Knowledge. *International Journal of Game Theory*, 28(3):263–300, 1999.

[3] P. Battigalli. Rationalizability in infinite, dynamic games with incomplete information. *Research in Economics*, 57(1):1–38, 2003. TY - JOUR.

[4] P. Battigalli. Rationalization in signaling games: Theory and applications. *mimeo*, 2003.

[5] P. Battigalli and M. Siniscalchi. Hierarchies of conditional beliefs and interactive epistemology in dynamic games. *Journal of Economic Theory*, 88(1):188–230, 1999.

[6] P. Battigalli and M. Siniscalchi. Strong belief and forward induction reasoning. *Journal of Economic Theory*, 106(2):356–391, 2002. TY - JOUR.

[7] P. Battigalli and M. Siniscalchi. Rationalization and incomplete information. *Advances in Theoretical Economics*, 3, 2003.

[8] J. P. Benoit and J. Dubra. Information revelation in auctions. *mimeo*, 2004.

[9] A. Brandenburger and E. Dekel. Rationalizability and correlated equilibria. *Econometrica*, 55:1391–1402, 1987.

[10] J. Broome. Utilitarianism and expected utility. *The journal of Philosophy*, 84:405–422, 1987.

[11] G. Cappelletti. On the equivalence between semantic and syntactic approach. 2003.

[12] G. G. L. Cappelletti. A note on the characterization of strong d-rationalizability in static and extensive form games. 2004.

[13] C. Chang and H. J. Keisler. *Model Theory*. North Holland, 1998.

[14] W. B. Charles Blackorby and D. Donaldson. HarsanyiŠs social aggregation theorem: A multi-profile approach with variable-population extensions. *mimeo*, 2003.

[15] V. Crawford. Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Economic Review*, 93(1):133–149, 2003.

[16] V. P. Crawford and J. Sobel. Strategic information transmission. *Econometrica: Journal of the Econometric Society*, 50(6):1431–1451, 1982.

[17] C. D'Aspremont and L. Gevers. Equity and the infromational basis of collective choice. *Review of Economic Studies*, 44:199–209, 1977.

[18] R. M. Dudley. *Real Analysis and Probability*. Cambridge Press, 2002.

[19] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about knowledge*. MIT Press, 1995.

[20] J. Farrell. Meaning and credibility in cheap-talk games. 5, 1993.

[21] C. Genest and J. V. Zidek. Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 1:114–135, 1986.

[22] P. R. Halmos. *Measure Theory*. Springer Verlag, 1974.

[23] P. J. Hammond. Equity in two person situations: Some consequences. *Econometrica*, 47:1127–1136, 1979.

[24] P. J. Hammond. Ex-ante and ex-post welfare optimality under uncertainty. *Economica*, 48:235–250, 1981.

[25] P. J. Hammond. Equity, arrow's conditions, and rawls' difference principle. *Econometrica*, 44:793–804, 76.

[26] J. Harsanyi. Cardinal welfare, individualistic ethics, and interpersonal comparison of utility. *Journal of Political Economics*, 63:309–321, 1955.

[27] A. Heifetz and D. Samet. Knowledge spaces with arbitrarily high rank. *Games and Economic Behavior*, 22(2):260–273, 1998. TY - JOUR.

[28] A. Heifetz and D. Samet. Topology-free typology of beliefs. *Journal of Economic Theory*, 82(2):324–341, 1998. TY - JOUR.

[29] A. Heifetz and D. Samet. Coherent beliefs are not always types. *Journal of Mathematical Economics*, 32(4):475–488, 1999. TY - JOUR.

[30] A. Heifetz and D. Samet. Hierarchies of knowledge: An unbounded stairway. *Mathematical Social Sciences*, 38(2):157–170, 1999. TY - JOUR.

[31] T.-W. Hu. General rationalizability and its robustness for strategic form games with incomplete information. 2004. Mimeo.

[32] A. Hylland and R. Zeckhauser. The impossibility of bayesian group decision making with separate aggregation of beliefs and values. *Econometrica*, 47:1321–1336, 1979.

[33] D. S. I. Gilboa and D. Schmeidler. Utilitarian aggregation of beliefs and tastes. *mimeo*, March 2001.

[34] M. e. A. Kaneko. A map of common knowledge logics. *Studia Logica*, 71:57–86, 2002.

[35] H. J. Keisler. *Model Theory for Infinitary Logic.* North Holland, 1991.

[36] M. C. Kemp and Y. Ng. On the existence of social welfare functions, social orderings and social decision functions. *Economica*, 43:59–66, 1976.

[37] V. Krishna and J. Morgan. The art of conversation: Eliciting information from experts through multi-stage communication. *mimeo*, 2003.

[38] D. Marker. *Model Theory: An Introduction.* Springer, 2002.

[39] J. Mertens and S. Zamir. Formulation of bayesian analysus for games with incomplete information. *International Journal of Game Theory*, 14:1–29, 1985.

[40] B. D. Meyer and P. Mongin. A note on affine aggregation. *Economics Letters*, 47:177–183, 1995.

[41] P. Mongin. Harsanyi's aggregation theorem: Multiple-profile version and unsettled questions. *Social Choice and Welfare*, 11:331–354, 1994.

[42] P. Mongin. Consistent bayesian aggregation. *Journal of Economic Theory*, 66:313–351, 1995.

[43] J. Nash. Non-cooperative game. *The Annals of Mathematics*, 54:286–295, 1951.

[44] M. Okuno-Fujiwara, A. Postlewaite, and K. Suzumura. Strategic information revelation. *The Review of Economic Studies*, 57(1):25–47, 1990. FLA 00346527 Society for Economic Analysis Limited Copyright 1990 The Review of Economic Studies Ltd.

[45] R. P. Parks. An impossiblility theorem for fixed preference: A dictatorial bergson-samuelson welfare function. *The Review of Economic Studies*, 43:447–4450, 1976.

[46] P. Pollak. Bergson - samuelson social welfare functions and the theory of social choice. *The Quarterly journal of Economics*, 93:73–90, 1979.

[47] M. Rabin. Communication between rational agents. *Journal of Economic Theory*, 51(1):144–170, 1990.

[48] H. Raiffa. *Decision Analysis*. Addion-Wesley, 1970.

[49] A. Rubinstein. The single profile analogues to multi profile theorems: Mathematical logic's approach. *International Economic Review*, 25:719–730, 1984.

[50] L. Savage. *The Foundation of Statistics*. Dover Publications, 1954.

[51] A. Sen. *Collective Choice and Social Welfare*. Holden Day, 1974.

[52] M. Shimoji and J. Watson. Conditional dominance, rationalizability, and game forms. *Journal of Economic Theory*, 83:161 195, 1998.

[53] J. B. K. T. Seidenfeld and M. Schervish. On the shared preferences of two bayesian decision makers. *The Journal of Philosophy*, 86:225–244, 1989.

[54] T. C. Tan and S. R. D. C. Werlang. The bayesian foundations of solution concepts of games. *Journal of Economic Theory*, (45), 1988.