INCOMPLETE INFORMATION AND ROBUSTNESS IN STRATEGIC ENVIRONMENTS

Antonio Penta
University of Pennsylvania, apenta@ssc.wisc.edu

Follow this and additional works at: http://repository.upenn.edu/edissertations
Part of the Economic Theory Commons

Recommended Citation

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/edissertations/134
For more information, please contact repository@pobox.upenn.edu.
INCOMPLETE INFORMATION AND ROBUSTNESS IN STRATEGIC ENVIRONMENTS

Abstract
Game theoretic modeling involves making assumptions on agents' infinite hierarchies of beliefs. These assumptions are understood to be only approximately satisfied in the actual situation. Thus, the significance of game theoretic predictions depend on robustness properties of the solution concepts adopted. Chapter 1 discusses recent results in this research area and their relations with the results obtained in the subsequent chapters. Chapter 2 explores the impact of misspecification of higher order beliefs in static environments, when arbitrary common knowledge assumptions on payoffs are relaxed. (Existing literature focuses on the extreme case in which all such assumptions are relaxed.) Chapter 3 provides a characterization of the strongest predictions, for dynamic games, that are "robust" to possible misspecifications of agents' higher order beliefs, and shows that such characterization depends on modeling assumptions that have hitherto received little attention in the literature (namely, the distinction between knowledge and certainty), raising novel questions of robustness. Chapter 4 develops a methodology to address classical questions of implementation, when agents' beliefs are unknown to the designer and their private information changes over time. The key idea is the identification of a solution concept that allows a tractable analysis of the full implementation problem: Full "robust" implementation requires that, for all models of agents' beliefs, all the perfect Bayesian equilibria of a mechanism induce outcomes consistent with the social choice function (SCF). It is shown that, for a weaker notion of equilibrium and for a general class of games, the set of all such equilibria can be computed by means of a "backwards procedure" that combines the logic of rationalizability and backward induction reasoning. It is further shown that a SCF is (partially) implementable for all models of beliefs if and only if it is ex-post incentive compatible. In environments with single crossing preferences, strict ex-post incentive compatibility and a "contraction property" are sufficient to guarantee full robust implementation in direct mechanisms. This property limits the interdependence in agents' valuations.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Economics

First Advisor
George J. Mailath

Keywords
Robustness; Mechanism Design; Robust Implementation; Dynamic Games; Incomplete Information; Rationalizability; Hierarchies of beliefs

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/134
Subject Categories
Economic Theory
INCOMPLETE INFORMATION AND ROBUSTNESS IN STRATEGIC ENVIRONMENTS

Antonio Penta

A DISSERTATION

in

Economics

Presented to the Faculties of the University of Pennsylvania

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

2010

Supervisor of Dissertation

______________________________

George J. Mailath, Walter H. Annenberg Professor in the Social Sciences

Graduate Group Chairperson

______________________________

Dirk Krueger, Professor of Economics

Dissertation Committee:

George J. Mailath, Walter H. Annenberg Professor in the Social Sciences
Andrew Postlewaite, Harry P. Kamen Professor of Economics
Qingmin Liu, Assistant Professor of Economics
INCOMPLETE INFORMATION AND ROBUSTNESS
IN STRATEGIC ENvironments

COPYRIGHT

2010

Antonio Penta
Per Sara, il suo amore, e il suo sorriso.
ACKNOWLEDGMENTS

I want to thank Prof. George Mailath for his dedication, his support, the valuable insights, and above all for having constantly challenged me. As a true mentor, he has set an example of intellectual honesty and professionalism which will accompany me throughout my own career.

I also thank Prof. Qingmin Liu, for the careful and always constructive criticisms, and Prof. Andrew Postlewaite, for the sharpness of his thinking.

Special thanks go to Prof. Pierpaolo Battigalli, who more than anyone has shaped my view and understanding of game theory. This very dissertation testifies the extent of my intellectual debt to him.

I owe a lot to the intellectual vivacity of Larbi Alaoui, to the inspiring passion of Dave Cass, and to a long list of colleagues and friends who have made these years at Penn a unique life experience.

Everything I achieved though would not have been possible without the invaluable and unconditional support of my parents, the loyalty of my sister, and the friendship of Gil, Larbi, Deniz, Dionissi, Francesca, Leonardo and Michela. Of the many things I should thank them for, there is one that is above the others: their patience.

Finally, I thank Sara, without whom everything would have had just a different flavor.
ABSTRACT
INCOMPLETE INFORMATION AND ROBUSTNESS IN STRATEGIC ENVIRONMENTS
Antonio Penta
George J. Mailath

Game theoretic modeling involves making assumptions on agents’ infinite hierarchies of beliefs. These assumptions are understood to be only approximately satisfied in the actual situation. Thus, the significance of game theoretic predictions depend on robustness properties of the solution concepts adopted. Chapter 1 discusses recent results in this research area and their relations with the results obtained in the subsequent chapters. Chapter 2 explores the impact of misspecification of higher order beliefs in static environments, when arbitrary common knowledge assumptions on payoffs are relaxed. (Existing literature focuses on the extreme case in which all such assumptions are relaxed.) Chapter 3 provides a characterization of the strongest predictions, for dynamic games, that are “robust” to possible misspecifications of agents’ higher order beliefs, and shows that such characterization depends on modeling assumptions that have hitherto received little attention in the literature (namely, the distinction between knowledge and certainty), raising novel questions of robustness. Chapter 4 develops a methodology to address classical questions of implementation, when agents’ beliefs are unknown to the designer and their private information changes over time. The key idea is the identification of a solution concept that allows a tractable analysis of the full implementation problem: Full “robust” implementation requires that, for all models of agents’ beliefs, all the perfect Bayesian equilibria of a mechanism induce outcomes consistent with the social choice function (SCF). It
is shown that, for a weaker notion of equilibrium and for a general class of games, the set of all such equilibria can be computed by means of a “backwards procedure” that combines the logic of rationalizability and backward induction reasoning. It is further shown that a SCF is (partially) implementable for all models of beliefs if and only if it is ex-post incentive compatible. In environments with single crossing preferences, strict ex-post incentive compatibility and a “contraction property” are sufficient to guarantee full robust implementation in direct mechanisms. This property limits the interdependence in agents’ valuations.
## Contents

Acknowledgements iv

1 Incomplete Information and Robustness in Strategic Environments 1

1.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2

1.2 Complete Information: Rationalizability and Equilibria . . . . . . . 4

1.2.1 (Static) Games with Complete Information. . . . . . . . . . . 5

1.2.2 Common Belief in Rationality and Equilibrium . . . . . . . . 7

1.3 (Static) Games with Incomplete Information . . . . . . . . . . . . . . 9

1.3.1 Harsanyi’s approach: Bayesian Games and Equilibrium . . . . . 10

1.3.2 Non-Equilibrium Approach . . . . . . . . . . . . . . . . . . . . . 16

1.4 Interim Robustness in Static Games . . . . . . . . . . . . . . . . . . . 24

1.4.1 Relaxing all CK assumptions on payoffs . . . . . . . . . . . . . 25

1.4.2 Belief-Free Models and Equilibrium . . . . . . . . . . . . . . . 32

1.4.3 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33

1.5 Dynamic Games . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33

1.5.1 Relaxing all CK-assumptions on payoffs (II) . . . . . . . . . . . 37

1.5.2 Equilibria in Belief-free environments . . . . . . . . . . . . . . 47

2 On the Structure of Rationalizability on Arbitrary Spaces of Uncertainty 62
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>62</td>
</tr>
<tr>
<td>2.2</td>
<td>Game Theoretic Framework</td>
<td>65</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Structure of Rationalizability without <em>Richness</em></td>
<td>67</td>
</tr>
<tr>
<td>2.3</td>
<td>Discussion</td>
<td>74</td>
</tr>
<tr>
<td>3</td>
<td>Higher Order Beliefs in Dynamic Environments</td>
<td>76</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>77</td>
</tr>
<tr>
<td>3.2</td>
<td>Relaxing CK-assumptions and Robustness in Dynamic Games</td>
<td>82</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Preliminaries and Examples</td>
<td>82</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Non-technical Presentation of the Approach and Results</td>
<td>88</td>
</tr>
<tr>
<td>3.3</td>
<td>Game Theoretic Framework</td>
<td>93</td>
</tr>
<tr>
<td>3.4</td>
<td>Interim Sequential Rationalizability</td>
<td>99</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Example: Finitely Repeated Prisoner’s Dilemma</td>
<td>102</td>
</tr>
<tr>
<td>3.5</td>
<td>Robustness(-es)</td>
<td>105</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Upper Hemicontinuity</td>
<td>105</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Type Space Invariance</td>
<td>106</td>
</tr>
<tr>
<td>3.5.3</td>
<td>Model Invariance</td>
<td>106</td>
</tr>
<tr>
<td>3.6</td>
<td>The structure of <em>ISR</em> in the <em>Universal Model</em></td>
<td>107</td>
</tr>
<tr>
<td>3.6.1</td>
<td>Sensitivity of Multiplicity to higher order beliefs</td>
<td>109</td>
</tr>
<tr>
<td>3.6.2</td>
<td>Genericity of Uniqueness</td>
<td>113</td>
</tr>
<tr>
<td>3.7</td>
<td>Related Literature and Concluding Remarks</td>
<td>115</td>
</tr>
<tr>
<td>4</td>
<td>Robust Dynamic Mechanism Design</td>
<td>119</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>120</td>
</tr>
<tr>
<td>4.2</td>
<td>A Dynamic Public Goods Problem</td>
<td>123</td>
</tr>
<tr>
<td>4.3</td>
<td>Environments</td>
<td>130</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4.4</td>
<td>Mechanisms</td>
<td>132</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Games with Payoff Uncertainty</td>
<td>132</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Direct Mechanisms</td>
<td>134</td>
</tr>
<tr>
<td>4.5</td>
<td>Models of Beliefs</td>
<td>135</td>
</tr>
<tr>
<td>4.6</td>
<td>Solution Concepts</td>
<td>136</td>
</tr>
<tr>
<td>4.6.1</td>
<td>Mechanisms in Environments with Beliefs: ( (\mathcal{E}, \mathcal{M}, \mathcal{B}) )</td>
<td>137</td>
</tr>
<tr>
<td>4.6.2</td>
<td>Mechanisms in Environments without Beliefs: ( (\mathcal{E}, \mathcal{M}) )</td>
<td>140</td>
</tr>
<tr>
<td>4.6.3</td>
<td>Results</td>
<td>143</td>
</tr>
<tr>
<td>4.7</td>
<td>Partial Implementation</td>
<td>146</td>
</tr>
<tr>
<td>4.8</td>
<td>Full Implementation in Direct Mechanisms</td>
<td>148</td>
</tr>
<tr>
<td>4.8.1</td>
<td>Environments with Monotone Aggregators of Information</td>
<td>148</td>
</tr>
<tr>
<td>4.8.2</td>
<td>Aggregator-Based SCF</td>
<td>153</td>
</tr>
<tr>
<td>4.8.3</td>
<td>Relaxing “Aggregator-Based”</td>
<td>155</td>
</tr>
<tr>
<td>4.8.4</td>
<td>“Quasi-direct” Mechanisms</td>
<td>159</td>
</tr>
<tr>
<td>4.9</td>
<td>Further Remarks on the Solution Concepts</td>
<td>160</td>
</tr>
<tr>
<td>4.10</td>
<td>Concluding Remarks</td>
<td>162</td>
</tr>
<tr>
<td>A</td>
<td>Appendix to Chapter 3</td>
<td>165</td>
</tr>
<tr>
<td>A.1</td>
<td>Proofs of results from Section 3.5</td>
<td>165</td>
</tr>
<tr>
<td>A.1.1</td>
<td>Proof of proposition 3.1</td>
<td>165</td>
</tr>
<tr>
<td>A.1.2</td>
<td>Proof of proposition 3.2</td>
<td>166</td>
</tr>
<tr>
<td>A.1.3</td>
<td>Proof of proposition 3.3</td>
<td>167</td>
</tr>
<tr>
<td>A.2</td>
<td>Proofs of results from Section 3.6</td>
<td>168</td>
</tr>
<tr>
<td>A.2.1</td>
<td>Proof of lemma 3.2</td>
<td>168</td>
</tr>
</tbody>
</table>
List of Figures

1.1 Example 1.5 ................................................................. 19
1.2 Examples 1.7 and 1.8 ......................................................... 34
1.3 The type space in Example 1.8 ........................................... 40
1.4 Examples 1.9 and 1.10 ....................................................... 41
1.5 Example 1.11 ................................................................. 50
1.6 Example 1.12 ............................................................... 56
1.7 Example 1.13 ............................................................... 59

3.1 Examples 3.1 and 3.2 ......................................................... 82
3.2 The type space in Example 3.2 ........................................... 85
3.3 Examples 3.3 and 3.4 ....................................................... 87

4.1 Single Crossing Conditions .............................................. 157
Chapter 1

Incomplete Information and Robustness in Strategic Environments

Abstract: Standard models for games with incomplete information impose common knowledge on agents’ payoffs and beliefs. Recent research has explored questions of robustness of game theoretic predictions with respect to possible misspecifications of the model. In static games, the solution concept of rationalizability, corresponding to the epistemic assumptions of Rationality and Common Belief in Rationality (RCBR), provides a unique answer to several questions of robustness. When the same questions are addressed in dynamic games, such unity is lost: different ‘dynamic extensions’ of RCBR provide answers to different robustness questions.

Keywords: equilibrium – hierarchies of beliefs – incomplete information – rationalizability – robustness

JEL Codes: C72; C73; D82.
1.1 Introduction

Game theory can be defined as the study of mathematical models of interactions of rational, strategic and intelligent agents, in situations in which agents’ payoffs depend on the actions of others.

Rationality here means that agents are assumed to choose behavior that maximizes exogenously given, stable preferences. (More specifically, agents are assumed expected utility maximizers.) The interdependence of agents’ decision problems is not exclusive to game theory: Models of competitive economies do exhibit this feature. What is specific to game theory is the fact that agents are aware of such interdependence: That is, agents are strategic. Intelligent agents not only take strategic considerations into account, but also the fact that their opponents are strategic.

Classical game theory is based upon two fundamental methodological tenets: first, the assumption that the rules of the game and the payoff structure are commonly known by the players (that is, everybody knows them, everybody knows that everybody knows them, and so on); second, the equilibrium hypothesis, for which agents play best-responses to correct conjectures about the opponents’ behavior (i.e. a Bayes-Nash equilibrium).

For a long time, authoritative views ascribed to Nash equilibrium the value of a necessary consequence of the assumption of common knowledge of the game and of the agents’ rationality, to the point that the equilibrium hypothesis almost acquired the status of a methodological assumption.\footnote{For example, in Schelling (1960, p.115):}

\begin{quote}
“Whether one agrees explicitly to a bargain or agrees tacitly or accepts it by default, he must, if he has his wits with him, expect that he could do no better and recognize that the other party must reciprocate the feeling. Thus the fact of an outcome, which is simply a coordinated choice, should be analytically characterized by the notion of
\end{quote}

1 This view was forcefully challenged by
Bernheim (1984) and Pearce (1984), who introduced the concept of rationalizability:

“far from being a consequence of rationality, equilibrium arises from certain restrictions on agents’ expectations which may or may not be plausible [...]” (Bernheim 1984, p. 1007).

In contrast, rationalizability is based on the idea that agents are rational, they believe that everybody is rational, and so on (that is, rationality and common belief in rationality, RCBR hereafter), without imposing coordination of expectations.\(^2\)

Brandenburger and Dekel (1987) showed that RCBR and equilibrium are indeed connected, but in a more subtle way. Brandenburger and Dekel’s result occupies a central position in relation with more recent research, that explored the premises of traditional solution concepts when the other fundamental tenet of game theory is addressed: Namely, that the rules of the game, including the payoff structure and agents’ beliefs, are common knowledge. This literature produced a number of notions of rationalizability (all based on the idea of RCBR) for incomplete information games, i.e. strategic situations in which some features of the environment are not commonly known to players.

Overall, RCBR (in its different declinations) emerged as the unique answer to different questions of robustness for static games. In my dissertation I address some of these robustness questions in the context of dynamic games. What emerges is that this unity is lost: different analogues of RCBR for dynamic games answer different converging expectations.”

More recently, in Myerson (1991, p. 108):

“[...] being a Nash Equilibrium is [...] a necessary condition for a theory to be a good prediction of the behavior of intelligent rational players.”

\(^2\)Tan and Werlang (1988) provided the formal epistemic characterization.
questions of robustness. The source of the “discrepancy” that arises in dynamic games lies in the agents’ beliefs updating process, and in the possibility that they observe unexpected moves, possibly inconsistent with their previous beliefs about everyone’s rationality. This possibility, inherent to dynamic games, also raises novel robustness questions that can only be addressed adopting an extensive-form approach.

This chapter surveys the main concepts and results in this literature, serving as a background for the results obtained in the subsequent chapters. The rest of the paper is organized as follows: Section 1.2 introduces the basic concepts, and presents the main result from Brandenburger and Dekel (1987); Section 1.3 introduces games with incomplete information; Section 1.4 presents the main results on robustness for static games; Section 1.5 concludes discussing the main results of the dissertation concerning the analysis of robustness problems in dynamic games.

1.2 Complete Information: Rationalizability and Equilibria

Bernheim and Pearce clarified that equilibrium is not the logical consequence of the assumptions of rationality and common belief in rationality (RCBR), and proposed the concept of rationalizability. However, Brandenburger and Dekel (1987) showed that rationalizability and the equilibrium approach are more closely related than at first it might appear, proving that a certain kind of equilibrium could be given a decision theoretic foundation based on RCBR alone.

Brandenburger and Dekel’s result applies to games with complete information, but is important for understanding the subsequent research on robustness in games of incomplete information. This section introduces the basic notation and illustrates the main concepts and results from Brandenburger and Dekel (1987, BD hereafter).
1.2.1 (Static) Games with Complete Information.

For simplicity, we will consider here only games with two players, indexed by \( i = 1, 2 \). A (static) game with complete information is a tuple \( (A_i, u_i)_{i=1,2} \) where \( A_i \) is a (finite) set of actions (or strategies) for agent \( i \), and \( u_i : A_1 \times A_2 \to \mathbb{R} \) is agent \( i \)'s payoff function. The tuple \( (A_i, u_i)_{i=1,2} \) is assumed to be common knowledge among the players: that is, everybody knows \( (A_i, u_i)_{i \in N} \), everybody knows that everybody knows \( (A_i, u_i)_{i=1,2} \), and so on.

**Rationalizability.** A subset \( C_1 \times C_2 \) of \( A_1 \times A_2 \) is a best reply set if for every \( i \) and each \( a_i \in C_i \), there exist conjectures \( \mu^i \in \Delta (C_{-i}) \) to which \( a_i \) is a best reply. The set of rationalizable action profiles \( R_1 \times R_2 \subseteq A_1 \times A_2 \) is the (finite) union of all best reply sets \( C_1^\alpha \times C_2^\alpha \). An equivalent definition of the set \( R_1 \times R_2 \) can be given in terms of iterated deletion of strongly dominated actions.\(^3\)

**Example 1.1.** Consider for instance the following game, with \( A_1 = \{U, M, D\} \), \( A_2 = \{L, C, R\} \), and payoffs as represented in the following matrix (payoffs in bold correspond best responses):

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>C</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>2, 3</td>
<td>5, 1</td>
<td>2, 0</td>
</tr>
<tr>
<td>M</td>
<td>3, 1</td>
<td>2, 3</td>
<td>2, 0</td>
</tr>
<tr>
<td>D</td>
<td>1, 4</td>
<td>1, 3</td>
<td>6, 0</td>
</tr>
</tbody>
</table>

Notice that strategy \( R \) is dominated by strategies \( L \) and \( C \), hence a rational player 2 would never play \( R \). Hence, if player 1 believes that 2 is rational, he would never play \( R \).

\(^3\)In the case of more than two players, this definition translates into correlated rationalizability, in which agents’ conjectures about the opponents’ behavior allow for correlation, i.e. \( \mu^i \in \Delta (\times_{j \neq i} A_j) \). Bernheim and Pearce instead defined independent rationalizability, in which opponents’ actions were independent in the eyes of a player, hence conjectures are of the form \( \mu^i \in \times_{j \neq i} \Delta (A_j) \). For the case of two players there is no distinction.
play D, which is only justified by the belief that 2 plays R. The set of rationalizable strategy profiles is therefore \(\{U, M\} \times \{L, C\}\).

**A Posteriori Equilibrium.** The equilibrium concept that Brandenburger and Dekel relate to rationalizability is a refinement of Aumann’s (1974) subjective correlated equilibrium, called a posteriori equilibrium.

To define an a posteriori equilibrium, we must introduce a (finite) space \(\Omega\), and for each player \(i\) a partition \(\mathcal{I}_i\) of \(\Omega\). A strategy now is a \(\mathcal{I}_i\)-measurable function \(\sigma_i : \Omega \rightarrow A_i\). For each player \(i\), also specify a prior \(P_i\) on \(\Omega\) and a proper and regular version of conditional probability on \(\mathcal{I}_i\): for each \(I_i \in \mathcal{I}_i\), \(P_i(\cdot | I_i)\) is required to be a probability measure on \(\Omega\) and \(P_i(I_i | I_i) = 1\). Clearly, if \(P_i(I_i) > 0\), Bayes rule guarantees that \(P_i(\cdot | I_i)\) satisfies both requirements, but \(P_i(\cdot | I_i)\) is required to satisfy them also if \(P_i(I_i) = 0\). For each \(i\) and \(\omega \in \Omega\), let \(\mathcal{I}_i(\omega)\) denote the cell of \(i\)’s partition that contains \(\omega\).

**Definition 1.1.** A strategy profile \((\sigma_1, \sigma_2)\) is an a posteriori equilibrium if for each \(i\) and for each \(\omega \in \Omega\),

\[
\sum_{\omega' \in \Omega} P_i(\{\omega\} | \mathcal{I}_i(\omega)) \cdot u_i(\sigma_i(\omega), \sigma_{-i}(\omega')) \geq \sum_{\omega' \in \Omega} P_i(\{\omega\} | \mathcal{I}_i(\omega)) \cdot u_i(a'_i, \sigma_{-i}(\omega')) \text{ for all } a'_i \in A_i.
\]

Notice that, similar to Aumann’s (1974) subjective correlated equilibrium, in an a posteriori equilibrium agents have subjective beliefs over the space \(\Omega\), that is agents are allowed to have inconsistent priors over \(\Omega\). A posteriori equilibrium though refines subjective correlated equilibrium because it requires every agent to best respond at every \(I_i \in \mathcal{I}_i\), including those cells that receive zero probability under the prior. Subjective correlated equilibrium instead maintains an ex ante perspective, and con-
ditional probabilities on zero probability events don’t need to be specified.

1.2.2 Common Belief in Rationality and Equilibrium

The next result states an equivalence between rationalizable and equilibrium strategies.\(^4\)

**Proposition 1.1** (cf. Proposition 2.1 in BD). *Strategy* \(a_i\) *is rationalizable if and only if there exists an* a posteriori equilibrium *in which* \(a_i\) *is played.*\(^5\)

The following example illustrates the proposition.

**Example 1.2.** *Take the game in example 1.1, and consider the following equilibrium, in which all the rationalizable strategies are played: let* \(\Omega = \{a, b, c, d\}\), *and agents’ partitions* \(\mathcal{I}_1 = \{(a, b), (c, d)\}\) *and* \(\mathcal{I}_2 = \{(a, c), (b, d)\}\). *Let agents’ priors be such that* \(P_1(\{a\}) = 1\) *and* \(P_2(\{c\}) = 1\). *Let conditional probabilities be such that* \(P_1(\{a\} \mid \{a, b\}) = 1\) *and* \(P_1(\{d\} \mid \{c, d\}) = 1\), *while for player 2 they are* \(P_2(\{c\} \mid \{a, c\}) = 1\) *and* \(P_2(\{b\} \mid \{b, d\}) = 1\). *Consider the following strategy profile:

\[
\begin{align*}
\sigma_1^*(a) &= \sigma_1^*(b) = U, \quad \text{and} \quad \sigma_1^*(c) = \sigma_1^*(d) = M; \\
\sigma_2^*(a) &= \sigma_2^*(c) = C, \quad \text{and} \quad \sigma_2^*(b) = \sigma_2^*(d) = L.
\end{align*}
\]

*To see that* \((\sigma_1^*, \sigma_2^*)\) *is an a posteriori equilibrium, consider state a first: at that state, player 1 puts probability one on the true state, and given* \(\sigma_2^*\) *he also puts probability one on 2 playing* \(C\): \(\sigma_1^*(a) = U\) *is thus a best response. Player 2 instead puts probability 1 on* \(c\) *at state a, hence he expects player 1 to play* \(M\), *hence* \(\sigma_2^*(a) = C\) *is a best response.*

\(^4\)BD expressed the equivalence in terms of payoffs. For our purposes, it is more useful to present it in terms of strategies.

\(^5\)With more than two players, this result holds for *correlated rationalizability*. Using a modified version of the equilibrium concept, BD also provide an analogous for *independent rationalizability* (footnote 3.)
response. At state $c$, the opposite is true: 2 puts probability one on the true state, and expects 1 to play $M$, and he best responds playing $C$; player 1 instead (mistakenly) puts probability one on $a$, expects 2 to be playing $C$, and best responds playing $U$. In state $b$, player 1 assigns probability one to $a$, hence he expects 2 to play $C$, hence $\sigma_1^*(b) = U$ is a best response; at $b$ instead 2 puts probability one on $b$, i.e. he expects 1 to play $U$, hence $\sigma_2^*(b) = L$ is a best response. Finally, at $d$, player 1 believes the true state is $d$, hence expects 2 to play $L$, so that $\sigma_1^*(d) = M$ is a best response; at $d$ instead player 2 believes that the state is $b$, i.e. he expects 1 to play $U$, so that $\sigma_2^*(d) = L$ is a best response.

The equilibrium hypothesis is based on two assumptions: 1) rationality; 2) coordination of expectations: agents hold correct conjectures about the opponents’ strategies. On the contrary, rationalizability does not impose coordination and derives restrictions on agents’ conjectures from the hypothesis of common belief in rationality. These restrictions may or may not lead to coordination of expectations. Hence, in general, rationalizability is weaker than equilibrium concepts. The equivalence result of Proposition 1.1 though shows that the two approaches are intimately connected: Proposition 1.1 can be interpreted as saying that if no restrictions on the agents’ beliefs about the correlating device are imposed (i.e. if agents play not Nash, but a posteriori equilibrium), and if the analyst has no information about the equilibrium correlating device (i.e. he does not know $\langle \Omega, \mathcal{I}_1, \mathcal{I}_2 \rangle$), then the assumption that agents best respond to correct conjectures about the opponents’ strategies (i.e. the equilibrium hypothesis) has no bite beyond rationalizability alone.
1.3 (Static) Games with Incomplete Information

Games of incomplete information are situations in which some features of agents’ payoffs are not common knowledge.\(^6\) This situation can be modelled parametrizing the payoff functions on a space of uncertainty \(\Theta\), letting \(u_i : A \times \Theta \rightarrow \mathbb{R}\) denote agent \(i\)’s payoff as a function of the action profile \(a \in A\) and the payoff state \(\theta \in \Theta\). In general, agents may have private information about the payoff-relevant states \(\Theta^*\). To represent this, write \(\Theta\) as

\[
\Theta = \Theta_0 \times \Theta_1 \times \Theta_2.
\]

For each \(i \in \{1, 2\}\), \(\Theta_i\) is the set of player \(i\)’s payoff types, i.e. possible pieces of information that \(i\) may have about the payoff state; \(\Theta_0\) instead represents the set of states of nature, the residual uncertainty that is left after pooling all players’ information. The interpretation is that when the payoff state is \(\theta = (\theta_0, \theta_1, \theta_2)\), player \(i\) knows that \(\theta \in \Theta_0 \times \{\theta_i\} \times \Theta_{-i}\). Special cases of interest are those of distributed knowledge, in which \(u_i\) is constant in \(\theta_0\) (or \(\Theta_0\) is a singleton); private values, in which each \(u_i\) depends on \(\Theta_i\) only; and the case in which the \(u_i\)’s depend on \(\Theta_0\) only, so that players have no information about payoffs (or, without loss of generality, \(\Theta = \Theta_0\)).

We refer to the tuple \(G = \left\langle \Theta_i, (A_i, u_i)_{i=1,2} \right\rangle\) as game with payoff uncertainty. (If \(\Theta = \{\theta^*\}\), the game has complete information, and it’s denoted by \(G(\theta^*).\) Notice that this is not the standard way of representing a game with incomplete information, i.e. \(G\) is not a Bayesian game. Bayesian games are introduced next.

\(^6\)Harsanyi (1967-68) argued that, in static games, lack of common knowledge of any feature of the game (i.e. payoffs, rules, etc.) can be reduced to this case.
1.3.1 Harsanyi’s approach: Bayesian Games and Equilibrium

Extending the classical (equilibrium) approach to situations with incomplete information entails specifying agents’ infinite hierarchies of beliefs: consider player 1, who knows $\theta_1$ and has beliefs about what he doesn’t know, $(\theta_0, \theta_2)$. In any equilibrium, 1’s behavior will in general depend on his information and on such first order beliefs. But this is true for agent 2 as well, so to form beliefs about 2’s behavior player 1 also needs to have beliefs about 2’s beliefs about $(\theta_0, \theta_1)$ (that is, beliefs about 2’s first order beliefs, or second order beliefs). Clearly, this reasoning can be iterated ad infinitum. This hierarchies-of-beliefs approach is mathematically feasible (see e.g. Mertens and Zamir, 1985), but it does not provide a tractable framework for a direct analysis of incomplete information games.

Harsanyi (1967-68) showed that a convenient parametrization of agents’ beliefs avoids the difficulties of the hierarchies-approach, and so incomplete information games can be analyzed with the standard tools of game theory. Key to Harsanyi’s approach is the notion of “type space”: For each player $i$ and each payoff-type $\theta_i \in \Theta_i$, we add a parameter $e_i$ from some space $E_i$ that encodes the purely epistemic components of player $i$’s attributes. In general, different values of $e_i$ can be attached to a given payoff-type $\theta_i$. This way we obtain a set $T_i \subseteq \Theta_i \times E_i$ of possible attributes, or Harsanyi-types, of player $i$. A Harsanyi-type $t_i = (\theta_i, e_i)$ encodes the payoff-type and the epistemic type of a player. The epistemic components are given content by functions that specify agents’ beliefs about what they don’t know: The state of nature $\theta_0$ and the opponent’s Harsanyi-type $t_{-i} = (\theta_{-i}, e_{-i})$. Such functions are denoted by $\tau_i : T_i \rightarrow \Delta (\Theta_0 \times T_{-i})$, $i = 1, 2$. The tuple $T = \left\{(T_i, \tau_i)_{i=1,2}\right\}$ is a type space, and is assumed common knowledge. More formally, and to emphasize the underlying space of uncertainty, we define:
**Definition 1.2.** Given an underlying space of uncertainty $\Theta = \Theta_0 \times \Theta_1 \times \Theta_2$, and parameter spaces $E_1, E_2$, a $\Theta$-based type space is a tuple $T = \langle (T_i, \tau_i)_{i=1,2} \rangle$ such that $T_i \subseteq \Theta_i \times E_i$ satisfies $\text{proj}_{\Theta_i} T_i = \Theta_i$ and $\tau_i : T_i \rightarrow \Delta (\Theta_0 \times T_{-i})$.

First, notice that each type in a type space induces an infinite hierarchy of beliefs: the first-order beliefs $\pi^1_i (t_i)$ of type $t_i$ is simply the marginal of $\tau_i (t_i)$ on $\Theta_0 \times \Theta_{-i}$; the $(k+1)$-order beliefs implicit in $t_i$ is derived from $\tau_i (t_i)$ and knowledge of the function $\pi^k_i (\cdot)$ mapping the opponent's type into his $k$-order beliefs. Notice also that the array of probabilities $(\pi_i (t_i))_{t_i \in T_i}$ can always be derived from some "prior", i.e. there is at least one probability measure $P_i \in \Delta (\Theta_0 \times T_1 \times T_2)$ such that $\tau_i (t_i) = P_i (\cdot | t_i)$ for all $t_i$. Thus, within this framework, players' situation in the incomplete information game is formally similar to the interim stage of a game with complete but imperfect and asymmetric information about a chance move, whereby $t_i$ represents agent $i$'s private information. But the "prior" $P_i$ here does not represent $i$'s beliefs in a hypothetical ex-ante stage, it is only a (unnecessary) technical device to express the belief function $\tau_i (\cdot)$.

When we attach a type space $T$ to the game $G$, we obtain a Bayesian game: $B = \langle \Theta, (A_i, T_i, \tau_i, u^*_i)_{i=1,2} \rangle$, where $(T_i, \tau_i)_{i=1,2}$ are as in $T$, and $u^*_i$ extends $u_i$ to the payoff-irrelevant epistemic components (i.e. $u^*_i : \Theta_0 \times T \times A \rightarrow \mathbb{R}$ is such that $u^*_i (\theta_0, t_1, t_2, a) = u_i (\theta_0, \theta_1, \theta_2, a)$ whenever $t_j = (\theta_j, e_j)$, $j = 1, 2$). In the rest of the paper, notation will be slightly abused and $u_i$ will be used for both domains.

**Definition 1.3.** A Bayesian equilibrium is a profile of strategies $\sigma_i : T_i \rightarrow \Delta (A_i)$, $i = 1, 2$ such that for each player $i$ and each Harsanyi-type $t_i$, strategy $\sigma_i (t_i)$ maximizes $i$'s expected payoff given the payoff-type $\theta_i$, the subjective belief $\tau_i (t_i) \in \Delta (\Theta_0 \times T_{-i})$, and the function $\sigma_{-i} : T_{-i} \rightarrow \Delta (A_{-i})$.

Throughout the paper, the convention is maintained that "beliefs" are about $\Theta$.
and the opponents’ beliefs about $\Theta$. That is, “beliefs” are about \textit{exogenous variables} only. The term “conjectures” instead refers to beliefs that also encompass the opponents’ strategies.

\textbf{The universal type space.} Harsanyi (1967-68) argued that assuming the existence of a commonly known Bayesian model of the environment (namely, a type space) entails no loss of generality: any set of hierarchies of beliefs can find an implicit representation as a type in a type space. Mertens and Zamir (1985) showed that this is indeed the case, proving the existence of a \textit{universal type space} that essentially contains all possible hierarchies of beliefs on a fundamental space of uncertainty $\Theta$. The $\Theta$-based universal type space, $T^*_\Theta$, can be thought of as the set of all possible hierarchies of beliefs over $\Theta$: Let $T^*_i,\Theta$ denote the set all hierarchies of beliefs $t_i = (\theta_i, \pi^1_i, \pi^2_i,...)$ that player $i$ may have that satisfy some consistency conditions.\footnote{A hierarchy satisfies “coherency” if higher order beliefs agree with lower order beliefs on their common support. Hierarchies in $T^*_i,\Theta$ are those consistent with \textit{common certainty of coherency}: i.e. they are coherent, put probability one on the opponent’s beliefs that are coherent and concentrated on coherent beliefs, and so on. (See Mertens and Zamir, 1985, or Brandenburger and Dekel, 1993).} Mertens and Zamir showed that if $T^*_i,\Theta$ are endowed with the product topology, then there exists a homeomorphism $\tau^*_i : T^*_i,\Theta \rightarrow \Delta (\Theta_0 \times T^*_i,\Theta)$. Thus, hierarchies of beliefs can be represented by a single joint probability measure over the space $\Theta_0$ and the opponents’ hierarchies. That is, sets of hierarchies $T^*_1,\Theta, T^*_2,\Theta$ can be thought of as sets of types in a $\Theta$-based type space. The type space thus obtained, $T^*_\Theta = \langle (T^*_{i,\Theta},\tau^*_i)_{i=1,2} \rangle$, is the $\Theta$-based universal type space.

Hence, Harsanyi’s indirect approach is without loss of generality in the sense that, modeling a situation with incomplete information $G$ as the Bayesian game $B^* = \langle G, T^*_\Theta \rangle$ imposes no more restrictions than those that are already implicit in $G$. In applied work though it is common to work with smaller, non-universal type spaces,
which do entail some loss in generality. Mertens and Zamir showed that any such
(non-universal) type space without redundant types (i.e. such that no two distinct
types induce the same hierarchy of beliefs) can be seen as a belief-closed subspace of
the universal type space.

Ex-ante vs interim perspective. Harsanyi’s implicit representation of hierar-
chies is convenient because it ultimately transforms the problem of incomplete infor-
mation into one of asymmetric information about an initial chance move.\footnote{Harsanyi
pushed the analogy even further by assuming that all the subjective beliefs
\((\tau_i (t_i))_{t_i \in T_i} \) can be derived from a common prior \( P \in \Delta (\Theta_0 \times T) \),
so that \( \tau_i (t_i) = P (\cdot | t_i) \). In this case, Bayesian equilibrium simply corre-
ponds to a Nash equilibrium of a companion game with imperfect information about a
fictitious chance move selecting the vector of attributes according to probability
measure \( P \). For games with “genuine” incomplete information, such common prior
assumption can hardly be justified. For more on this see, for example, Gul (1998) and
assumption in incomplete information games, characterizing it as a very strong ‘agreement’
property.} But the
formal analogy should not overshadow the fundamental distinction between the two
situations: If agents are facing “genuine” incomplete information, the ex-ante stage
is merely fictitious; the relevant objects are the hierarchies of beliefs, and therefore
the natural perspective is the interim one. Also, hierarchies of beliefs (or types) are
purely subjective states describing a player’s view of the strategic situation he is fac-
ing. As such, they enter the analysis as a datum and should be regarded in isolation
(i.e. player by player and type by type). Nothing prevents players’ views of the world
to be inconsistent with each other; they are part of the environment (exogenous vari-
ables), and as such game theoretic reasoning cannot impose restrictions on them; it
is given such beliefs that we can apply game theoretic reasoning to make predictions
about players’ behavior (the endogenous variables).

Also, it is important to emphasize that in situations with “genuine” incomplete
information Harsanyi-types are mere parametrizations of agents’ beliefs, and cannot
in general be interpreted as information: When his type is \( t_i = (\theta_i, e_i) \), agent \( i \)'s information about the environment is \( \theta_i \), not \( t_i \); the epistemic component is just a parameter for \( i \)'s subjective beliefs about what \( i \) does not know, i.e. the state of nature, the opponents’ information and their beliefs. It is only in some special cases that Harsanyi-types can be interpreted as “information” (see Section 1.3.2).

Both the applied and the theoretical literature have often overlooked the distinction between the \textit{ex-ante} and the \textit{interim} perspectives, as well as the related distinction between “types” viewed as “information” or as “parameters”. The distinction though has great importance when questions of robustness are addressed: In situations of \textit{asymmetric information}, where the ex-ante stage is real, a meaningful robustness question would be to consider perturbations of agents’ priors \( P_i \in \Delta (\Theta_0 \times T_1 \times T_2) \), and study how solution concepts behave in such “nearby models” (e.g., Kaji and Morris, 1997); with “genuine” \textit{incomplete information} instead, models should be considered “nearby” if the agents’ hierarchies of beliefs are “close”, not their priors.

\textbf{Redundant Types, Bayesian Equilibria and Correlated Equilibria}

\textit{Redundant types} in a Bayesian game are types that correspond to the same information and to the same hierarchy of beliefs over \( \Theta \).

\textbf{Example 1.3.} Consider a situation in which \( \Theta_0 = \{0, 1\} \), while \( \Theta_1 \) and \( \Theta_2 \) are singletons (hence, an environment with no information). Consider a type space in which \( T'_1 = \{t'_1, t''_1\} \), \( T'_2 = \{t'_2, t''_2\} \), and agents share a common prior over \( \Theta_0 \times T'_1 \times T'_2 \), described by the following matrices:

\begin{tabular}{|c|c|c|}
\hline
\( \theta_0 = 0 \) & \( t'_2 \) & \( t''_2 \) \\
\hline
\( t'_1 \) & 1/6 & 1/12 \\
\hline
\( t''_1 \) & 1/12 & 1/6 \\
\hline
\end{tabular}

\begin{tabular}{|c|c|c|}
\hline
\( \theta = 1 \) & \( t'_2 \) & \( t''_2 \) \\
\hline
\( t'_1 \) & 1/12 & 1/6 \\
\hline
\( t''_1 \) & 1/6 & 1/12 \\
\hline
\end{tabular}
Notice that in this example, both types for every player correspond to the same hierarchy of beliefs: every type puts $1/2$ probability on each $\theta \in \{0, 1\}$, and this is common certainty (i.e. all types put probability one on the opponent putting probability one on... on $\theta = 0, 1$ being equally likely).

Consider now a situation with complete information, i.e. such that $\Theta = \{\theta^*\}$. It is still possible to define type spaces in such an environment. Clearly, all such type spaces would have redundant types, all corresponding to common certainty (common knowledge, in fact) of $\theta^*$, but they would induce different Bayesian games, with possibly different sets of Bayesian equilibria. What is the relationship between the Bayesian equilibria of these games, and the equilibria of the underlying complete information game $G(\theta^*)$, without redundant types? Essentially the answer is that the correlated equilibria of $G(\theta^*)$ coincide with the Bayesian equilibria of the Bayesian games thus obtained. The point is illustrated by the following example:

**Example 1.4.** Consider the correlated equilibrium discussed in example 1.2: it is easy to see that it only takes a little more than a relabeling of the correlating device, to obtain a type space for which the correlated equilibrium is a Bayesian equilibrium of the induced Bayesian game. Let $T_1 = \{t_{11}^{ab}, t_{11}^{cd}\}$ and $T_2 = \{t_{22}^{ac}, t_{22}^{bd}\}$, and let beliefs be such that $\tau_1(t_1)[t_{12}^{ac}] = 1$ if $t_1 = t_{11}^{ab}$, 0 otherwise; for player 2 instead, let $\tau_2(t_2)[t_{21}^{cd}] = 1$ if $t_2 = t_{21}^{ac}$, 0 otherwise. It is easy to check that the following strategy profile is a Bayesian equilibrium of the induced Bayesian game: $\sigma_1(t_{11}^{ab}) = U$, $\sigma_1(t_{11}^{cd}) = M$, $\sigma_2(t_{22}^{ac}) = C$, $\sigma_2(t_{22}^{bd}) = L$. (The superscripts of these types refer to the corresponding cells in the partition of the correlating device in example 1.2.)

In fact, the result of Proposition 1.1 can be restated as follows:
Proposition 1.2. The set of rationalizable actions of a complete information game $G(\theta^*)$ characterizes the set of actions played as part of some Bayesian equilibrium of the Bayesian game obtained attaching some type space to $G(\theta^*)$.

(Proposition 1.2 is a special case of Proposition 1.9 below.)

1.3.2 Non-Equilibrium Approach

If we envision a game as an isolated interaction, there is no reason to believe that agents will be able to coordinate on a given equilibrium, i.e. that they will have correct conjectures about the opponent’s strategy. This is the essence of Bernheim (1984) and Pearce’s (1984) critique. If, on the other hand, we envision a game $G$ as the description of a stable, repeated interaction, then Nash equilibrium could be interpreted as the steady state of a learning process in which players repeatedly play $G$.\(^9\) This argument is problematic for situations with incomplete information, making the non-equilibrium approach even more compelling.\(^{10}\)

This section introduces some versions of rationalizability for incomplete information games, and discusses the relationship between them. The next section discusses some results and robustness properties of these solution concepts.

Rationalizability in Bayesian games

A non-equilibrium approach does not require the machinery developed by Harsanyi, but the notions of type space and Bayesian game are so entrenched in the literature

\(^9\)In general, convergence is not guaranteed and, even if the play converges, the limit outcome is a selfconfirming (or conjectural) equilibrium, which need not be equivalent to a Nash equilibrium. See Fudenberg and Levine (1998) and references therein.

\(^{10}\)Dekel, Fudenberg and Levine (2004) argue that the Bayesian equilibrium concept is hard to justify in terms of learning even for games with asymmetric information where the ex ante stage is real, but players have subjective heterogeneous priors on the state of Nature.
that in recent years several versions of rationalizability for Bayesian games have been put forward. Given our focus on the \textit{interim perspective}, we focus on two versions of rationalizability, both defined for the interim normal form of the game: these are \textit{interim correlated rationalizability} (Dekel, Fudenberg and Morris, 2007) and \textit{interim independent rationalizability} (Ely and Peski, 2006)

\textbf{Interim Correlated Rationalizability (ICR).} \textit{ICR} is (correlated) rationalizability applied to the interim normal form of the Bayesian game $B = \langle G, T \rangle$. Denote $i$’s “conjectures” in the Bayesian game by $\psi^i \in \Delta (\Theta_0 \times A_{-i} \times T_{-i})$. For each type $t_i$, his consistent conjectures are

$$\Psi_i(t_i) = \{ \psi^i \in \Delta (\Theta_0 \times A_{-i} \times T_{-i}) : \text{marg}_{\Theta \times T_{-i}} \psi^i = \tau_i(t_i) \} .$$

Let $BR_i(\psi^i)$ denote the set of best responses to conjecture $\psi^i$. If $a_i \in BR_i(\psi^i)$, we say that $\psi^i$ justifies action $a_i$.

**Definition 1.4.** For each $t_i \in T_i$ and $i \in N$, set $\text{ICR}^0_i(t_i) = A_i$. For $k = 0, 1, \ldots$, let $\text{ICR}^k_i(t_i)$ be such that $(a_i, t_i) \in \text{ICR}^k_i(t_i)$ if and only if $a_i \in \text{ICR}^k_i(t_i)$ and $\text{ICR}^k_{-i} = \times_{j \neq i} \text{ICR}^k_j$. Then recursively, for $k = 1, 2, \ldots$

$$\text{ICR}^k_i(t_i) = \{ a_i \in A_i : \exists \psi^{a_i} \in \Psi_i(t_i) \text{ s.t.: } a_i \in BR_i(\psi^{a_i}) \text{ and } \text{supp} (\text{marg}_{A_{-i} \times T_{-i}} \psi^{a_i}) \subseteq \text{ICR}^{k-1}_{-i} \} .$$

Then, let $\text{ICR}_i(t_i) = \bigcap_{k \geq 0} \text{ICR}^k_i(t_i)$.

\textbf{Interim Independent Rationalizability (IIR).} $\Psi_i(t_i)$ allows for correlation between the opponent’s action, his type, and the state of nature $\theta_0$. \textit{IIR} imposes the extra requirement that at each round, for an action to survive, it must be justified by a consistent conjecture that treats $a_{-i}$ as independent of $\theta_0$, conditional on $t_{-i}$. For
each \( t_i \), define the set of consistent and conditionally independent conjectures for type \( t_i \) as
\[
\Psi_i^{CI} (t_i) = \{ \psi^i \in \Delta (\Theta_0 \times A_{-i} \times T_i) : \exists \sigma_{-i} : T_{-i} \rightarrow \Delta (A_{-i}) \text{ s.t.} \psi^i [\theta_0, a_{-i}, t_{-i}] = \tau_i (t_i) [t_{-i}] \cdot \sigma_{-i} (t_{-i}) [a_{-i}] \}.^{11}
\]

**Definition 1.5.** For each \( t_i \in T_i \) and \( i \in N \), set \( IIR_0^i (t_i) = A_i \). For \( k = 0, 1, ..., \) let \( IIR_k^i \) be such that \( (a_i, t_i) \in IIR_k^i \) if and only if \( a_i \in IIR_k^i (t_i) \) and \( IIR_{k-1}^i = \times_{j \neq i} IIR_j^k \).

Then recursively, for \( k = 1, 2, ... \)
\[
IIR_k^i (t_i) = \{ a_i \in A_i : \exists \psi^{a_i} \in \Psi_i^{CI} (t_i) \text{ s.t.: } a_i \in BR_i (\psi^{a_i}) \text{ and } \text{supp} (\text{marg}_{A_{-i} \times T_i} \psi^{a_i}) \subseteq IIR_{k-1}^i \}.
\]

Then, let \( IIR_i (t_i) = \bigcap_{k \geq 0} IIR_k^i (t_i) \).

**Interim rationalizability and redundant types.**

**Example 1.5.** Consider a game with payoff uncertainty with payoffs represented as in figure 1.1.\(^{12}\) Suppose that players have no information about the states \( \Theta = \{0, 1\} \), that they both put \( \frac{1}{2} \) probability on each state and that this is common knowledge.

In figure 1.1 two type spaces are used to model this situation: In the bottom type space \( T^* \), each player has only one type, \( t^*_i \), which puts probability \( \frac{1}{2} \) on each pair \((0, t^*_{-i})\) and \((1, t^*_{-i})\). In the top type space \( T' \), each player has two types, \( t'_i \) and \( t''_i \).

The two matrices represent the common prior on \( \Theta \times T' \). Notice that each type in \( T'_i \) corresponds to the same \( \Theta \)-hierarchy as \( t^* \), i.e. represents the beliefs that the two states are equally likely and that this is common knowledge. Clearly, either action is rational for all types of player 2, as she is indifferent between both actions. Now, consider

\(^{11}\)In case of more than two players, ICR also allows correlation among all of the opponents’ behavior and information, while IIR imposes conditional independence also among the opponents’ strategies.

\(^{12}\)This example is borrowed from Dekel, Fudenberg and Morris (2007).
the Bayesian game induced by the singleton type space: there is no conditionally independent conjecture over actions, states, and types that justifies $U$ for type $t_1^*$. Thus $D$ is the only IIR action for type $t_1^*$. On the other hand, if type $t_1'$ conjectures that type $t_2'$ will play $L$ and type $t_2''$ will play $R$, then he attaches probability $1/3$ to each of the state-action pairs $(1, L)$ and $(0, R)$. This is enough to make $U$ a best response. Thus both $U$ and $D$ are interim independent rationalizable for types $t_1^*$ and $t''_1$. Hence, although $t_1^*$ and $t''_1$ correspond to the same hierarchy of beliefs, IIR delivers different predictions for the two type spaces. Consider now ICR instead: it’s easy to see that in both type spaces, both $U$ and $D$ are rationalizable for all types of player 1: in either type space, suppose that 1 believes that with probability $1/2$ the true state is 0 and player 2 plays $L$, and with probability $1/2$ the true state is $\theta = 1$ and that 2 plays $R$. Then, $U$ is a best response.

The types in $T'$ are redundant (cf. example 1.3); nonetheless they differ in their conjectures about their opponents and this is potentially important depending on the choice of solution concept. Redundant types can serve as a correlating device, and so these types are not truly “redundant” unless the addition of correlating devices has
no effect.

Under the assumption of conditional independence of IIR, the correlation implicit in type space $T'$ affects the solution concept predictions in the example above. In contrast, ICR allows players to have correlated conjectures about the opponent’s actions, types, and the state, so the ability of “redundant types” to support such correlation is, truly, redundant. Hence, unlike IIR, the predictions of ICR only depend on the hierarchies of beliefs, not on the type space used to represent them. This is the content of the next result, from Dekel, Fudenberg and Morris (2007, DFM07):

**Proposition 1.3** (Proposition 1 in DFM07). If types $t_i, t'_i$ (in possibly different type spaces) induce the same $\Theta$-hierarchy, then $ICR_i(t_i) = ICR_i(t'_i)$.

**Incomplete information games: A direct approach**

The specification of a type space is unnecessary for a non-equilibrium approach: Agents’ subjective situations can be explicitly modeled, and their conjectures about the opponents’ information and beliefs can be incorporated in the solution concept, without need to specify a type space. The most direct approach, is to model agents’ information explicitly. Agents may have both payoff-relevant and payoff-irrelevant information. Payoff-relevant information should be understood as knowledge of components that directly affect agents’ payoffs, such as a firm’s knowledge of her own productivity. This information is represented by the payoff-types $\theta_i$ introduced above. Agents though may also have payoff irrelevant information: this information does not affect payoffs directly, but may be strategically relevant nonetheless, either because it may be thought to be correlated with the payoff-state $\theta$ or with the opponent’s actions. Economic examples abound: geological information, or satellite photographs
of a tract of land on sale are thought to be correlated with the value of the recoverable resources; experts’ reports on an object are thought to be correlated with its value; personality traits and propensities may be thought to be correlated with ability, etc.

A complete model should include all the potentially relevant aspects of which players are commonly aware. Let \( \xi_i \in \Xi_i \) denote a realization of all the payoff-irrelevant but (potentially) strategically relevant aspects known by player \( i \). The pair \( x_i = (\theta_i, \xi_i) \in X_i \) (with \( X_i = \Theta_i \times \Xi_i \)) describes \( i \)'s private information. Hence, a direct representation of the environment consists of a tuple

\[
\mathcal{E} = \left( \Theta_0, (\Theta_i, \Xi_i, A_i, u_i)_{i=1,2} \right).
\]

Battigalli and Siniscalchi (2003, 2007) introduced the solution concept of \( \Delta \)-rationalizability: an “umbrella notion” defined on the environment \( \mathcal{E} \) and parameterized by information-dependent restrictions on players’ beliefs about the primitives. Formally, for each player \( i = 1, 2 \) and each \( x_i \in X_i \), let \( \Delta_{x_i} \subseteq \Delta(\Theta_0 \times X_{-i} \times A_{-i}) \) denote a set of possible beliefs about the exogenous state \( (\theta_0, x_{-i}) \) and the opponent’s action, representing restrictions on \( i \)'s conjectures when his information is \( x_i \). For every \( x_i \) and every \( \mu^i \in \Delta_{x_i} \), denote by \( BR_i(\mu^i, x_i) \) the set of \( i \)'s best responses to conjectures \( \mu^i \), when his information is \( x_i \). Then, define an iterative deletion procedure that takes such restrictions on beliefs into account:

**Definition 1.6.** The set \( R^\Delta \) of \( \Delta \)-rationalizable profiles of information-action pairs is recursively defined as follows: for every player \( i \) let \( R^\Delta_{i,0} = X_i \times A_i = \Theta_i \times \Xi_i \times A_i \) and then, for all \( k \geq 0 \),

\[
R^\Delta_{i,k+1} = \left\{ (\theta_i, \xi_i, a_i) \in X_i \times A_i \left| \begin{array}{l}
\exists \mu_i \in \Delta(\theta_i, \xi_i) \text{ such that } \\
\text{supp } \mu_i \subseteq \Theta_0 \times R^\Delta_{-i,k} \\
a_i \in BR_i(\mu^i, x_i)
\end{array} \right. \right\}.
\]

Finally, let \( R^\Delta_i = \cap_{k \geq 0} R^\Delta_{i,k} \) and \( R^\Delta = R^\Delta_1 \times R^\Delta_2 \).
Battigalli and Siniscalchi show that \( R^\Delta \) is characterized by the epistemic assumptions of rationality and common belief in rationality and the \( \Delta \)-restrictions. Hence, RCBR obtains setting \( \Delta_{x_i} = \Delta(\Theta_0 \times X_{-i} \times A_{-i}) \) for every \( x_i \).

**A unified view**

Battigalli, Di Tillio, Grillo and Penta (2009, BDGP hereafter) provide a unified view of these and other versions of rationalizability. An important notion in BDGP is that of *information-based type space* (or type space with *information types*): type space \( T = \langle (T_i, \tau_i) \rangle_{i=1,2} \) has *information types* if, for each \( i \), the set of types \( T_i \) is (isomorphic to) \( X_i \). In this case, Harsanyi-types can be interpreted as information, because every epistemic component corresponds to some payoff irrelevant signal \( \xi_i \) that belongs to the environment.\(^{13}\) Hence, in this case, the type space restrictions on agents’ beliefs find an immediate representation in the environment, as a set \( \Delta \) of (information-dependent) restrictions on beliefs about the primitives: for each \( i \) and \( x_i \), the set \( \Delta_{x_i} \) is the set of measures \( \mu^i \in \Delta(\Theta_0 \times X_{-i} \times A_{-i}) \) such that (M) \( \text{marg}_{\Theta_0 \times X_{-i}} \mu^i = \tau_i(x_i) \) (that is, \( \text{marg}_{\Theta_0 \times X_{-i}} \mu^i \) is the belief of information-type \( x_i \) in \( T \)). Based on this observation, \( \Delta \)-rationalizability and ICR coincide in the following sense:

**Proposition 1.4** (Proposition 1 in BDGP). *If \( T \) has information types and \( \Delta \) is the set of restrictions corresponding to \( T \), then \( \Delta \)-rationalizability coincides with ICR.*

For the case of *information types*, it is possible to relate \( \Delta \)-rationalizability to IIR as well: Say that a set \( \Delta \) of restrictions *CI-corresponds* to \( T \) if, for each \( i \) and \( x_i \), \( \Delta_{x_i} \) is the set of measures \( \mu^i \in \Delta(\Theta_0 \times X_{-i} \times A_{-i}) \) such that (M) \( \text{marg}_{\Theta_0 \times X_{-i}} \mu^i = \)

\(^{13}\)In *information-based type spaces*, “redundant” types that induce the same \( \Theta \)-hierarchy are easily interpreted as corresponding to the same payoff-type but to different payoff-irrelevant information, \( \xi_i \) (notice though that they would not be redundant in terms of \( \Theta \times \Xi \)-hierarchies.)
\( \tau_i(x_i) \) and (CI) \( \mu^i \left[ x_{-i} \right] > 0 \) implies \( \mu^i[\theta_0, a_{-i}|x_{-i}] = \mu^i[\theta_0|x_{-i}]\mu^i[a_{-i}|x_{-i}] \), that is, \( i \) believes that \( \theta_0 \) and \( a_{-i} \) are independent conditional on \( x_{-i} \). The following result shows that the procedure of taking the interim strategic form of a Bayesian game with information types and then computing the rationalizable strategies amounts to imposing the conditional independence restriction (CI) on top of the restrictions (M) implied by the type space:

**Proposition 1.5** (Proposition 3 in BDGP). If \( T \) has information types and \( \Delta \) is the set of restrictions that CI-correspond to \( T \), then \( \Delta \)-rationalizability coincides with IIR.

Hence, for the case of information types, ICR corresponds to RCBR and common certainty of the restrictions imposed by the type space (M), while IIR adds to these conditions the assumption of conditional independence (CI) and common belief in conditional independence.

For type spaces that are not “information-based”, Harsanyi-types do not find a natural interpretation in terms of the primitives of the environment, because in that case the epistemic components are mere parametrizations of agents’ beliefs that do not correspond to any information present in the environment. In this case, providing epistemic foundations to solution concepts defined for Bayesian games is more problematic.\(^{14}\) Moreover, without information types, it is not clear whether the common practice of taking Bayesian games (or type spaces) as “primitive objects” is meaningful at all: Harsanyi-types are just parameters, imposed by the modeler for convenience, they have nothing to do with the fundamentals of the economic problem.\(^{15}\)

\(^{14}\)BDGP provide an epistemic characterization of ICR also for the case without information types.

\(^{15}\)In particular, it is not clear what it means to have redundant types in a non-information based
1.4 Interim Robustness in Static Games

The implicit assumption that a game (with or without complete information; modelled following Harsanyi’s or the “direct” approach) is “common knowledge” among the agents inherently imposes restrictions on agents’ beliefs of all orders. The existence of a universal type space guarantees that, in principle, this assumption entails no loss of generality. In applied work though assumptions are imposed that do entail some loss of generality. To the extent that these assumptions are understood to hold only approximately in the actual situation, the significance of game theoretic predictions relies crucially on robustness properties of the solution concepts adopted.

Section 1.4.1 discusses the premises of adopting refinements of rationalizability when no common knowledge (CK) assumptions on payoffs are imposed. The main result in the section implies that, when all common knowledge assumptions on payoffs are relaxed, the strongest “robust” predictions are those based on ICR alone.

In Section 1.4.2 instead we consider a somewhat complementary question: rather than relaxing common knowledge assumptions on payoffs, we maintain common knowledge of the game structure $G$, but we impose no further assumptions on agents’ beliefs about the environment (that is, we don’t specify a type space). In such belief-free environments, we address the premises of the equilibrium hypothesis when no assumptions on agents’ beliefs are imposed. It will be shown that the only results of the Bayesian (equilibrium) analysis that are independent of the exact specification of the type space are those based on RCBR alone (that is, on $\Delta$-rationalizability with $\Delta_{x_i} = \Delta (\Theta_0 \times X_{-i} \times A_{-i})$ for all $x_i$.)

Hence, RCBR essentially provides the answer to both robustness problems.

type space. For a related point, see Liu (2009)
1.4.1 Relaxing all CK assumptions on payoffs

Any game theoretic analysis is based on some commonly known structure. So, what does it mean to “relax all common knowledge assumptions on payoffs”? It essentially means to assume an underlying space of uncertainty that is “sufficiently rich” to allow enough freedom to agents’ preferences over the outcomes: it is making precise the meaning of “sufficiently rich” that clarifies what assumptions are being relaxed. For instance, consider the game in example 1.6, p. 27: in the complete information model, with \( \Theta^* = \{2/5\} \), the game is a coordination game, and it is commonly known that no action is dominant; if we set \( \Theta^* = \{-2/5, 2/5\} \) then it is not common knowledge that ‘Not’ is not dominant, but common knowledge that ‘Attack’ is not dominant is maintained. If instead we set \( \Theta^* = \{-2/5, \theta^*, 6/5\} \), we relax common knowledge that either action is not dominant. Hence, relaxing all CK assumptions essentially means to consider all possible hierarchies of beliefs players may have about a sufficiently rich space of uncertainty \( \Theta^* \):\(^{16}\) assuming common knowledge of the \( \Theta^* \)-based universal type space, \( T_{\Theta^*} \), entails no loss of generality. \( T_{\Theta^*} \) will thus be referred to as the universal model. A solution concept \( S \) can thus be seen as a correspondence that assigns to each player’s hierarchy of beliefs (the exogenous variables) a set of strategies (the endogenous variables), i.e., \( S = (S_i)_{i=1,2} \) such that \( S_i : T_{i,\Theta^*} \rightrightarrows A_i \), \( i = 1,2 \).

The universal model is a benchmark, but it is never used in practice: In modeling a strategic situation, applied theorists typically select only a subset of the possible

\(^{16}\)The representation \( G^* = \langle \Theta^*, (A_i, u_i)_{i=1,2} \rangle \), where \( u_i : \Theta^* \times A \rightarrow \mathbb{R} \), is without loss of generality: For example, taking the underlying space of uncertainty \( \Theta^* \equiv \left( \mathbb{R}^2 \right)^A \) imposes no restrictions on agents’ preferences (beyond the assumption that they both are expected-utility maximizers of course.)
hierarchies to focus on. To the extent that the “true” hierarchies are understood to be only close to the ones considered in a specific model, the concern for robustness of the theory’s predictions translates into a continuity property of the solution concept correspondence. Here we consider a solution concept “robust” if it never rules out strategies that are not ruled out for arbitrarily close hierarchies of beliefs: This is equivalent to requiring upper hemicontinuity of the solution concept correspondence on $T_{i,\Theta}$. Dekel, Fudenberg and Morris (2006, DFM06) showed that $ICR$ is “robust” in this sense:

**Proposition 1.6 (DFM06).** The correspondence $ICR_i : T^*_{i,\Theta} \Rightarrow A_i$ is upper hemicontinuous.

*Upper hemicontinuity in the universal model* addresses a specific robustness question: Robustness, with respect to small “mistakes” in the modeling choice of which subset of players’ hierarchies to consider. Though, as discussed in Section 1.3.1, when applied theorists choose a subset of $\Theta^*$-hierarchies to focus on, they typically represent them as Harsanyi-types in (non universal) *type spaces*, rather than elements of $T^*_{\Theta}$, imposing common knowledge assumptions that do entail some loss of generality. The particular type space chosen to represent a given set of hierarchies of beliefs may potentially affect the predictions of a solution concept (cf. example 1.5). Invariance of the predictions with respect to the type space is thus a distinct robustness property. This property is called *type space-invariance*. Proposition 1.3 can thus be interpreted as a robustness result for $ICR$: namely, that $ICR$ is *type space-invariant*.

$ICR$ is thus robust in two senses: it is type space-invariant, and it is upper hemicontinuous. Clearly, a solution concept that never rules out anything is upper hemicontinuous (hence “robust”), but not interesting. One way to solve this trade-off is to look for a “strongest robust” solution concept. In fact, if $\Theta^*$ is “sufficiently
rich” (i.e., if “enough” common knowledge assumptions are being relaxed), Weinstein and Yildiz (2007, WY hereafter) proved that ICR is the strongest robust solution concept. The following example illustrates the point:

**Example 1.6.** Consider a game $G(\theta)$ with payoffs parametrized by $\theta \in \mathbb{R}$ as in the following matrix:

\[
\begin{array}{c|cc}
& \text{Attack} & \text{Not} \\
\text{Attack} & \theta, \theta & \theta - 1, 0 \\
\text{Not} & 0, \theta - 1 & 0, 0 \\
\end{array}
\]

Let $\theta^* = 2/5$ and let $T_{\text{CK}} = \{t_{\text{CK}}\}$ denote the model in which $\theta^*$ is common knowledge. This delivers a coordination game, with (Attack, Attack) and (Not, Not) as the two pure strategy Nash Equilibria. All actions are rationalizable in this game, while only the equilibrium (Not, Not) is risk-dominant.

Now let the space of uncertainty be $\Theta^* = \{\theta^*, -2/5, 6/5\}$, so that any strategy is dominant in some state, and consider type space $T$ with set of types $T_1 = \{-1, 1, 3, \ldots\}$ and $T_2 = \{0, 2, 4, \ldots\}$, and beliefs as follow: type $-1$ puts probability one on state $\theta' \neq \theta^*$ and type $0$; type $0$ puts probability $p \in (0, 1)$ on $(\theta', -1)$ and $(1 - p)$ on $(\theta^*, 1)$; types $k = 1, 2, \ldots$ all put probability $p \in (0, 1)$ on $(\theta^*, k - 1)$ and probability $(1 - p)$ on $(\theta^*, k + 1)$. The sequence of types profiles $\{(k, k + 1)\}_{k/2 \in \mathbb{N}}$ converges to common certainty of $\theta^*$ as $k \to \infty$: Type $k = 2$ is certain of $\theta^*$, and also puts probability one on the opponent’s type being 1 or 3, who are also certain of $\theta^*$. Hence type 2 is consistent with mutual certainty of $\theta^*$, but not with common certainty: Type 2 puts positive probability on type 1, which puts positive probability on type 0, who assigns probability $p$ to the state being $\theta' \neq \theta^*$. Similarly, any type $k$ is consistent with $k$ iterations of mutual certainty of $\theta^*$, so that the first $k$ orders of beliefs of type $k$ are the same as type $t_{\text{CK}}$. But as long as $p \in (0, 1)$, any finite $k$ is not consistent with
common certainty of $\theta^*$: Common certainty is only approached letting $k \to \infty$.\footnote{This convergence is in the product topology of hierarchies of beliefs.}

Now, suppose that $\theta' = -2/5$ and $p = 2/3$: Type $-1$ plays Not, as he is certain that it is dominant; type $0$ is certain of $\theta^*$, but puts probability $2/3$ on type $-1$, who plays Not, hence playing Not is his unique rationalizable strategy. The argument can be iterated, so that for each $k > 1$, despite there are $k$ levels of mutual certainty of $\theta^*$, playing Not is the unique ICR action. It is easy to check that the same reasoning can be repeated to obtain Attack as uniquely ICR, simply by letting $\theta' = 6/5$.\footnote{From an \textit{ex ante} perspective, the type space in example 1.6 cannot be considered “close” to the complete information model $T_{\text{CK}}$: Remember though that here we are focusing on “genuine” incomplete information, hence the natural perspective is the \textit{interim} one. For robustness analysis from an \textit{ex ante} point of view, see e.g. Kaji and Morris (1997).}

Hence, the strongest robust predictions are those based on rationalizability alone: For instance, any solution concept that ruled out $(\text{Attack}, \text{Attack})$ in the complete information model (e.g., risk-dominance) would not deliver “robust predictions”, because there exists a sequence of hierarchies of beliefs, arbitrarily close to $t_{\text{CK}}$, in which $(\text{Attack}, \text{Attack})$ is uniquely selected. Similarly, any solution concept that ruled out $(\text{Not}, \text{Not})$ would not be robust.

For their general result, WY adopt the following notion of “richness”:

\textbf{WY’s Richness Condition:} For every agent and every $a_i \in A_i$, there exists a payoff state $\theta^{a_i} \in \Theta^*$ that makes $a_i$ strictly dominant in $G(\theta^{a_i})$.

WY’s main result is the following:

\textbf{Proposition 1.7.} If $\Theta^*$ is “rich”, then for every type $t_i \in T_{i, \Theta}^*$, and for every $a_i \in ICR_i(t_i)$, there exists a sequence $\{t_{i, \nu}^\nu\}_{\nu \in \mathbb{N}} \subseteq T_{i, \Theta}^*$ such that $t_{i, \nu}^\nu \to t_i$ and $ICR_i(t_{i, \nu}^\nu) =$
\{a_i\} for each \( \nu \in \mathbb{N} \). Furthermore, the set of types for which ICR is unique is open and dense in \( T_{i,\Theta} \).

The generic uniqueness result in the second part of the proposition generalizes an important insight from the literature on global games, i.e. that multiplicity can be seen as the direct consequence of the common knowledge assumptions.\(^{19}\) Relaxing such assumptions, RCBR is (generically) sufficient to explain agents’ coordination and deliver uniqueness. But notice here that the unique rationalizable outcome in the perturbation depends on the choice of the sequence. Hence, WY’s result does not support a selection argument: Generically, ICR delivers a unique prediction, but once we have multiplicity for a given type, it is not possible to refine it in a “robust” way.

**WY’s program.** To summarize what we said so far, WY’s program involves the following steps: 1) Start with a standard model (such as \( t^{CK} \)), which makes implicit common knowledge assumptions on the payoff structure and players’ hierarchies of beliefs. 2) To relax all the CK assumptions, this model is embedded in a larger one, with an underlying space of uncertainty \( \Theta^* \): relaxing all CK assumptions, essentially means to consider all possible hierarchies of beliefs players may have about a sufficiently rich space of uncertainty. In example 1.6, type \( t^{CK} \) was “embedded” in a larger model as a type (call it \( t^* \)) corresponding to common certainty of \( \theta^* \). 3) Once the common knowledge assumptions of the original model are embedded as common certainty assumptions in the “universal model”, the robustness of a solution concept can be formulated as a continuity property in this space. Hence, the properties of

\(^{19}\)On global games, see Morris and Shin (2003) and references therein. Carlsson and Van Damme (1993) is the seminal paper of the literature.
sequences converging to the common certainty type $t^*$ could be used to address the robustness of the predictions of the “common knowledge” model $t^{CK}$.

The possibility of envisioning $t^{CK}$ as $t^*$ is central to the argument, but rests on the interchangeability of knowledge and certainty. Modeling knowledge as certainty is innocuous for the purpose of WY’s analysis, but as we will see in Section 1.5.1 the distinction becomes crucial when the analysis is extended to dynamic settings. As a consequence, the properties of a solution concept on the “universal model” do not provide an exhaustive answer to the original robustness question.

**Relaxing some CK assumptions.**

WY proved that when all CK assumptions are relaxed (i.e. the richness condition is imposed), the strongest robust predictions are those based on ICR alone. In many situations though, imposing richness may be an unnecessarily demanding robustness test: The richness condition on $\Theta^*$ implies that it is not common knowledge that any strategy is not dominant. However, as modelers, we may wish to assume that some features of the environment actually are common knowledge. For example: common knowledge that some strategies are not dominant. In that case, the underlying space of uncertainty does not satisfy richness. Chapter 2 explores the structure of ICR on the universal type space based on arbitrary spaces of uncertainty, i.e. without assuming richness.

Let $\mathcal{A}_i^0 \subseteq A_i$ be the set of actions of player $i$ for which there exists a dominance state $\theta^{\alpha_i} \in \Theta^*$. For each $k = 1, 2, ..., $ let $\mathcal{A}_i^k$ be equal to $\mathcal{A}_i^{k-1}$ union the set of actions that are strict best responses to conjectures concentrated on $\mathcal{A}_i^{k-1}$. Finally, let $\mathcal{A}_i^\infty = \bigcup_{k \geq 0} \mathcal{A}_i^k$.

---

20WY do not consider payoff types. Hence, in our terminology, players have no information in WY’s setting.
**Proposition 1.8** (Proposition 2.1). *Whenever a type $t_i$ has multiple rationalizable actions, for any $a_i$ of these actions that (i) belong to $A_{i}^{\infty}$ and (ii) are justified by conjectures concentrated on $A_{i}^{\infty}$, there is a sequence of types converging to $t_i$ for which $a_i$ is uniquely rationalizable.*

Notice that WY’s results obtain setting $A_i^0 = A_i$ for each $i$, i.e. assuming (strict) dominance regions for each player’s action (WY’s richness condition). Notice also the following: suppose that (a) there exists a payoff state $\theta^* \in \Theta^*$ for which payoff functions are supermodular, with each player’s higher and lower actions $a_i^h$ and $a_i^l$ respectively, and such that no action is dominated; and (b) for each $i$, $A_i^0 = \{a_i^l, a_i^h\}$. Then, it is easy to see that under these conditions $A^\infty = A$, and WY’s full results are obtained. Conditions (a) and (b) correspond to the case considered by the global games literature, in which the underlying game has strategic complementarities and dominance regions are assumed for the extreme actions only. The difference is that in that literature supermodularity is assumed at all states (so that it is CK). In contrast, here it may be assumed for only one state, which only entails relaxing CK that the game is *not* supermodular. This observation clarifies that, on the one hand, the equilibrium selection results obtained in the global games literature, which contrast with WY’s non-robustness result, are exclusively determined by the particular class of perturbations that are considered, not by the fact that some (as opposed to all) CK-assumptions are relaxed. On the other hand, the generic uniqueness result can be obtained without assuming CK of supermodularity or relaxing all CK-assumptions: for instance, relaxing CK that the game is *not* supermodular and that the corresponding extreme actions are *not* dominant would suffice to obtain the full results of WY.
1.4.2 Belief-Free Models and Equilibrium

The specification of the type space is often unrelated to the fundamentals of the economic problem (cf. Section 1.3.1). Yet, it may crucially affect the set of equilibrium outcomes. This raises several related theoretical questions. Can we analyze incomplete information games without specifying a type space? Which results of the Bayesian analysis are independent of the exact specification of the type space? Is it possible to provide a relatively simple characterization of the set of all Bayesian equilibrium outcomes?

The notion of $\Delta$-rationalizability introduced above (definition 1.6) provides an answer to these questions: Battigalli and Siniscalchi (2003, BS) showed that the set of $\Delta$-(rationalizable strategies in a game with payoff uncertainty $G$ characterizes the set of strategies played as part of Bayesian equilibria in Bayesian games obtained attaching some type space to $G$:\textsuperscript{21}

Proposition 1.9 (Propositions 4.2 and 4.3 in BS). Fix a game with payoff uncertainty $G$, and let $\Delta_{x_i} = \Delta(\Theta_0 \times X_{-i} \times A_{-i})$ for all $x_i$. Then: $a_i \in R^\Delta(\theta_i, \xi_i)$ if and only if there exist a type space $T$ and a type $t_i = (\theta_i, e_i)$ in $T$ that plays $a_i$ in a Bayesian equilibrium of the Bayesian game $B = (G, T)$.\textsuperscript{22}

This proposition is the incomplete information counterpart of Proposition 1.1. The analogy between the two is made transparent by Proposition 1.2, relating the rationalizable strategies and the set of Bayesian equilibria obtained adding redundant

\textsuperscript{21}See also Bergemann and Morris (2009) for a related result.

\textsuperscript{22}Battigalli and Siniscalchi proved a more general result, for arbitrary restrictions: they show that $\Delta$-rationalizability exactly characterizes the set of Bayesian equilibrium outcomes “consistent with arbitrary restrictions $\Delta$.”
types to a complete information game.\footnote{In fact, proposition 1.2 is just a special case of proposition 1.9.}

Proposition 1.9 implies that, if no assumptions on the agents’ beliefs are imposed, the \textit{equilibrium hypothesis} has essentially no bite: the set of equilibrium strategies coincides with the set of rationalizable strategies. Hence, RCBR characterizes the implication of equilibrium analysis in belief-free environments.

1.4.3 Discussion

We see that RCBR provides a unique answer to several notions of robustness. Going to dynamic environments, there are several “natural” analogues of RCBR. As we will see, such unity is lost in dynamic games: different extensions of RCBR provide the answer to different questions.

1.5 Dynamic Games

In a dynamic game, agents’ information about the environment is endogenous, and may change as the game unfolds. Moreover, this may happen in ways that are unexpected to the agents. This opens important questions concerning the modeling activity, and the assumptions on agents’ process of revising beliefs, both about exogenous variables (such as the payoff states) and about the opponents’ behavior. Furthermore, as the game unfolds, agents’ beliefs about the exogenous variables will in general depend on their beliefs on the endogenous: for instance, agents may draw inferences about the opponents’ private information from their behavior, and notions of equilibrium for dynamic games will have to take this possibility into account.

While mostly inconsequential in static games, the distinction between \textit{knowledge} and \textit{certainty} (probability-one belief) is crucial in dynamic games. If an agent
“knows” something, he would be certain of it after observing any event (expected or not). Not so if “knowledge” is replaced by probability-one belief, as Bayes’ rule puts no restrictions on the conditional beliefs after unexpected events.

Related to this, also assumptions about rationality (and interactive beliefs about it) must be formulated with greater care. First of all, the natural notion of rationality in dynamic settings is “sequential rationality”: an agent is rational not only if he plans to play optimally at the beginning of the game, but also if he does so conditional on each possible piece of information he may observe later in the game, expected or not.

As the very notion of rationality in dynamic games involves restrictions on agents’ behavior at each point in the game, so also beliefs about the opponents’ rationality must be made explicit at each point of the game. Consider the following example:

**Example 1.7.** Consider the game in figure 1.2, and suppose that it is common knowledge that \( \theta = 0 \) (hence, the game has complete information. Denote this model by \( T^{CK} = \{t^{CK}\} \)). Then, strategy \( a_3 \) is dominated by \( a_1 \). Thus, if at the beginning of the game player 2 thinks that 1 is rational, he assigns zero probability to \( a_3 \) being played. For example, 2 could assign probability one to \( a_1 \), so that the next information set is unexpected. We can consider two different hypothesis on the agents’ “reasoning”
in the game:

- [H.1] 2 believes that 1 is rational even after an unexpected move; or

- [H.2] 2 believes that 1 is rational as long as he is not surprised, but he is willing to consider that 1 is not rational if he observes an unexpected move.

If [H.1] is true, in the subgame player 2 would still assign zero probability to \(a_3\), and play \(b_1\) if rational. If 1 believes [H.1] and that 2 is rational, he would expect \(b_1\) to be played. Then, if 1 is also rational, he would play \(a_2\). This is the logic of Pearce’s (1984) Extensive Form Rationalizability (EFR), which delivers \((a_2, b_1)\) as the unique outcome in this game.

Now, let’s maintain that 2 is rational, but assume [H.2] instead: once surprised, player 2 is willing to consider that 1 is not rational. Hence, in the subgame, he may assign probability one to \(a_3\) being played, which would justify \(b_2\). If at the beginning 2 assigned positive probability to \(a_2\), then the subgame would not be unexpected, and player 2 would still assign zero probability to \(a_3\), making \(b_1\) the unique best response. Thus, if [H.2] is true, either \(b_1\) or \(b_2\) may be played by a rational player 2. If 1 believes that 2 is rational and that [H.2] is true, he cannot rule out either \(b_1\) or \(b_2\), and so both \(a_1\) and \(a_2\) may be played by a rational player 1. This is the logic of Ben-Porath’s (1997) Common Certainty of Rationalizability (CCR), which selects \(\{a_1, a_2\} \times \{b_1, b_2\}\) in this game.

Formally, a dynamic game is defined by an extensive form \(\langle N, \mathcal{H}, \mathcal{Z} \rangle\) (\(N = \{1, ..., n\}\) is the set of players; \(\mathcal{H}\) and \(\mathcal{Z}\) the sets of partial and terminal histories, denoted by \(h\) and \(z\) respectively) and players’ payoffs, defined over the terminal histories. Incomplete information is modelled as in Section 1.3, parametrizing the payoff functions on a space of uncertainty \(\Theta\), letting \(u_i : \mathcal{Z} \times \Theta \to \mathbb{R}\). The tuple

35
\[ \langle \Theta_0, (\Theta_i, u_i)_{i \in N} \rangle, \] where \( u_i : \mathcal{Z} \times \Theta \to \mathbb{R} \) for each \( i \in N \), represents players’ information about everyone’s preferences, and is referred to as information structure.

An extensive form and an information structure induce a dynamic game with payoff uncertainty, \( \Gamma = \langle N, \mathcal{H}, \mathcal{Z}, \Theta_0, (\Theta_i, u_i)_{i \in N} \rangle \). This is not a dynamic Bayesian game, because a model of agents’ beliefs is not specified.

**Strategic Forms.** Agents’ strategies in the game \( \Gamma \) are functions \( s_i : \mathcal{H} \to A_i \) such that \( s_i (h) \in A_i (h) \) for each \( h \in \mathcal{H} \) (\( A_i (h) \) denotes the actions that \( i \) may choose at \( h \)), and \( A_i = \bigcup_{h \in \mathcal{H}} A_i (h) \). The set of player \( i \)’s strategies is denoted by \( S_i \), and as usual we define the sets \( S = \times_{i \in N} S_i \) and \( S_{-i} = \times_{j \neq i} S_j \). Any strategy profile induces a terminal allocation \( z (s) \in \mathcal{Z} \). Hence, we can define strategic-form payoff functions \( U_i : S \times \Theta \to \mathbb{R} \) as \( U_i (s, \theta) = u_i (z (s), \theta) \) for each \( s \) and \( \theta \). If strategies \( s_i \) and \( s'_i \) only differ at histories that are inconsistent with \( s_i \), then they are said to be “outcome equivalent”. Reduced form strategies, denoted by \( S_i \), are the equivalence classes of the outcome equivalence relation. Reduced strategies \( s_i \in S_i \) are also called *plans of actions*: each \( s_i \) only specifies actions at histories that are reachable by \( s_i \) itself.

As the game unfolds, agents learn about the environment from the private signals, but they also learn about the opponents’ behavior through the public histories: for each history \( h \) and player \( i \), let \( S_i (h) \) denote the set of player \( i \)’s strategies that are consistent with history \( h \) being observed.

**Bayesian Games** Dynamic Bayesian games are defined as for the static case, attaching a \( \Theta \)-based type space \( T_\Theta \) that implicitly represents a set of hierarchies of beliefs. All comments about Harsanyi-types, information, ex-ante vs. interim approach, etc., made in Section 1.3.1, obviously apply to these settings as well. (Pure) strategies in the Bayesian game are mappings \( \sigma_i : T_i \to S_i \). To distinguish \( \sigma_i \in \Sigma_i \)
from \( s_i \in S_i \), the latter will be referred to as “interim strategies”.

The next two sections explore robustness questions in dynamic games analogous to those discussed in Sections 1.4.1 and 1.4.2 for static games.

1.5.1 Relaxing all CK-assumptions on payoffs (II)

The results of Section 1.4.1 do not cover dynamic games for one technical reason: WY’s “richness condition” cannot be satisfied by the normal form of a dynamic game, as strategies that only differ at nodes that the opponents can prevent from being reached cannot strictly dominate each other. An obvious way to overcome the problem is to modify the notion of dominance: Strategy \( s_i \) is conditionally dominant at state \( \theta \in \Theta^* \) if, for every history \( h \) on its path, and every \( s'_i \in S_i(h) \), if \( s_i(h) \neq s'_i(h) \), then

\[
u_i\left(z(s_i, s_{-i}), \theta\right) > u_i\left(z(s'_i, s_{-i}), \theta\right) \text{ for all } s_{-i} \in S_{-i}(h).
\]

“Richness” for dynamic games: for every \( s \in S \), there exists \( \theta^* = (\theta^*_0, \theta^*_i, \theta^*_z) \in \Theta^* \) such that \( s_i \) is conditionally dominant at \( \theta^* \) for every \( i \).

Based on this richness condition, in chapter 3 I conduct an analysis analogous to the one discussed in Section 1.4.1. To do this, a new solution concept for Bayesian games in extensive form is introduced, Interim Sequential Rationalizability (ISR).\(^{24}\) ISR extends the logic of Ben-Porath’s (1997) CCR (see example 1.7) to dynamic Bayesian games: similar to CCR, ISR expresses the assumption that agents are (sequentially) rational and share common certainty of (sequential) rationality at the

\(^{24}\) Independent work of Chen (2009) addressed the same question, but maintaining WY’s normal form approach: that is, Chen applies ICR to the reduced interim normal form of a dynamic game, only substituting WY’s richness condition with the version for dynamic games. In chapter 3 it is shown that such a normal form approach is only viable for the case of no-information environments, while the extensive form approach applies to general information structures. This point is discussed below, p. 45.
beginning of the game, but no restrictions are imposed on their conjectures after unexpected events; similar to ICR, ISR is defined as a “type-by-type” iterative deletion procedure for interim (reduced form) strategies $s_i$, that imposes the **consistency condition** that a type’s conjectures at the beginning of the game agree with his beliefs (as specified in the type space), and conjectures allow correlation between the opponents’ behavior and the states of nature.

Hence, ISR represents the assumptions of **(sequential) rationality, initial common certainty of (sequential) rationality and of the type space restrictions**: “initial”, because no restrictions are imposed after unexpected events.25

Proposition 3.4 below provides a result analogous to the first part of Proposition 1.7 above:

**Proposition 1.10** (Proposition 3.4). If $\Theta^*$ satisfied the “richness condition for dynamic games”, then for every type $t_i \in T_i^{*}_{\Theta^*}$ and for every $s_i \in ISR_i (t_i)$, there exists a sequence $\{t^\nu_i\}_{\nu \in \mathbb{N}} \subseteq T_i^{*}_{\Theta^*}$ such that $t^\nu_i \to t_i$ and $ISR_i (t^\nu_i) = \{s_i\}$ for each $\nu \in \mathbb{N}$.

This result implies, for instance, that despite the fact that the EFR logic is compelling in example 1.7, its predictions are not robust. To illustrate the point we will construct a sequence of hierarchies of beliefs, converging to the hierarchies in example 1.7, in which $(a_1, b_2)$ is the unique ISR-outcome (hence, also the unique EFR outcome). Since $(a_1, b_2)$ is ruled out by EFR in the limit, but uniquely selected along the converging sequence, EFR is not “robust”.

25 Battigalli and Siniscalchi (2007) notion of weak $\Delta$-rationalizability, characterized by the epistemic assumptions of initial common certainty in rationality and in the $\Delta$-restrictions, is also related to ISR. That solution concept though is defined for games with payoff uncertainty, not Bayesian games. In the case of information-types (Section 1.3.2), a result analogous to Proposition 1.4 can be proved that relates ISR and weak $\Delta$-rationalizability, when the restrictions $\Delta$ are derived from the type space.
Example 1.8. In the game of figure 1.2, let the space of uncertainty be \( \Theta = \{0, 3\} \).

Suppose that player 1 knows the true state, while 2 doesn’t (and this is common knowledge). Let \( t^* = (t_1^*, t_2^*) \) represent the situation in which there is common certainty that \( \theta = 0 \): “type” \( t_1^* \) knows that \( \theta = 0 \), and puts probability one on 2 being type \( t_2^* \); type \( t_2^* \) puts probability one on \( \theta = 0 \), and player 1 being type \( t_1^* \). A reasoning similar to that in example 1.7 implies that \( \{a_1, a_2\} \) and \( \{b_1, b_2\} \) are the sets of ISR strategies for \( t_1^* \) and \( t_2^* \), written \( \text{ISR}(t^*) = \{a_1, a_2\} \times \{b_1, b_2\} \).

Now, a sequence of types \( \{t^m\} \) will be constructed, converging to \( t^* \), such that \( (a_1, b_2) \) is the unique ISR-outcome for each \( t^m \): Since it is the unique ISR-outcome along the sequence, any (strict) refinement of ISR that rules out \( (a_1, b_2) \) for type \( t^* \) would not be upper hemicontinuous, hence not “robust”. The type spaces used for the construction of the sequence \( \{t^m\} \) can be viewed as a “perturbed” version of Rubinstein’s (1989) e-mail game, in which there is a small probability that the original e-mail is sent in the wrong state of nature.

Fix \( \varepsilon \in (0, \frac{1}{6}) \) and let \( p \in \left(0, \frac{\varepsilon}{(1-2\varepsilon)}\right) \). Consider the set of type profiles \( T_1^\varepsilon \times T_2^\varepsilon \subseteq T_\Theta^* \), where \( T_1^\varepsilon = \{-1^3, 1^0, 1^3, 3^0, 3^3, 5^0, 5^3, \ldots\} \) and \( T_2^\varepsilon = \{0, 2, 4, \ldots\} \). Types \( k^\theta \) (\( k = -1, 1, 3, \ldots, \theta = 0, 3 \)) are player 1’s types who know that the true state is \( \theta \); 2’s types only know their own payoffs, which are constant across states, but don’t know the opponent’s type. Suppose that beliefs are described as follows. Type \(-1^3 \) puts probability one on facing type 0; type 0 assigns probability \( \frac{1}{1+p} \) to type \(-1^3 \), and complementary probability to types \( 1^0 \) and \( 1^3 \), with weights \((1-\varepsilon)\) and \( \varepsilon \), respectively. Similarly, for all \( k = 2, 4, \ldots \) player 2’s type \( k \) puts probability \( \frac{1}{1+p} \) on 1’s types \( (k-1)^0 \) and \( (k-1)^3 \), with weights \((1-\varepsilon)\) and \( \varepsilon \) respectively, and complementary probability \( \frac{p}{1+p} \) on the \( (k+1) \)-types, with weight \((1-\varepsilon)\) on \( (k+1)^0 \) and \( \varepsilon \) on \( (k+1)^3 \). For all other types of player 1, with \( k = 1, 3, \ldots \), and \( \theta = 0, 3 \), type \( k^\theta \) puts probability \( \frac{1}{1+p} \) on
2’s type $k-1$, and complementary probability on 2’s type $k+1$. (The type space is represented in figure 1.2.) Notice that the increasing sequence of even $k$’s and odd $k^0$’s converges to $t^*$ as we let $\varepsilon$ approach 0.\textsuperscript{26} It will be shown that player 2’s types 0, 2, 4, ... only play $b_1$, while 1’s types $1^0, 3^0, ...$ only play $a_1$: All types $k^3$ ($k = -1, 1, 3, ...$) would play $a_3$, for they know it is dominant. Type 0 puts probability $\frac{1}{1+p}$ on type $-1$, who plays $a_3$; given these initial beliefs, type 0’s conditional conjectures after $I_n$ must put probability at least $\frac{1}{1+p}$ on $a_3$ being played, which makes $b_2$ optimal for him. Type $1^0$ also puts probability $\frac{1}{1+p}$ on type 0, who plays $b_2$, thus $a_1$ is the unique best response. Type 2’s initial beliefs are such that type $1^0$ plays $a_1$ and types $1^3$ and $3^3$ play $a_3$. Hence, the probability of $a_3$ being played, conditional on $I_n$ being observed, must be no smaller than

$$\Pr(\theta = 3 \mid \text{not } 1^0) = \frac{\varepsilon}{1 - \frac{1}{1+p} (1 - \varepsilon)} = \frac{(1 + p) \varepsilon}{p + \varepsilon}$$

Given that $p < \frac{\varepsilon}{(1 - 2\varepsilon)}$, this probability is greater than $\frac{1}{2}$. Hence, playing $b_2$ is the unique

\textsuperscript{26}This convergence of the hierarchies of beliefs is in the product topology.
best response, irrespective of type 2’s conjectures about 3’s behavior. Given this, type 3 also plays $a_1$. The reasoning can be iterated, so that for all types $1^0, 3^0, 5^0, ..., a_1$ is the unique $\text{ISR}$ strategy, while for all types $0, 2, 4, ...$ of player 2, strategy $b_2$ is.

Hence, when all common knowledge assumptions are relaxed, any refinement of $\text{ISR}$ is not upper hemicontinuous, i.e. delivers non-robust predictions. Together with Proposition 1.10, the next proposition implies that, under the richness condition, $\text{ISR}$ is the strongest robust solution concept for dynamic Bayesian games: it is robust, and any refinement of it is not.

**Proposition 1.11** (Proposition 3.1). For any $\Theta$, $\text{ISR}$ is upper hemicontinuous on the universal type space $\mathcal{T}_\Theta^*$.

In a dynamic game, the details of the modeling assumptions on agents’ information about $\Theta^*$ are crucial. The next example shows how the distinction between knowledge and certainty may affect the predictions of a solution concept, raising novel questions of robustness.

**Example 1.9.** Consider the game in figure 1.4, and assume that $\Theta^* = \{-3, 3\}$. Let’s consider two alternative setups.

- **[A]: Player 2 observes the realization of $\theta$, and this is common knowledge.** (This situation can be modelled letting $\Theta^*_0$ and $\Theta^*_1$ be singletons, and
\( \Theta^*_2 = \{-3, 3\} \). Suppose that agents are commonly certain that \( \theta = 3 \), denoted by \( t^A = (t^A_1, t^A_2) \in T^*_A \) if type \( t^A_2 \) player 2 knows that \( \theta = 3 \), and puts probability one on player 1 being type \( t^A_1 \); if type \( t^A_1 \), agent 1 puts probability one on \( \theta = 3 \) and agent 2 being type \( t^A_2 \). (Notice that \( t^A \) is not “common knowledge of \( \theta = 3 \)” because \( t^A_1 \) does not know that \( \theta = 3 \).) Since 1 knows that 2 knows the true state of \( \theta \), if 1 is rational and believes that 2 is rational, he must play \( \text{Out} \): if 2 is rational and knows \( \theta \), 1 obtains \(-3 \) in the subgame irrespective of the realization of \( \theta \).

- \([B]\): Players observe nothing about \( \theta \), and this is common knowledge.
  (This situation can be modelled letting \( \Theta^*_1 \) and \( \Theta^*_2 \) be singletons, and \( \Theta^*_0 = \{-3, 3\} \).) Now, let’s maintain that agents share common certainty of \( \theta = 3 \), denoted by \( t^B = (t^B_1, t^B_2) \in T^*_B \): each type \( t^B_i \) \( (i = 1, 2) \) puts probability one on \( \theta = 3 \) and the opponent being type \( t^B_j \) \( (j \neq i) \). Hence, agents in this setup share the same hierarchies as types \( t^A \). If player 1 believes that \( \theta = 3 \), and believes that player 2 plays \( D \), then 1’s optimal response is to play \( \text{Out} \). If 2 believes this, then his information set is unexpected: If called to move, player 2 would be surprised and, if he revises his beliefs in favor of \( \theta = -3 \), he may play \( U \). If 1 anticipates this, then playing \( \text{In} \) is optimal even if 1 believes that \( \theta = 3 \). That is because 1 knows that 2, although certain of \( \theta = 3 \), does not know that \( \theta = 3 \). So, types \( t^B \) share the same hierarchies as types \( t^A \), but under this specification of agents’ information, \( \mathcal{ISR} \quad (t^B) = \{ \text{In, Out} \} \times \{ U, D \} \) (while \( \mathcal{ISR} \quad (t^A) = \{ (\text{Out, D}) \} \)).

Hence, in a dynamic game, the details of agents’ information about \( \Theta^* \) may crucially affect our predictions. So, for instance, types \( \Theta^* \) in example 1.8 share the
same hierarchies of beliefs as types $t^{CK}$ in example 1.7, but agents have different information about $\theta$ in the two examples. Yet, in that case $\mathcal{ISR}(t^*) = \mathcal{ISR}(t^{CK})$. Then, why is it the case that $\mathcal{ISR}(t^{CK}) = \mathcal{ISR}(t^*)$ while $\mathcal{ISR}(t^A) \neq \mathcal{ISR}(t^B)$?

What if we change agents’ information in example 1.8? How do differences in agents’ information affect the properties of “robustness” with respect to perturbations of agents’ hierarchies of beliefs?

**Robustness(-es): Type Spaces, Models and Invariance.** Modeling a strategic situation with incomplete information requires a specification of agents’ entire hierarchies of beliefs. A “complete model” includes a description of all possible hierarchies of beliefs agents may have over a “rich” space of uncertainty, $\Theta^*$; then, a solution concept assigns to each hierarchy of beliefs the set of strategies that the agent might play (possibly a singleton). In practice, it is common to select a small subset of all the possible hierarchies of beliefs to focus on. Then, the “robustness” of our predictions depends on properties of continuity of the solution concept correspondence on the space of all possible hierarchies of beliefs. Upper hemicontinuity in the *universal model* $\mathcal{T}_\Theta^*$ addresses precisely this robustness problem, and we have seen that Propositions 1.10 and 1.11 imply that $\mathcal{ISR}$ is a “strongest robust” solution concept in this sense.

When applied theorists choose a subset of $\Theta^*$-hierarchies to focus on, they typically represent them by means of (non universal) $\Theta^*$-based type spaces (definition 1.2) rather than elements of $\mathcal{T}_\Theta^*$. Representing a hierarchy as a type in a (non-universal) type space $\mathcal{T}_{\Theta^*}$ rather than an element of $\mathcal{T}_{\Theta^*}$ does not change the common knowledge assumptions on the *information structure* $\langle \Theta^*, (u_i)_{i \in N} \rangle$, but it does impose common knowledge assumptions on players’ hierarchies of beliefs, and their correlation with
the states of nature \( \theta_0 \). A solution concept is type space-invariant if the behavior prescribed for a given hierarchy does not depend on whether it is represented as an element of \( T^*_0 \) or of a different \( T^*_\theta \). Thus, type space-invariance is also a robustness property: Robustness, with respect to the introduction of the extra common knowledge assumptions on players’ beliefs (and their correlations with \( \theta_0 \)) imposed by non-universal type spaces.

**Proposition 1.12** (Proposition 3.2). \( ISR \) is type space-invariant under all information structures.

The intuition behind the type space-invariance of \( ISR \) is the same as in example 1.5 for \( ICR \): since \( ISR \) allows all possible correlations in agents’ conjectures about the state of nature and the opponents’ strategies, the possibility of redundant types to serve as correlating devices does not expand the set of feasible conjectures. Hence, \( ISR \) only depends on the hierarchies of beliefs, not on the type space.

In writing down a game, as analysts we typically make common knowledge assumptions not only on players’ beliefs, but also on payoffs. For instance, suppose that all types in \( T^*_0 \) have beliefs concentrated on some strict subset \( \Theta \subset \Theta^* \), i.e. there is common certainty of \( \Theta \) in \( T^*_0 \). In applied work, states that receive zero probability (\( \theta \in \Theta^* \setminus \Theta \)) are usually excluded from the model. That is: common certainty of \( \Theta \) is made into common knowledge of \( \Theta \).

**Definition 1.7.** A model of the environment is a \( \Theta \)-based type space, where \( \Theta \) is such that \( \Theta_k \subseteq \Theta^*_k \) for each \( k = 0, ..., n \).

Each type in a model induces a \( \Theta \)-hierarchy, and hence a \( \Theta^* \)-hierarchy. A solution concept is model invariant if the behavior is completely determined by the \( \Theta^* \)-hierarchies irrespective of the model they are represented in. Model invariance is
a stronger robustness property than type space-invariance, as it also requires robustness to the introduction of extra common knowledge assumptions on the information structure.

**Example 1.10.** Consider example 1.9 again, and suppose that the “true” situation is the one described by case B, in which agents share common certainty that $\theta = 3$, but they have no information about $\theta$. As argued, for these types the solution concept delivers $\mathcal{ISR}(t^B) = \{\text{In}, \text{Out}\} \times \{U, D\}$. If that situation is modelled as there being common knowledge of $\theta = 3$, then the prediction of $\mathcal{ISR}$ is the backward induction outcome $\mathcal{ISR}(\tilde{t}^{CK}) = (\text{Out}, D)$. Hence, despite $\tilde{t}^{CK}$ and $t^B$ share the same hierarchies of beliefs, $\mathcal{ISR}$ delivers different predictions depending on the model in which those hierarchies are represented.

Hence, in environments with no-information (NI-environments), $\mathcal{ISR}$ is not model-invariant. The next proposition instead shows that $\mathcal{ISR}$ is model-invariant in environments with private-values, or PV-environments (That is why $\mathcal{ISR}(t^{CK}) = \mathcal{ISR}(t^*)$ in example 1.8.):

**Proposition 1.13** (Proposition 3.3). In environments with private values, $\mathcal{ISR}$ is model-invariant.

**On NI- and PV-settings.** In NI-settings, all common knowledge-assumptions are relaxed. In particular, the assumption that players know their own preferences: in these environments, agents don’t know their own preferences over the terminal nodes, they merely have beliefs about that. Under an equivalent richness condition, independent work by Chen (2009) studied the structure of $\mathcal{ICR}$ for dynamic NI-environments, extending Weinstein and Yildiz’s (2007) results. Together with the upper hemicontinuity of $\mathcal{ISR}$, Chen’s results imply that the two solution concepts coincide on the
universal model in these settings. Outside of the realm of NI-environments, \( \mathcal{ISR} \) generally refines \( ICR \) (imposing \textit{sequential rationality} restrictions). The fact that, under the richness condition, \( \mathcal{ISR} \) coincides with \( ICR \) in NI-environments, implies that \textit{sequential rationality} has no bite in these settings. The intuition is simple: In NI-environments players don’t know their own payoffs, they merely have beliefs about them. Once an unexpected information set is reached, Bayes’ rule does not restrict players’ beliefs, which can be set arbitrarily. Under the richness conditions, there are essentially no restrictions on players’ beliefs about their own preferences, so that any behavior can be justified. Hence, the only restrictions that retain their bite are those imposed by (normal form) rationality alone. This also provides the main intuition for \( \mathcal{ISR} \)’s failure of model invariance in NI-settings (1.10).

To the extent that the interest in studying extensive form games comes from the notion of sequential rationality, PV-settings, in which the assumption that \textit{players know their own payoffs} (and this is common knowledge) is maintained, seem most meaningful for dynamic environments. As shown by proposition 1.13, \( \mathcal{ISR} \) is \textit{model invariant} in these settings.\textsuperscript{27}

\textbf{On the Impact of higher order beliefs on multiplicity.} A growing literature is exploring to what extent the main insights from the theory of global games can be extended to dynamic environments. These contributions are mainly from an applied perspective, and do not pursue a systematic analysis of these problems. Consequently, the conclusions are diverse: for instance, Chamley (1999) and Frankel

\textsuperscript{27}Another closely related paper is Dekel and Fudenberg (1990): In that paper, solution concept \( S^{\infty}W \) is shown to be “robust” to the possibility that players entertain small doubts about their opponents’ payoff functions. The robustness result for \( \mathcal{ISR} \) is in the same spirit. Dekel and Fudenberg (1990) maintain the assumption that players \textit{know} their own payoffs: This corresponds to the PV-case in this paper. In Appendix A.3 it is shown that, in PV-settings, \( \mathcal{ISR} \) coincides with \( S^{\infty}W \) applied to the interim normal form.
and Pauzner (2000) obtain the familiar uniqueness result in different setups, under different sets of assumptions. On the other hand, few recent papers seem to question the general validity of these results: For instance, Angeletos, Hellwig and Pavan (2007) and Angeletos and Werning (2006) apply the global games’ signals structure to dynamic environments, and obtain non-uniqueness results that contrast with the familiar ones in static settings. The origin of such multiplicity lies in a tension between the global games’ signals structure and the dynamic structure: by relaxing common knowledge-assumptions, the former favors uniqueness; in dynamic games, some information endogenously becomes common knowledge (e.g. public histories), thus mitigating the impact of the signals structure.

The following result shows that also in dynamic games multiplicity can be imputed to common knowledge assumptions: once CK-assumptions are relaxed, ISR generically delivers uniqueness.

**Proposition 1.14** (Proposition 3.5). If $\Theta^*$ is “rich”, the set of types for which ISR is unique is open and dense in $T_{0^*}$.

### 1.5.2 Equilibria in Belief-free environments

In this section we explore the scope of the equilibrium hypothesis in dynamic games, when no assumptions on beliefs are imposed. The main goal is therefore a result analogous to Proposition 1.9, p. 32: we want to characterize the predictions of the equilibrium approach in dynamic games that do not depend on the formulation of a specific type space.

In dynamic Bayesian games, equilibrium concepts typically involve two assumptions about agents’ rationality: first, that agents are sequentially rational; second, that they hold consistent beliefs. That is, at each point in the game agents’ con-
ditional beliefs about the state of nature and the opponents’ types are obtained via Bayesian updating from the agent’s prior beliefs and the equilibrium strategy profiles. Several notions of “Perfect Bayesian Equilibrium” (PBE) have been put forward in the literature (see e.g. Section 8.2 in Fudenberg and Tirole, 1991b). Here we focus on the weakest possible version of PBE, that maintains the assumption of sequential rationality at all histories, and Bayesian updating whenever possible: this delivers the solution concept of interim perfect equilibrium (IPE), from Chapter 4.28

**Interim Perfect Equilibrium.**

An IPE is defined as an assessment (i.e. a strategy profile and a belief system) that satisfies sequential rationality and consistency of beliefs. Formally, given a dynamic Bayesian game \((\Gamma, T)\), a system of beliefs consists of collections \((p_{t_i}(h))_{h \in \mathcal{H}}\) for every type \(t_i\) of every agent \(i\), such that \(p_{t_i}(h) \in \Delta(\Theta_0 \times T_{-i})\) for every \(h \in \mathcal{H}\). A strategy profile and a belief system \((\sigma, p)\) form an assessment. For each agent \(i\), a strategy profile \(\sigma\) and conditional beliefs \(p_{t_i}\) induce, at each private history \(h_i\), a probability measure \(P^{\sigma,p_{t_i}}(h)\) over the histories \(h'\) following \(h\).

**Definition 1.8.** Fix a strategy profile \(\sigma \in \Sigma\). Assessment \((\sigma, p)\) is consistent if for each \(i \in N\) and \(t_i \in T_i\), (BC1) \(p_{t_i}(\phi) = \tau_i(t_i)\) and if (BC2) for each \(h'\) such that \(h \prec h'\) (i.e. \(h\) precedes \(h'\)), \(p_{t_i}(h')\) is obtained from \(p_{t_i}(h)\) and \(P^{\sigma,p_{t_i}}(h)\) via Bayesian updating (whenever possible).

Condition (BC1) requires that each type’s beliefs at the beginning of the game agree with those specified in the type space; condition (BC2) requires that the belief system \(p_{t_i}\) is consistent with Bayesian updating whenever possible. Also, notice that

28The environment in Chapter 4 is one of distributed knowledge, i.e. \(\Theta_0\) is a singleton. Chapter 4 also allows for the possibility that agents learn their payoff-type over time.
\( P^\sigma,p_i(h) \) is obtained taking both \( \sigma_i \) and \( \sigma_{-i} \) into account, hence also “own-deviations” are zero probability events, and leave an agent’s beliefs at the subsequent history unrestricted.\(^{29}\) This point is discussed below.

The notion of sequential rationality is completely standard: Given beliefs \( p, (\sigma, p) \) is *sequentially rational* if for every agent \( i \), every type \( t_i \in T_i \), and for every history \( h \in \mathcal{H} \), behavior \( \sigma_i(t_i)(h) \) is a best response to \( \sigma_{-i} \) at \( h \), given beliefs \( p_{t_i}(h) \).

**Definition 1.9.** An assessment \( (\sigma, p) \) is an Interim Perfect Equilibrium (IPE) if it is consistent and sequentially rational.

Notice that the notion of consistency in definition 1.8 imposes no restrictions on the beliefs held at histories that receive zero probability at the preceding node. Hence, even if agents’ initial beliefs admit a common prior, IPE is weaker than Fudenberg and Tirole’s (1991a) *perfect Bayesian equilibrium* (hence also of Kreps and Wilson’s (1982) *sequential equilibrium*). Also, notice that from the point of view of the belief system, any player’s deviation is a zero probability event, and treated the same way. In particular, if history \( h' \) is precluded by \( \sigma_i(h) \) alone (for \( h \prec h' \)), \( h' \notin \text{supp} P^\sigma,p_i(h) \), and agent \( i \)'s beliefs at \( h' \) are unrestricted the same way they would be after an unexpected move of the opponents: it is *as if* player \( i \) is surprised by his own deviation, and after such a “surprise” his beliefs about \( \Theta_0 \times T_{-i} \) may be upset.\(^{30}\) This feature of IPE is not standard, but it is key to some of the results in Chapter 4. This point will be further discussed below.

\(^{29}\)Penta (2009a) considers a more standard notion of equilibrium in which unilateral “own-deviations” to not trigger this change in beliefs: i.e., the belief system is obtained considering the probability distribution induced by \( \sigma_{-i} \) and \( p_i \), written \( P^{\sigma_{-i},p_i}(h) \).

\(^{30}\)IPE is consistent with a “trembling-hand” view of unexpected moves, in which no restrictions on the possible correlations between “trembles” and other elements of uncertainty are imposed. Unlike other notions of *weak perfect Bayesian equilibrium* (e.g. definition 9.C.3, p. 285 in Mas-Colell, Whinston and Green, 1995), in IPE agents’ beliefs are consistent with Bayesian updating also off-the-equilibrium path.
Finally, notice the following:

**Remark 1.1.** *In complete information games, IPE coincides with subgame perfect equilibrium (SPE).*

**Backwards Rationalizability.**

As SPE refines Nash equilibrium in dynamic games, so IPE refines Bayesian equilibrium (it also generalizes SPE to incomplete information games the same way Bayesian equilibrium generalizes Nash equilibrium). Similarly, for a result analogous to Proposition 1.9, a good candidate would be some dynamic version of Δ-rationalizability (definition 1.6), i.e. some dynamic version of the idea of *common belief in rationality (RCBR)*. The solution concept that answered the robustness questions addressed in Section 1.5.1 (*ISR*) represents the weakest “dynamic extension” of RCBR: that is, sequential rationality takes the place of (ex ante) rationality, but *common certainty* of it is only assumed *at the beginning of the game*; whenever agents are surprised, they may believe that their opponents are irrational. *Initial common certainty of rationality* is not the only possible “dynamic extension” of the idea of RCBR:

**Example 1.11.** *Consider the complete information game in figure 1.5: Strategy rD*
is dominated by \( l \). Thus, if at the beginning of the game player 2 thinks that 1 is rational, he assigns zero probability to \( rD \) being played. For example, 2 could assign probability one to \( l \), so that the next information set is unexpected.

If we adopt a solution concept based on (only) initial common certainty of rationality, like ISR for instance (Section 1.5.1), then in the proper subgame nothing is ruled out for player 2: If 2 puts probability one on \( l \), the subgame is unexpected, and 2 is allowed to believe that 1 could play anything in the subgame (including \( D \)). Hence, 2’s best response could be either \( L \), \( C \) or \( R \). Given this, both \( rU \) and \( rM \) are “justifiable” for a rational player 1, as well as strategies \( lU \), \( lM \) and \( lD \). Hence, under initial common certainty in rationality, reduced strategy profiles \( \{rU,rM,l\} \times \{L,C,R\} \) are selected. (ISR is a solution concept defined for reduced-form strategies, \( l \) denotes the equivalence class of strategies \( lU \), \( lM \) and \( lD \), which are all outcome equivalent.)

Now, besides initial common certainty of rationality, consider the following assumptions:

\[ [H-1] \text{ Agents believe that the opponent is rational even after unexpected moves, AND after such moves:} \]

\[ [H-1.1] \text{ They believe that the opponent is rational and, if possible, that he was rational in the past; OR} \]

\[ [H-1.2] \text{ They accept that the opponent may have played irrationally in the past, but will be rational in the future.} \]

Suppose that \([H-1]\) is true, and this is common certainty: then, after observing \( r \), player 1 would still put zero probability on \( D \), hence he would never play \( R \) if rational. If 1 anticipates this, then \( rU \) is dominated by \( l \), because the highest payoff from \( rU \) if player 2 does not play \( R \) is 2.
If [H-1.1] is true, upon observing \( r \) player 2 would assign zero probability to \( U \): the only way playing \( r \) would be rational for 1, is that 1 expects 2 to play \( C \), which means that 1 is playing strategy \( rM \). Given this, 2 plays \( C \). If 1 anticipates this, his unique best response is in fact to play \( rM \). This reasoning corresponds to Pearce’s (1984) Extensive Form Rationalizability (cf. example 1.7 above), and selects profile \((rM, C)\) as the only outcome in this game.

Suppose that [H-1.2] is true instead: now, if 2 observes the unexpected move \( r \), he may believe that 1 has “made a mistake”. Then, from the observation of \( r \) player 2 does not infer that 1 expects 2 to play \( C \), hence he also does not infer that 1 would play \( M \). Hence, if agents are willing to accept that the opponents have played irrationally in the past, but they share common certainty of future rationality at every history, then the strategy profiles selected in this game are \( \{rU, rM, lU, lM\} \times \{L, C\} \).

Chapter 4 introduces a solution concept for dynamic games with payoff uncertainty (such as \( \Gamma \), i.e. not Bayesian games) based on the idea of rationality and common certainty of future rationality at every history, called Backwards Rationalizability (\( BR \)).\(^{31}\)

\( BR \) consists of an iterated deletion procedure similar to \( \Delta \)-rationalizability. One important difference is that conjectures here are represented by Conditional Probability Systems \( \mu^i = (\mu^i(h))_{h \in H} \in \Delta^{H_i}(\Theta_{-i} \times S) \), i.e. collections of conditional probability distributions, one for each history, such that \( \mu^i(h) \in \Delta(\Theta_{-i} \times S) \) for every \( h \), and that satisfy two conditions: (C.1) agents are certain of what they know: \( i \)’s conjectures are concentrated on \( \Theta_{-i} \times S(h) \) at history \( h \); (C.2) conditional probabilities are consistent with Bayesian updating (whenever possible.)

\(^{31}\)Perea (2010) independently introduced a solution concept for complete information games, Common Belief in Future Rationality, that coincides with \( BR \) in this class of games.
Notice that in this specification agents entertain conjectures about the payoff state, the opponents’ as well as their own strategies. Nonetheless, it is maintained that an agent is always certain of his own strategy (hence the notion of sequential rationality does not change.) This points will be discussed below.

$BR_i$ is defined as follows: At each round, strategy $\hat{s}_i$ survives if it is justified by conjectures $\mu^i$ that satisfy two conditions: condition (A) states that at the beginning of the game, the agent must be certain of his own strategy $\hat{s}_i$ and have conjectures concentrated on opponents’ strategies that survived the previous rounds of deletion; condition (B) restricts the agent’s conjectures at unexpected histories, and it’s made of two parts: condition (B.1) states that agent $i$ is always certain of his own continuation strategy, $\hat{s}_i|h$; condition (B.2) requires conjectures to be concentrated on opponents’ continuation strategies $s_{-i}|h$ that are consistent with the previous rounds of deletion.\footnote{That is: if $(\theta_{-i}, s_{-i})$ is in the support of the conjectures $\mu^i(h)$ at round $k$, then there exist $\hat{s}_{-i} \in BR_{-i}^{k-1}(\theta_{-i})$ s.t. $\hat{s}_{-i}|h = s_{-i}|h$.} However, at unexpected histories, agents’ conjectures about $\Theta_{-i}$ are unrestricted. Thus, condition (B) embeds two conceptually distinct kinds of assumptions: one concerning agents’ conjectures about $\Theta_{-i}$; the other concerning their conjectures about the continuation behavior. For ease of reference, they are summarized as follows:

- **Unrestricted-Inference Assumption (UIA):** At unexpected histories, agents’ conjectures about $\Theta_{-i}$ are unrestricted. In particular, agents are free to infer anything about the opponents’ private information from the public history $h$.

For example, conditional conjectures may be such that $\text{marg}_{\Theta_{-i}}\mu^i(h)$ is concentrated on a payoff-type $\theta_{-i}$ for which some of the previous moves in $h$ are irrational.
Nonetheless, condition (B.2) implies that it is believed that $\theta_{-i}$ will behave rationally in the future. From an epistemic viewpoint, it can be shown that $\mathcal{BR}$ can be interpreted as common certainty of future rationality at every history.

- **Common Certainty in Future Rationality (CCFR):** at every history (expected or not), agents share common certainty in future rationality.

**Results**

The next result provides the analogue of Proposition 1.9, and shows that $\mathcal{BR}$ characterizes the set of IPE-strategies across models of beliefs:

**Proposition 1.15** (Proposition 4.1). Fix a game $\Gamma$. For each $i$: $\hat{s}_i \in \mathcal{BR}_i$ if and only if there exists a type space $T$ and a type $t_i \in T_i$ and $(\hat{\sigma}, \hat{p})$ such that: (i) $(\hat{\sigma}, \hat{p})$ is an IPE of $(\Gamma, T)$ and (ii) $\hat{s}_i \in \text{supp} \hat{\sigma_i}(t_i)$.

As emphasized above, in $\mathcal{BR}$ agents hold conjectures about both the opponents’ and their own strategies. First, notice that conditions (A) and (B.2) in the definition of $\mathcal{BR}$ maintain that agents are always certain of their own strategy; furthermore, the definition of sequential best response depends only on the marginals of the conditional conjectures over $\Theta_{-i} \times S_{-i}$. Hence, this particular feature of $\mathcal{BR}$ does not affect the standard notion of rationality. The fact that conjectures are elements of $\Delta^H(\Theta_{-i} \times S)$ rather than $\Delta^H(\Theta_{-i} \times S_{-i})$ corresponds to the assumption, discussed above, that IPE treats all deviations the same; its implication is that both histories arising from unexpected moves of the opponents and from one’s own deviations represent zero-probability events, allowing the same set of conditional conjectures about $\Theta_{-i} \times S$, with essentially the same freedom that IPE allows after anyone’s deviation. Hence Proposition 1.15.
Backwards Procedure and Backward Induction Reasoning. Condition (B.1), hence CCFR, can be interpreted as a condition of belief persistence on the continuation strategies: Both at expected and unexpected histories, agents’ possible conjectures about the opponents’ continuation strategies are unchanged. In games of complete information, an instance of the same principle is provided by subgame perfection, where agents believe in the equilibrium continuation strategies both on- and off-the-equilibrium path. The belief persistence hypothesis goes hand in hand with the logic of backward induction, allowing to envision each subgame “in isolation”: SPE is defined as a Nash equilibrium that induces a Nash equilibrium in each subgame. Hence, SPE-implications in a subgame can be drawn looking at the subgame in isolation. This “modular” structure abides Harsanyi’s and Selten (1988) requirement of “subgame consistency”, according to which ‘the behavior prescribed on a subgame is nothing else than the solution of the subgame itself’ (ibid., p.90).

Assumptions UIA and CCFR extend these ideas to incomplete information environments: under UIA, the set of beliefs agents are allowed to entertain about the opponents’ payoff types is the same at every history: $\Theta_{-i}$. Hence, in this respect, their information about the opponents’ payoff-types in the subform starting from (public) history $h$ is the same as if the game started from $h$. Also, CCFR implies that agents’ epistemic assumptions about everyone’s behavior in the continuation is the same that would hold if the game started at $h$: namely, CCFR.

Thus, UIA and CCFR combined imply that, from the point of view of BR, a continuation from history $h$ is equivalent to a game with the same space of uncertainty and strategy spaces equal to the continuation strategies, which justifies the possibility of analyzing continuation games “in isolation”. This discussion provides the main insights for the following result:
\( \Theta = \Theta_1 = \{10, -10\} \)

\[
\begin{array}{c|c|c|c|c|c|c|}
 & \text{a}_1 & \text{a}_2 & \text{a}_3 \\
\hline
L_1 & 2; 1 & 1; 1 + \theta_1 & 0; -1 \\
L_2 & 0; 1 & 2; 1 + \theta_1 & 1; -1 \\
L_3 & -1; 2 & -1; 1 & -1; 3 \\
\end{array}
\]

\[
\begin{array}{c|c|c|c|c|c|c|}
 & \text{b}_1 & \text{b}_2 & \text{b}_3 \\
\hline
R_1 & 2 - \theta_1; -1 & 1 - \theta_1; 1 & -\theta_1; 1 + \theta_1 \\
R_2 & -\theta_1; -1 & 2 - \theta_1; 1 & 1 - \theta_1; 1 + \theta_1 \\
R_3 & -1 - \theta_1; 3 & -1 - \theta_1; 0 & -1 - \theta_1; 2 \\
\end{array}
\]

Figure 1.6: Example 1.12

**Proposition 1.16** (Proposition 4.2). \( \mathcal{BR} \) can be computed as follows: for every almost-terminal node \( h \), apply \( \Delta \)-rationalizability in the continuation game; then, proceed backwards applying \( \Delta \)-rationalizability to the normal form of the subgame, with strategies restricted to be \( \Delta \)-rationalizable in the corresponding continuations. Proceed backwards, until the initial node is reached (Denote this backwards procedure by \( \mathcal{R}^\phi \)).

To illustrate the solution concept, consider the following example:

**Example 1.12.** Consider the game in figure 1.6, and let \( \Theta_1 = \{10, -10\} \), while \( \Theta_0 \) and \( \Theta_2 \) are singletons (hence, player 1 knows the true state, while player 2 has no information). Notice that playing \( l \) (resp. \( r \)) is dominated for payoff-type \( \theta_1 = -10 \) (resp. \( \theta_1 = 10 \)). Hence, if we apply the logic of hypothesis \( [H - 1.1] \), that is Battigalli and Siniscalchi (2007) Strong (Extensive Form) Rationalizability, whatever 2’s prior beliefs about \( \theta_1 \), he would assign probability one to \( \theta_1 = 10 \) if he observes \( l \), and probability one to \( \theta_1 = -10 \) if he observes \( r \). Given this, and the fact that rational player 1 would never play \( L_3 \) or \( R_3 \) in the continuations, the only “rationalizable” strategy for 2 is to play \( a_2 b_2 \). Given this, only strategy \( lL_2 R_2 \) survives for type \( \theta_1 = 10 \),
and only strategy $rL_2R_2$ survives for type $\theta_1 = -10$. (Notice that hypothesis $[H - 1.1]$ incorporates a forward induction logic)

Now, let’s apply $BR$ instead: at the first round, strategies involving $L_3$ and $R_3$ would be deleted for all types of player 1, while strategies involving $l$ (resp. $r$) would be deleted for type $\theta_1 = -10$ (resp. $\theta_1 = 10$). So, for instance, after the first round strategy $rL_1R_2$ survives for type $\theta_1 = -10$. Now, suppose that 2’s initial conjectures are concentrated on payoff state-strategy pair $(\theta_1, s_1) = (-10, rL_1R_2)$: then, $l$ is unexpected, so after observing it we are free to specify 2’s beliefs, who could for example assign probability one on pair $(\theta_1, s_1) = (10, lL_1R_2)$, hence play $a_2$, or on pair $(\theta_1, s_1) = (1, lL_1R_2)$, hence play $a_1$. Similarly, both $b_1$ and $b_2$ in the right-most continuation game can be justified if 2’s prior assigns probability one to $\theta_1 = 10$. Hence, strategies that survive $BR$ are $\{a_1b_2, a_1b_3, a_2b_2, a_2b_3\}$ for player 2, $\{lL_1R_1, lL_1R_2, lL_2R_1, lL_2R_2\}$ for type $\theta_1 = 10$ and $\{rL_1R_1, rL_1R_2, rL_2R_1, rL_2R_2\}$ for type $\theta_1 = -10$.

Notice that this solution is precisely that delivered by the backwards procedure: if we apply $\Delta$-rationalizability to the continuation game following $r$, $R_3$ is deleted at the first round for both types of player 1, and $b_1$ at the second round for player 2. In this continuation game, the procedure selects continuations $\{b_2, b_3\}$ for player 2 and continuations $\{R_1, R_2\}$ for both types of player 1. Similarly, after $l$, only $\{a_1, a_2\}$ and $\{L_1, L_2\}$ for both types survive $\Delta$-rationalizability. So, now we apply $\Delta$-rationalizability with the normal form in which it is maintained that continuation strategies are rationalizable in the corresponding continuations, i.e. the relevant strategy sets now are $\{a_1b_2, a_1b_3, a_2b_2, a_2b_3\}$ for player 2, and

$$\{lL_1R_1, lL_1R_2, lL_2R_1, lL_2R_2\} \cup \{rL_1R_1, rL_1R_2, rL_2R_1, rL_2R_2\}$$

for (both types of) player 1: at this stage, type $\theta_1 = 10$ deletes all strategies involving
r at the first round, and so does type \( \theta_1 = -10 \) with those involving \( l \), but player 2
doesn’t delete anything: if he expects \( \theta_1 = 10 \) (hence 1 to play \( l \)), then both \( a_2b_2 \) and
\( a_2b_3 \) are best responses in the normal form; similarly, if he expects \( \theta_1 = 10 \) (i.e. 1 to
play \( r \)), then both \( a_2b_3 \) and \( a_1b_3 \) are optimal.

**Further Comments:** A result analogous to Proposition 1.15 can be obtained
for the more standard refinement of IPE mentioned in footnote 29, in which uni-
lateral own deviations leave an agents’ beliefs unchanged, applying to a modified
version of \( \mathcal{BR} \): such modification entails assuming that agents only form conjectures
about \( \Theta_{-i} \times S_{-i} \) (that is: \( \mu^i \in \Delta^\mathcal{H}(\Theta_{-i} \times S_{-i}) \), and consequently adapting conditions
(A) and (B) in the definition of \( \mathcal{BR} \) (See Penta, 2009a): The assumption that IPE
treats anyone’s deviation the same (and, correspondingly, that in \( \mathcal{BR} \) agents hold
conjectures about their own strategy as well) is not crucial to characterize the set
of equilibrium strategies for all models of beliefs. It is crucial instead for the result
of Proposition 1.16, which shows that such set can be computed applying a proce-
dure that extends the logic of *backward induction* to environments with incomplete
information: Treating own deviations the same as the opponents’ is key to the poss-
sibility of considering continuation games “in isolation”, necessary for the result. In
the alternative specification of IPE (in Penta 2009a), the “backwards procedure” to
compute the set of equilibria must be modified, so to keep track of the restrictions
the extensive form imposes on the agents’ beliefs at unexpected nodes. Losing the
possibility of envisioning continuation games “in isolation”, the modified procedure
is more complicated.
Backwards procedure, Subgame-Perfect Equilibrium and IPE.

In games with complete and perfect information, the “backwards procedure” coincides with the backward induction solution, hence with subgame perfection.\textsuperscript{33} The next example (borrowed from Perea, 2010) shows that if the game has complete but imperfect information, strategies played in Subgame-Perfect Equilibrium (SPE) may be a strict subset of $R^\phi$:

Example 1.13. Consider the complete information game in figure 1.7. The backward procedure delivers the following sets of strategies for the two agents: $R_1^\phi = \{bc, bd, ac\}$ and $R_2^\phi = \{f, g\}$. The game though has only one SPE, in which player 1 chooses $b$: in the proper subgame, the only Nash equilibrium entails the mixed (continuation) strategies $\frac{1}{2}c + \frac{1}{2}d$ and $\frac{3}{4}f + \frac{1}{4}g$, yielding a continuation payoff of $\frac{11}{4}$ for player 1. Hence, player 1 chooses $b$ at the first node.

In games with complete information, IPE coincides with SPE (remark 1.1, p. 50), but $R^\phi$ in general is weaker than subgame perfection. At first glance, this may appear

\textsuperscript{33}For the special case of games with complete information, Perea (2010) independently introduced an equivalent procedure and showed that $R^\phi$ coincides with the backward induction solution if the game has perfect information.
in contradiction with propositions 1.15 and 1.16, which imply that $R^\phi$ characterizes
the set of strategies played in IPE across models of beliefs. The reason is that even if
the environment has no payoff uncertainty ($\Theta$ is a singleton), the complete informa-
tion model in which $T_i$ is a singleton for every $i$ is not the only possible: models with
*redundant types* may exist, for which IPE strategies differ from the SPE-strategies
played in the complete information model. The source of the discrepancy is analo-
gous to the one between Nash equilibrium and subjective correlated equilibrium (see
examples 1.2 and 1.4). We illustrate the point constructing a type space and an IPE
in which strategy $(ac)$ is played with positive probability by some type of player 1.\(^{34}\)

Let payoffs be the same as in example 1.13, and consider the type space $T$ such that
$T_1 = \{t_{bc}^1, t_{bd}^1, t_{ac}^1\}$ and $T_2 = \{t_2^f, t_2^g\}$, with the following beliefs:

$$
\tau_1 (t_1) \left[ t_2^f \right] = \begin{cases} 
1 & \text{if } t_1 = t_{bc}^1, t_{ac}^1 \\
0 & \text{otherwise}
\end{cases}
$$

and

$$
\tau_2 (t_2^g) \left[ t_1^{ag} \right] = 1, \beta_2 \left( t_2^f \right) \left[ t_1^{bc} \right] = 1
$$

The equilibrium strategy profile $\sigma$ is such that $\forall i, \forall t_i, \sigma_i (t_i^s) = s_i$. The system of
beliefs agrees with the model’s beliefs at the initial history, hence the beliefs of types
$t_2^g$ and $t_1^{ac}$ are uniquely determined by Bayesian updating. For types $t_i^s \neq t_2^g, t_1^{ac}$,
it is sufficient to set $p_i (t_i^s, a_i) = \tau_i (t_i^s)$ (i.e. maintain whatever the beliefs at the
beginning of the game were) Then, it is easy to verify that $(\sigma, p)$ is an IPE.

On the other hand, if $|\Theta^*| = 1$ and the game has *perfect* information (no stage
with simultaneous moves), then $R^\phi$ coincides with the set of SPE-strategies. Hence,
in environments with no payoff uncertainty and with perfect information, only SPE-

---

\(^{34}\) It is easy to see that such difference is not merely due to the possibility of zero-probability types. Also the relaxation of the common prior assumption is not crucial.
strategies are played in IPE for any model of beliefs.
Chapter 2

On the Structure of Rationalizability on Arbitrary Spaces of Uncertainty

Abstract: This note characterizes the set \( \mathcal{A}_i^\infty \) of actions of player \( i \) that are uniquely rationalizable for some hierarchy of beliefs on an arbitrary space of uncertainty \( \Theta \). It is proved that for any rationalizable action \( a_i \) for the type \( t_i \), if \( a_i \) belongs to \( \mathcal{A}_i^\infty \) and is justified by conjectures concentrated on \( \mathcal{A}_i^{\infty} \), then there exists a sequence of types converging to \( t_i \) for which \( a_i \) is uniquely rationalizable.

Keywords: Rationalizability, incomplete information, robustness, refinement, higher order beliefs, dominance solvability, richness.

JEL Codes: C72.

2.1 Introduction

This note characterizes the set \( \mathcal{A}_i^\infty \) of actions of player \( i \) that are uniquely rationalizable for some hierarchy of beliefs on an arbitrary space of uncertainty \( \Theta \). It is proved that for any rationalizable action \( a_i \) for the type \( t_i \), if \( a_i \) belongs to \( \mathcal{A}_i^\infty \) and is justified
by conjectures concentrated on $A_i^\infty$, then there exists a sequence of types converging to $t_i$ for which $a_i$ is uniquely rationalizable.

Assuming that $\Theta$ contains regions of strict dominance for each player’s strategy (the richness condition), Weinstein and Yildiz (2007) prove a version of the above result:

- **(Non-) Robustness (R.1):** whenever a model has multiple rationalizable outcomes, any of these is uniquely rationalizable in a model with beliefs arbitrarily close to the original ones.

Result (R.1) essentially corresponds to the case $A_i^\infty = A_i$ for each $i$.

An important implication of (R.1) is that, under the richness condition, the strongest predictions that are robust to perturbations of higher order beliefs are those based on rationalizability alone. In many situations though, imposing richness may be an unnecessarily demanding robustness test: The richness condition on $\Theta$ implies that it is not common knowledge that any strategy is not dominant. However, as modelers, we may wish to assume that some features of the environment actually are common knowledge. For example: common knowledge that some strategies are not dominant. In that case, the underlying space of uncertainty does not satisfy richness. The main purpose of this paper is to explore the structure of rationalizability without assuming richness.

By guaranteeing that $A_i^\infty = A_i$ for each $i$, the richness condition also delivers the following striking result (Weinstein and Yildiz, 2007):

- **Generic Uniqueness (R.2):** in the space of hierarchies of beliefs, the set of types with a unique rationalizable action is open and dense (i.e. models are generically dominance-solvable);
Result (R.2) generalizes an important insight from the global games literature, that multiplicity is often the consequence of the modeling assumptions of common knowledge.\(^1\) If such assumptions are relaxed (e.g. assuming richness), hierarchies of beliefs “typically” have a unique rationalizable outcome. The case of multiplicity corresponds to a knife-edge situation, at the boundary of regions of uniqueness for each of the rationalizable actions.

Multiplicity is pervasive in applied models. Yet, from a theoretical point of view, there is a sense in which a complete model should be able to deliver a unique prediction. By way of analogy, consider the dynamics of a coin toss: If all the information about the intervening forces, the mass and shape of the coin, air pressure and so on (the initial conditions) were available, according to Newtonian mechanics we could predict the outcome of the coin toss. The “practical” unpredictability of a coin toss is rather a consequence of the “imperfection” of our model for the initial conditions: indeterminacy does not pertain to the underlying phenomenon; rather, it stems from the modeling activity.\(^2\)

Result (R.2) can be interpreted as saying that the typical indeterminacy of standard game theoretical models does not pertain to the object of study; rather, it is a consequence of the simplifying assumptions that we make on higher order beliefs.

It will be shown that very weak conditions on \(\Theta\) suffice to obtain Weinstein and

\(^1\)Notice though that result (R.1) is in sharp contrast with that literature: In the global games’ approach, the relaxation of common knowledge assumptions supports the robust selection of a unique equilibrium. In contrast, (R.1) implies that if one considers a richer class of perturbations, any selection from rationalizability is not robust. See Morris and Shin (2003) for a thorough survey of the literature.

\(^2\)The last paragraph presumes that the underlying phenomenon, i.e. the object of the model, is not “intrinsically indeterminate”. It is not a statement that no such indeterminate objects exist. If one believes that the object of study is intrinsically indeterminate, then the statement should be rejected. A debate in philosophy of science is open on the issue.
Yildiz’s results in their full strength, without imposing the *richness condition*. For instance, it suffices to assume that there exists a state in $\Theta$ for which payoff functions are supermodular, plus dominance regions for the corresponding extreme actions only. In other words, if it is *not* common knowledge that the game is *not* supermodular, and that the corresponding extreme actions are *not* dominant, then the strongest predictions that are robust to perturbations of higher order beliefs are those based on rationalizability alone.

### 2.2 Game Theoretic Framework

I consider static games with payoff uncertainty, i.e. tuples $G = \langle N, \Theta, (A_i, u_i)_{i \in N} \rangle$ where $N$ is the set of players; for each $i \in N$, $A_i$ is the set of actions and $u_i : A \times \Theta \to \mathbb{R}$ is $i$’s payoff function, where $A := \times_{i \in N} A_i$ and $\Theta$ is a parameter space representing agents’ incomplete information about the payoffs of the game. Assume that the sets $N$, $A$ and $\Theta$ are all finite. As standard, hierarchies of beliefs can be represented by means of type spaces: a *type space* is a tuple $T = (T_i, \tau_i)_{i \in N}$ s.t. for each $i \in N$, $T_i$ is the (compact) set of *types* of player $i$, and the continuous function $\tau_i : T_i \to \Delta (\Theta \times T_{-i})$ assigns to each type of player $i$ his beliefs about $\Theta$ and the opponents’ types. Let $T_i^*$ be the set of all coherent hierarchies; we denote by $T^* = (T_i^*, \tau_i^*)_{i \in N}$ the $\Theta$-based universal type space (Mertens and Zamir, 1985). Elements of $T_i^*$ will be referred to as *types* or *hierarchies*. A type $t_i \in T_i^*$ is finite if $\tau_i^* (t_i) \in \Delta (\Theta \times T_{-i}^*)$ has finite support; the set of finite types in the universal type space is denoted by $\hat{T}_i \subseteq T_i^*$. The function $\pi_i^T : T_i \to T_i^*$ represents the belief morphism assigning to each type in a type space the corresponding hierarchy in the universal type space.

Attaching a type space-representation of the players’ hierarchies of beliefs to the game with payoff uncertainty $G$, one obtains a *Bayesian model*, i.e. the Bayesian game
\[ G^T = \langle N, \Theta, (A_i, T_i, \hat{u}_i)_{i \in N} \rangle, \] with payoff functions defined as \( \hat{u}_i : A \times \Theta \times T \to \mathbb{R} \).

Since players' types are payoff irrelevant, with a slight abuse of notation we write \( u_i \) and drop the dependence on \( T \).

Given a Bayesian model \( G^T \), player \( i \)'s conjectures are denoted by

\[ \psi^i \in \Delta (\Theta \times A_{-i} \times T_{-i}). \]

For each type \( t_i \), his consistent conjectures are

\[ \Psi_i (t_i) = \{ \psi^i \in \Delta (\Theta \times A_{-i} \times T_{-i}) : \text{marg}_{\Theta \times T_{-i}} \psi^i = \tau_i (t_i) \}. \]

For any \( B_i \subseteq A_i \), let \( \text{BR}_{iB}^i (\psi^i) \) denote the set of best responses in \( B_i \) to conjecture \( \psi^i \), and write \( \text{BR}_i (\psi^i) \equiv \text{BR}_{iA}^i (\psi^i) \). Formally:

\[ \text{BR}_{iB}^i (\psi^i) = \arg \max_{a_i \in B_i} \sum_{(\theta, a_{-i}, t_{-i})} u_i (\theta, a_i, a_{-i}) \cdot \psi^i (\theta, a_{-i}, t_{-i}) . \]

If \( a_i \in \text{BR}_{iB}^i (\psi^i) \), we say that \( \psi^i \) justifies action \( a_i \) in \( B_i \). Appealing again to the payoff-irrelevance of the epistemic types, with another abuse of notation we will write \( \text{BR}_i (\beta^i) \) for conjectures \( \beta^i \in \Delta (\Theta \times A_{-i}) \).

We define next the solution concept, \textit{Interim Correlated Rationalizability (ICR)}, introduced by Dekel et al. (2007):

---

3In Weinstein and Yildiz (2007) and in the present settings types are payoff-irrelevant, or purely epistemic (capturing beliefs). Chapter 3 below instead considers the general case which allows for \textit{payoff-types}: the space of uncertainty is \( \Theta = \Theta_0 \times \Theta_1 \times \ldots \times \Theta_n \) and each player \( i \) observes the \( i \)-th component of the realized \( \theta \). A player’s type \( t_i = (t_i, e_i) \) is made of two parts: a payoff-relevant component (what \( i \) knows, \( t_i \)) and a purely epistemic component (\( e_i \)), representing his beliefs about what he doesn’t know: his residual uncertainty about \( \theta \) and the opponents’ beliefs (i.e. \( \Theta_0 \times \Theta_{-i} \times E_{-i} \)). In Chapter 3 it is shown how the distinction between payoff-relevant and purely epistemic types is irrelevant for the purpose of Weinstein and Yildiz’s analysis of static settings. (For the same reason, payoff-types are not considered here.) The distinction instead is relevant in dynamic settings, as it affects the possibility that players have to revise their beliefs after observing unexpected moves. If \( \Theta \) is sufficiently \textit{rich}, and no payoff-types are allowed, \textit{sequential rationality} has no bite in dynamic settings. Not so if payoff-types are considered.

4Throughout the paper, the convention is maintained that “beliefs” are about \( \Theta \) and the opponents’ beliefs about \( \Theta \). That is, “beliefs” are about \textit{exogenous variables} only. The term “conjectures” instead refers to beliefs that also encompass the opponents’ strategies.
Definition 2.1. Fix a Bayesian model $G^T$. For each $t_i \in T_i$ and $i \in N$, set $ICR_i^0(t_i) = A_i$. For $k = 0, 1, \ldots$, let $ICR_i^k$ be such that $(a_i, t_i) \in ICR_i^k$ if and only if $a_i \in ICR_i^k(t_i)$ and $ICR_{-i}^k = \times_{j \neq i} ICR_j^k$. Then recursively, for $k = 1, 2, \ldots$

$$ICR_i^{k,T}(t_i) = \{a_i \in A_i : \exists \psi^{a_i} \in \Psi_i(t_i) \text{ s.t.: } a_i \in BR_i(\psi^{a_i}) \text{ and } \text{supp}_{a_i,T_i} \psi^{a_i} \subseteq ICR_{-i}^{k-1,T}(t_i)\}$$

Then, let $ICR_i^{\infty,T}(t_i) = \bigcap_{k \geq 0} ICR_i^{k,T}(t_i)$.

$ICR$ is a version of rationalizability (Pearce, 1984 and Bernheim, 1984) applied to the interim normal form, with the difference that the opponents’ strategies may be correlated in the eyes of a player.\(^5\) Dekel et al. (2007) proved that whenever two types $t_i \in T_i$ and $t'_i \in T'_i$ are such that $\pi_i^T(t_i) = \pi_i^{T'}(t'_i)$, $ICR_i^{\infty,T'}(t'_i) = ICR_i^{\infty,T}(t_i)$:

That is, $ICR$ is completely determined by a type’s hierarchies of beliefs, irrespective of the type space representation. Hence, we can drop the reference to the specific type space $T$, and without loss of generality work with the universal type space.

2.2.1 Structure of Rationalizability without Richness

Let $A_i^0 \subseteq A_i$ be the set of actions of player $i$ for which there exists a dominance state $\theta^{a_i} \in \Theta$. For each $k = 0, 1, \ldots$, set $A_i^{k-1} = \times_{j \neq i} A_j^k$ and $A_i^k = \times_{i \in N} A_i^k$. Recursively, for each $k = 1, 2, \ldots$, define

$$A_i^k = \{a_i \in A_i : \exists \beta^i \in \Delta(\Theta \times A_{-i}^{k-1}) \text{ s.t. } a_i = BR_i(\beta^i)\}$$

and let $A_i^\infty = \bigcup_{k \geq 0} A_i^k$.

In words, for each $k = 1, 2, \ldots$, the set $A_i^k$ is set of player $i$’s actions that are unique best response to conjectures concentrated on $A_i^{k-1}$. Actions in $A_i^0$ are those for which

\(^5\)Ely and Peski (2006) studied Interim (Independent) Rationalizability, that is simply Pearce’s solution concept applied to the interim normal form. Battigalli et al. (2008) studied the connections between these and other versions of rationalizability for incomplete information games.
there exists dominance states. For \( k \) then, each action in \( \mathcal{A}_i^k \) can be “traced back” to such dominance regions through a finite sequence of strict best responses.

**Remark 2.1.** It is easy to verify that, for each \( k = 1, 2, \ldots, A_i^{k-1} \subseteq A_i^k \). Also, since each \( A_i \) is finite, there exists \( K \in \mathbb{N} \) such that for each \( i \in N \), \( A_i^K = A_i^{K+1} = A_i^\infty \).

The next lemma shows that for each \( k \) and for each action \( a_i \in \mathcal{A}_i^k \), there exists a finite type for which \( a_i \) is the only action that survives after \((k + 1)\) rounds of iterated deletion of dominated actions.

**Lemma 2.1.** For each \( k = 0, 1, \ldots, \), for each \( a_i \in \mathcal{A}_i^k \) there exists a finite type \( t_i^k \in \hat{T}_i \) such that \( ICR_i^{k+1} (t_i^k) = \{a_i\} \).

**Proof.** The proof is by induction:

**Initial Step:** this is immediate, as for \( a_i \in \mathcal{A}_i^0 \), there exists \( \theta^{a_i} \in \Theta \) that makes \( a_i \) strictly dominant, and letting \( t_i^0 \) denote the type corresponding to common belief of \( \theta^{a_i} \), \( ICR_i^1 (t_i^0) = \{a_i\} \).

**Inductive Step:** Let \( a_i \in \mathcal{A}_i^k \), then there exists \( \beta_i^i \in \Delta (\Theta \times \mathcal{A}_i^{k-1}) \) such that \( \{a_i\} = BR_i (\beta_i^i) \). From the inductive hypothesis, there exists a function \( \kappa_i^{k-1} : \mathcal{A}_i^{k-1} \rightarrow \hat{T}_{-i} \) such that for each \( a_{-i} \in \mathcal{A}_{-i}^{k-1} \), \( \{a_{-i}\} = ICR_i^{k-1} (\kappa_i^{k-1} (a_{-i})) \). We want to show that there exists \( t_i^k \in \hat{T}_i \) such that \( ICR_i^{k+1} (t_i^k) = \{a_i\} \). Let \( \mu_i^i \in \Delta (\Theta \times \mathcal{A}_i^{k-1} \times \hat{T}_{-i}) \) be defined as

\[
\mu_i^i (\theta, a_{-i}, \kappa_i^{k-1} (a_{-i})) = \beta_i^i (\theta, a_i)
\]

and let \( t_i^k \) be defined as \( \tau_i^k (t_i^k) = \text{marg}_{\Theta \times \hat{T}_{-i}} \mu_i^i \). Then, by construction:

\[
\{\mu_i^i\} = \left\{ \psi_i \in \Psi_i (t_i^k) : \text{supp} \left( \text{marg}_{\mathcal{A}_{-i} \times \hat{T}_{-i}} \psi_i \right) \subseteq ICR_i^{k-1} \right\}
\]

and \( \{a_i\} = BR_i (\mu_i^i) \).

Hence: \( ICR_i^{k+1} (t_i^k) = \{a_i\} \). 

68
Definition 2.2. From lemma 1, let $\kappa_i^k : A_i^k \rightarrow \hat{T}_i$ be defined as a mapping such that for each $a_i \in A_i^k$, $\{a_i\} = ICR_i^k(\kappa_i^k(a_i))$ and $\kappa_i : A_i^\infty \rightarrow \hat{T}_i$ as a mapping such that for each $a_i \in A_i^\infty$, $\{a_i\} = ICR_i^\infty(\kappa_i(a_i))$. Given $\kappa_i : A_i^\infty \rightarrow \hat{T}_i$, define the set of types $\hat{T}_i \subseteq \hat{T}_i$ as follows:

$$\hat{T}_i := \left\{ t_i \in \hat{T}_i : t_i = \kappa_i(a_i) \text{ for some } a_i \in A_i^\infty \right\}.$$

Remark 2.2. Since $A_i^\infty$ is finite, the set $\hat{T}_i$ is finite.

As already mentioned, Weinstein and Yildiz assume richness, that is $A_i^0 = A_i$ for each $i$. Hence, it is immediate to construct types with a unique rationalizable action. Given such “dominance” types, they prove their main result through an “infection argument” to obtain the generic uniqueness result. Their proof is articulated in two main steps: first, a type’s beliefs are perturbed so that any rationalizable action for that type, is also “strictly rationalizable” for a nearby type (lemma 6 in Weinstein and Yildiz, 2007); then, they show that by a further perturbation, each “strictly rationalizable” action can be made uniquely rationalizable for an arbitrarily close type, perturbing higher order beliefs only (lemma 7, ibid.).

With arbitrary spaces of uncertainty (without richness), the argument requires two main modifications: first, the set of types $\hat{T}_i$ which will be used to start the “infection argument” had to be constructed (definition 2.2); then, a result analogous to Weinstein and Yildiz’s lemma 6 is proved (lemma 3 below), but with a different solution concept than “strict rationalizability”, which will be presented shortly (def 2.3). The difference with respect to “strict rationalizability”, parallels the difference between Weinstein and Yildiz’s “dominance types” and types $\hat{T}_i$ constructed above, so to be able to “trace back” a type’s hierarchies to the dominance regions for actions in $A^0$. Given these preliminary steps, the further perturbations of higher order beliefs
needed to obtain the result is completely analogous to Weinstein and Yildiz’s: lemma 4 below entails minor modifications of Weinstein and Yildiz’s equivalent (lemma 7).

The proof of the main result is based on the following solution concept:

**Definition 2.3.** For each \( i \in N \) and \( t_i \in T_i \), set \( \mathcal{W}_i^0 (t_i) = \mathcal{A}_i^0 \). For \( k = 0, 1, \ldots \), let \( \mathcal{W}_i^k \) be such that \( (a_i, t_i) \in \mathcal{W}_i^k \) if and only if \( a_i \in \mathcal{W}_i^k(t_i) \) and \( \mathcal{W}_i^k = \times_{j \neq i} \mathcal{W}_j^k \). For \( k = 1, 2, \ldots \), define recursively

\[
\mathcal{W}_i^k(t_i) = \{ a_i \in \mathcal{A}_i^k : \exists \psi_i \in \Delta (\Theta \times \mathcal{W}_{-i}^{k-1}) \text{ s.t. } \text{marg}_{\Theta \times T_{-i}} \psi_i = \tau_i(t_i) \text{ and } \{ a_i \} = BR_i(\psi_i) \}
\]

Let \( K \in \mathbb{N} \) be such that for each \( k \geq K \), \( \mathcal{W}_i^{k+1}(t_i) \subseteq \mathcal{W}_i^k(t_i) \) for all \( t_i \) and \( i \) (such \( K \) exists because of remark 1 above). Finally, define \( \mathcal{W}_i^\infty(t_i) := \bigcap_{k \geq K} \mathcal{W}_i^k(t_i) \).

Notice that for \( k < K \), \( \mathcal{W}_i^k(t_i) \) may be non-monotonic in \( k \), as for \( k < K \) the sets \( \mathcal{A}_i^k \) are increasing. Hence, up to \( K \), \( \mathcal{W}_i^k(t_i) \) may increase. When \( k \geq K \) though, \( \mathcal{A}_i^k = \mathcal{A}_i^\infty \) is constant, and the condition “\( \exists \psi_i \in \Delta (\Theta \times \mathcal{W}_{-i}^{k-1}) \)” becomes (weakly) more and more stringent, making the sequence \( \{ \mathcal{W}_i^k(t_i) \}_{k \geq K} \) monotonically (weakly) decreasing. Being always non-empty, \( \mathcal{W}_i^\infty(t_i) := \bigcap_{k \geq K} \mathcal{W}_i^k(t_i) \) is also non-empty and well-defined (as long as \( \mathcal{A}_i^0 \neq \emptyset \)).

The following lemma states a standard fixed point property, and it is an immediate implication of lemma 5 in Weinstein and Yildiz (2007).\(^6\)

**Lemma 2.2.** For any family of sets \( \{ V_i(t_i) \}_{t_i \in T_i, i \in N} \) such that \( V_i(t_i) \subseteq \mathcal{A}_i^\infty \) for all \( i \in N \) and \( t_i \in T_i \). If for each \( a_i \in V_i(t_i) \), there exists \( \psi^i \in \Delta (\Theta \times A_{-i} \times T_{-i}) \) such that \( \{ a_i \} = BR_i(\psi^i) \), \( \text{marg}_{\Theta \times T_{-i}} \psi^i = \tau_i(t_i) \) and \( \psi^i(\theta, a_{-i}, t_{-i}) > 0 \Rightarrow a_{-i} \in V_{-i}(t_{-i}) \), then \( V_i(t_i) \subseteq \mathcal{W}_i^\infty(t_i) \) for each \( t_i \).

\(^6\)This is because \( \mathcal{W}^\infty \) coincides with Weinstein and Yildiz’s \( W^\infty \) applied to the game with actions \( \mathcal{A}^\infty \).
We turn next to the analysis of higher order beliefs: the next lemma shows how for each $t_i$ and each action $a_i \in ICR_i^\infty(t_i) \cap \mathcal{A}_i^\infty$ that is justified by conjectures concentrated on $\mathcal{A}_i^\infty$, we can construct a sequence of types converging to $t_i$ for which $a_i$ survives to the iterated deletion procedure introduced in definition 2.3.

**Lemma 2.3.** Let $a_i \in ICR_i^\infty(t_i) \cap \mathcal{A}_i^\infty$ be such that there exists a justifying conjecture $\psi^{a_i} \in \Psi_i(t_i)$ such that $\text{supp}(\text{marg}_{A_i} \psi^{a_i}) \subseteq \mathcal{A}_i^\infty$. Then there exists $t_i(\varepsilon) \to t_i$ as $\varepsilon \to 0$ such that for each $\varepsilon > 0$, $a_i \in W_i^\infty(t_i(\varepsilon))$ and $t_i(\varepsilon) \in \hat{T}_i$ (hence $a_i \in W_i^k(t_i(\varepsilon))$ for all $k \geq K$)

**Proof.** Since $a_i \in ICR_i^\infty(t_i) \cap \mathcal{A}_i^\infty$, $\exists \psi^{a_i} \in \Delta(\Theta \times ICR_i^\infty)$ such that $a_i \in BR_i(\psi^i)$ and $\text{marg}_{\Theta \times T_i} \psi^i = \tau_i(t_i)$ and there exists $\beta^i \in \Delta(\Theta \times \mathcal{A}_i^\infty)$ such that $\{a_i\} = BR_i(\beta^i)$. Let $\kappa_{-i}$ be as in definition 2, and $v_i^{(t_i,a_i)} \in \Delta(\Theta \times \mathcal{T}_i)$ such that for each $(\theta, a_{-i}) \in \Theta \times \mathcal{A}_i^\infty$, $v_i^{(t_i,a_i)}(\theta, \kappa_{-i}(a_{-i})) = \beta^i(\theta, a_{-i})$. For each $\varepsilon \in [0,1]$, consider the set of types $T_i^\varepsilon \subseteq T_i^\circ$ such that each $T_i^\varepsilon = \hat{T}_i \cup T_i^\circ$. That is, $T_i^\varepsilon$ consists of two kinds of types: types $\bar{t}_i^{a_i} \in \hat{T}_i$ (see definition 2), which have a unique rationalizable action, and for each $t_i$, $a_i$ and $\varepsilon$ types $\bar{t}_i(t_i, a_i, \varepsilon) \in T_i^\circ$ with hierarchies of beliefs are implicitly defined as follows:

$$\tau_i^\circ(\bar{t}_i(t_i, s_i, \varepsilon)) = \varepsilon \cdot v_i^{(t_i,a_i)}(1 - \varepsilon) \left[\psi^{a_i} \circ \bar{t}_i^{-1, \varepsilon}\right],$$

where $\bar{t}_{-i, \varepsilon} : \Theta \times \mathcal{A}_{-i}^\infty \times T_{-i} \to \Theta \times T_{-i}^\varepsilon$ is the mapping given by $\bar{t}_{-i, \varepsilon}(\theta, a_{-i}, t_{-i}) = (\theta, t_{-i}(t_{-i}, a_{-i}, \varepsilon))$, and $\left[\psi^{a_i} \circ \bar{t}_i^{-1, \varepsilon}\right]$ denotes pushforward of $\psi^{a_i}$ given by $\bar{t}_i^{-1, \varepsilon}$.

Define $\gamma : \Theta \times T_{-i}^\varepsilon \to \Theta \times \mathcal{A}_{-i}^\infty \times T_{-i}^\varepsilon$ such that:

$$\forall \bar{t}_{-i}(t_{-i}, a_{-i}, \varepsilon) \in T_{-i}^\circ :$$

$$\gamma(\theta, \bar{t}_{-i}(t_{-i}, a_{-i}, \varepsilon)) = (\theta, a_{-i}, \bar{t}_{-i}(t_{-i}, a_{-i}, \varepsilon))$$

and for every $\bar{t}_{-i}^{a_{-i}} \in \hat{T}_{-i} \subseteq T_{-i}^\varepsilon$,

$$\gamma(\theta, \bar{t}_{-i}^{a_{-i}}) = (\theta, a_{-i}, \bar{t}_{-i}^{a_{-i}}).$$
Consider the conjectures $\psi^i \in \Delta \left( \Theta \times \mathcal{A}^{\infty}_i \times T^\varepsilon_i \right)$ defined as $\psi^i = \left( \tau^\varepsilon_{i_1(t_i,a_i,\varepsilon)} \circ \gamma^{-1} \right)$. By construction, they are consistent with type $\tau^\varepsilon_{i_1(t_i,a_i,\varepsilon)}$. Being a mixture of the beliefs $\psi^{a_i}$ (which made $a_i$ best reply) and of $\beta^i$ (which makes $a_i$ strict best reply), we have that $\{a_i\} = BR_i(\psi^i)$. Hence, setting $V_i^i(\tilde{t}_i(t_i,a_i,\varepsilon)) = \{a_i\}$ and $V_i^i(\tilde{t}_i) = \{a_i\}$ as in lemma 2, we have that $\{a_i\} \in \mathcal{W}^\infty_i(t_i)$ for all $t_i \in T^\varepsilon_i$. Finally, $\tilde{t}_i(t_i,a_i,\varepsilon) \rightarrow t_i$ as $\varepsilon \rightarrow 0$.

The next lemma shows that for any type $t_i$ and for any $a_i \in \mathcal{W}^k_i(t_i)$, $k = 0, 1, \ldots$, there exists a type that differs from $t_i$ only for beliefs of order higher than $k$, for which $a_i$ is the unique action which survives $(k + 1)$ rounds of the ICR-procedure.

For any type $t_i \in T_i^*$, let $t_i^m$ denote the $m$-th order beliefs of type $t_i$. (By definition of $T_i^*$, any $t_i \in T_i^*$ can be written as $t_i = (t_i^m)_{m=1}^\infty$.)

Lemma 2.4. For each $k = 0, 1, \ldots$, and for each $a_i \in \mathcal{W}^k_i(t_i)$, there exists $\tilde{t}_i : \tilde{t}_i^m = t_i^m$ for all $m \leq k$ and such that $\{a_i\} = ICR_i^{k+1}(\tilde{t}_i)$

Proof. The proof is by induction. For $k = 0$, $a_i \in \mathcal{W}^0_i(t_i) = \mathcal{A}_i^0$, so there exists a dominance state for action $a_i$, $\theta^{a_i}$. Let $\tilde{t}_i$ denote common belief of $\theta^{a_i}$, so that $\{a_i\} = ICR_i^1(\tilde{t}_i)$ (condition $\tilde{t}_i^0 = t_i^0$ holds vacuously). For the inductive step, write each $t_{-i}$ as $t_{-i} = (l, h)$ where

$$l = (t^1_{-i}, \ldots, t^k_{-i}) \quad \text{and} \quad h = (t^{k+1}_{-i}, t^{k+2}_{-i}, \ldots) .$$

Let $L = \{l : \exists h \text{ s.t. } (l, h) \in T_{-i}^* \}$. Let $a_i \in \mathcal{W}^k_i(t_i)$, and $\psi^{a_i} \in \Delta \left( \Theta \times \mathcal{W}^{k-1}_{-i} \right)$ the corresponding conjecture s.t.

$$\text{marg}_{\Theta \times T_{-i}} \psi^{a_i} = \tau_i(t_i)$$

72
and \( \{ a_i \} = BR_i (\psi^{a_i}) \). Under the inductive hypothesis, for each

\[
(a_{-i}, t_{-i}) \in \text{supp} (\text{marg}_{A_{-i} \times T_{-i}} \psi^{a_i}) ,
\]

\[
\exists \tilde{t}_{-i} (a_{-i}) = \left( l, \tilde{h} (a_{-i}) \right)
\]

s.t. \( ICR^{k}_{-i} (\tilde{t}_{-i} (a_{-i})) = \{ a_{-i} \} \).

Define the mapping

\[
\varphi : \text{supp} (\text{marg}_{\Theta \times A_{-i} \times L} \psi^{a_i}) \rightarrow \Theta \times T^*_{-i}
\]

by \( \varphi (\theta, a_{-i}, l) = (\theta, \tilde{t}_{-i} (a_{-i})) \). Define \( \tilde{t}_i \) by

\[
\tau_i^* (\tilde{t}_i) = (\text{marg}_{\Theta \times A_{-i} \times L} \psi^{a_i}) \circ \varphi^{-1}
\]

By construction,

\[
\text{marg}_{\Theta \times A_{-i} \times L} \tau_i^* (\tilde{t}_i) = \psi^{a_i} \circ \text{proj}_{\Theta \times A_{-i} \times L}^{-1} \circ \varphi^{-1} \circ \text{proj}_{\Theta \times L}^{-1}
\]

\[
= \psi^{a_i} \circ \text{proj}_{\Theta \times L}^{-1}
\]

\[
= \psi^{a_i} \circ \text{proj}_{\Theta \times A_{-i} \times T^*_{-i}}^{-1} \circ \text{proj}_{\Theta \times L}^{-1}
\]

\[
= \text{marg}_{\Theta \times A_{-i} \times L} \tau_i (t_i)
\]

where the first equality exploits the definition of lower order beliefs and the construction of type \( \tilde{t}_i \), the second follows from the definition of \( \varphi \), for which

\[
\text{proj}_{\Theta \times L \times A_{-i}}^{-1} \circ \varphi^{-1} \circ \text{proj}_{\Theta \times L}^{-1} = \text{proj}_{\Theta \times L}^{-1}
\]

The third is simply notational, and the last one by definition. Hence, by construction, we have \( ICR^{k+1}_i (\tilde{t}_i) = \{ a_i \} \), which completes the inductive step. \( \blacksquare \)

We are now in the position to present the main result:
Proposition 2.1. For each \( t_i \in \hat{T}_i \) and for each \( a_i \in ICR_i^\infty (t_i) \cap \mathcal{A}_i^\infty \) such that supp\( (\text{marg}_{A_i^\infty}^\psi(a_i)) \subseteq \mathcal{A}_i^\infty \), there exists a sequence \( \{t_i^\nu\} \subseteq \hat{T}_i \) s.t. \( t_i^\nu \to t_i \) and for each \( \nu \in \mathbb{N}, \{a_i\} = ICR_i^\infty (t_i^\nu) \).

Proof: Take any \( t_i \in \hat{T} \) and any \( a_i \in ICR_i^\infty (t_i) \) such that \( \text{supp} \text{marg} A_i^\nu \subseteq \)\( A_i^\infty \). From lemma 3, there exists a sequence of finite types \( t_i(\varepsilon) \to t_i \) (as \( \varepsilon \to 0 \)) such that \( a_i \in \mathcal{W}_i^\infty (t_i(\varepsilon)) \) for each \( \varepsilon > 0 \), hence, there exists a sequence \( \{t_i(n)\}_{n \in \mathbb{N}} \) converging to \( t_i \) such that \( a_i \in \mathcal{W}_i^k (t_i(n)) \) for all \( k \geq K \). Then we can apply lemma 4 to the types \( t(n) \): for each \( n \), for each \( k \geq K \) and for each \( a_i \in \mathcal{W}_i^k (t_i(n)) \), there exists \( \tilde{t}_i(k,n) \) such that \( \tilde{t}_i^k (k,n) = t_i^k (n) \) and \( \{a_i\} = ICR_i^{k+1} (\tilde{t}_i(k,n)) \). Hence, for each \( n \), the sequence \( \{\tilde{t}_i(k,n)\}_{k \in \mathbb{N}} \) converges to \( t_i(n) \) as \( k \to \infty \). Because the universal type-space \( T^* \) is metrizable, there exists a sequence \( t_i(n,k,n) \to t_i \) such that \( ICR_i^\infty (t_i(n,k,n)) \) = \( \{a_i\} \). Set \( t_i^\nu = t_i(n,k,n) \); \( t_i^\nu \to t_i \) as \( \nu \to \infty \) and \( ICR_i^\infty (t_i^\nu) = \{a_i\} \) for each \( \nu \).

2.3 Discussion

If it is common knowledge that no action is dominant \( (\mathcal{A}_i^0 = \emptyset) \), proposition 1 is vacuous. Weinstein and Yildiz’s richness condition amounts to assuming that \( \Theta \) is such that \( \mathcal{A}_i^0 = A_i \) for each \( i \): In this case, proposition 1 coincides with proposition 1 in Weinstein and Yildiz (2007).

Of more interest is the observation that all results in Weinstein and Yildiz (2007) (including the generic uniqueness result) hold true, without richness, whenever \( \mathcal{A}_i^\infty = A \).

Moreover, suppose that there exists a payoff state \( \theta^* \in \Theta \) for which payoff functions are supermodular, with player \( i \)'s higher and lower actions \( a^h_i \) and \( a^l_i \) respectively; and for each \( i, \mathcal{A}_i^0 = \{a^l_i, a^h_i\} \). Then under these conditions \( \mathcal{A}_i^\infty = A \), and Weinstein and
Yildiz’s full results are again obtained. This corresponds to the case considered by the global games literature, in which the underlying game has strategic complementarities and dominance regions are assumed for the extreme actions only. The difference is that in that literature supermodularity is assumed at all states (so that it is commonly known). In contrast, here it may be assumed for only one state, which only entails relaxing common knowledge that payoffs are not supermodular. This observation clarifies that, on the one hand, the equilibrium selection results obtained in the global games literature, which contrast with the non-robustness result (R.2), are exclusively determined by the particular class of perturbations that are considered. On the other hand, the generic uniqueness result can be obtained without assuming common knowledge of supermodularity or imposing richness: as argued, relaxing common knowledge that payoffs are not supermodular and that the corresponding extreme actions are not dominant would suffice to obtain the full results of Weinstein and Yildiz.

With minor changes, the proof above can be used to obtain a slightly stronger result (although not directly in terms of the primitives) just setting $A^0_i$ as the set of actions of player $i$ that are uniquely ICR for some type (clearly, this would always include the set of actions that are dominant in some state): Then, $A^\infty_i$ characterizes the set of actions that are uniquely rationalizable (hence robust) for some type.8

7See Morris and Shin (2003) and references therein.

8Since the ICR-correspondence is upper hemicontinuous in $T^*$, the uniqueness regions are open and locally constant: the corresponding predictions are thus robust.
Chapter 3

Higher Order Beliefs in Dynamic Environments

Abstract: The impact of higher order uncertainty in dynamic games is not as well understood as for static ones. This paper answers the following question: What are the strongest predictions that we can make, in a dynamic game, that are “robust” to possible misspecifications of agents’ higher order beliefs? The answer is provided by a new solution concept, Interim Sequential Rationalizability (ISR): It is shown that, if the space of uncertainty is “sufficiently rich”, ISR is the strongest upper hemi-continuous solution concept on the universal type space. Furthermore, ISR is generically unique. ISR is in general very weak, but its weakness depends on other modeling assumptions concerning agents’ information about the environment. These results reveal an interaction with modeling assumptions that have hitherto received little attention in the context of higher order uncertainty (namely, the distinction between knowledge and certainty), and raise further questions of robustness: It is shown that ISR is type space-invariant, that is it depends solely on agents’ hierarchies, not on the type space; In “private-values” environments, ISR is also invari-
ant to the representation of certainty as knowledge, a novel robustness property called model-invariance.

**Keywords:** dynamic games – hierarchies of beliefs – higher order beliefs – robustness – uniqueness

**JEL Codes:** C72; C73; D82.

### 3.1 Introduction

A large literature, beginning with Rubinstein (1989), has explored the impact of perturbations of common knowledge assumptions in static games. The overall message is negative: many predictions depend on the simplifying assumptions implicit in the modeling choices made. If common knowledge assumptions are relaxed, only weak predictions are robust to possible misspecifications of agents’ higher order beliefs. In particular, Weinstein and Yildiz (2007) showed that, when no common knowledge assumptions on the payoffs of a static game are imposed, the strongest “robust” predictions are those based on (interim correlated) rationalizability alone. In contrast, the role of higher order uncertainty in dynamic games is not well understood.

This paper characterizes the strongest predictions that are “robust” to possible misspecifications of agents’ higher order beliefs in dynamic games. This characterization depends on implicit assumptions that have hitherto received little attention. These assumptions concern the distinction between knowledge and certainty (probability-one belief). While mostly inconsequential in static games, in dynamic games the distinction is crucial. If an agent “knows” something, he would be certain of it after observing any event (expected or not). Not so if “knowledge” is replaced by probability-one belief, as Bayes’ rule puts no restrictions on the conditional beliefs after unexpected events.
Addressing robustness questions in dynamic games, it is then important to maintain agents’ information about the environment separate from their beliefs. A type in this paper consists of two components: one describing the agent’s information, or payoff-type; and one purely epistemic, parametrizing his beliefs. Different assumptions on the agents’ information affect the robustness of a model’s predictions: For example, it will be shown that in environments with no information, in which types are purely epistemic, the “strongest robust” predictions are weaker than in environments with private values, in which players know their own payoff function, and higher order uncertainty only concerns the opponents’ payoffs (and their beliefs about their opponents’). Hence, to provide a thorough answer to the question of robustness, it is necessary to “deconstruct” the modeling activity, so to isolate the different intervening forces.

Modeling a strategic situation with incomplete information requires a specification of agents’ entire hierarchies of beliefs. A “complete model” includes a description of all possible hierarchies of beliefs agents may have over the underlying space of uncertainty; then, a solution concept assigns to each hierarchy of beliefs the set of strategies that the agent might play (possibly a singleton). In practice though it is common to select a small subset of all the possible hierarchies of beliefs to focus on. Then, the “robustness” of our predictions depends on properties of continuity of the solution concept correspondence on the space of all possible hierarchies of beliefs.

Since Harsanyi (1967-68), it is common practice to represent hierarchies of beliefs implicitly, by means of (non-universal) type spaces. These type spaces impose common knowledge assumptions on agents’ hierarchies of beliefs. The particular type space chosen to represent a given set of hierarchies of beliefs may potentially affect the predictions of a solution concept. Invariance of the predictions with respect to the
type space is thus a distinct robustness property from that discussed in the previous paragraph. This property is called \textit{type space-invariance}.

Finally, the modeling activity typically imposes common knowledge restrictions not only on agents’ beliefs, but also on payoffs. For instance, if all types under consideration have beliefs concentrated on a subset $\Theta$ of all possible payoff states, i.e. if there is \textit{common certainty} of $\Theta$, it is common practice to ignore the states not in $\Theta$ in the specification of the model, thus transforming \textit{common certainty} into \textit{common knowledge} assumptions. I will say that a solution concept is \textit{model-invariant} if its predictions are invariant to the representation of \textit{certainty} as \textit{knowledge} (and vice-versa). Model-invariance is a novel robustness property, in general more demanding than type space-invariance, but equivalent to it in static environments.

To answer the questions above, a solution concept for dynamic games of incomplete information is introduced: \textit{Interim Sequential Rationalizability} (\textit{ISR} hereafter). \textit{ISR} captures the implications of agents’ \textit{sequential rationality} and \textit{initial common certainty of sequential rationality}, and extends ideas from solution concepts introduced by Ben-Porath (1997) and Battigalli and Siniscalchi (2007). It is proven that, if no common knowledge assumptions on payoffs are imposed, \textit{ISR} is the strongest upper hemicontinuous solution concept on the \textit{universal type space} (Mertens and Zamir, 1985): any (strict) refinement of \textit{ISR} delivers predictions that are not robust to perturbations of higher order beliefs.

This delivers a thoroughly negative message, as \textit{ISR} is in general a very weak solution concept.$^1$ Its weakness though is highly sensitive to the fine details of the

---

$^1$It is important to emphasize that the result holds once \textit{all} common knowledge assumptions are being relaxed: although common knowledge assumptions are intrinsically strong, in some cases analysts may have reasons to believe that some of them are actually satisfied. In that case, the robustness exercise performed here may be unnecessarily demanding, and stronger predictions may be robust if only \textit{some} such assumptions are relaxed. Chapter 2 above investigates the problem in
maintained assumptions about players’ knowledge and their beliefs: For instance, if common knowledge assumptions are relaxed in an environment with no information, ISR coincides with (interim correlated) rationalizability, that is iterated strict dominance in the interim normal form; If instead we consider private-values settings, then ISR strictly refines rationalizability, and it is shown to be equivalent to Dekel and Fudenberg’s (1990) $S^\infty W$-procedure applied to the interim normal form (that is, one round of deletion of weakly dominated strategies, followed by iterated strict dominance).

It is further shown that, irrespective of the information structure, ISR is type space-invariant. The type space-invariance and upper hemicontinuity results parallel analogous results in Dekel, Fudenberg and Morris (2007) on (interim correlated) rationalizability. Model-invariance instead is a novel robustness property. ISR is shown to satisfies this property in environments with private values, but not in environments with no information.

Weinstein and Yildiz (2007) showed that if no common knowledge assumptions are imposed, static games are generically dominance-solvable. This result generalizes an important insight from the literature on (static) global games: The pervasive multiplicity of equilibria that we observe in standard models stems from the high degree of coordination that common knowledge assumptions force on the agents’ beliefs.\(^2\) The existing literature is not conclusive as to whether the same insight applies to dynamic games: As the game unfolds, players may extract information about the environment from the history of play, updating their initial beliefs; furthermore, some information endogenously becomes common knowledge (e.g. public histories), possibly serving

\(^2\)Morris and Shin (2003) survey that literature.
as a coordination device and favoring multiplicity. This tension is at work in several studies on dynamic global games, in which the familiar uniqueness results don’t obtain, casting some shadow on the possibility of drawing tempting analogies from static to dynamic environments.\textsuperscript{3} This paper provides a generic uniqueness result for \textit{ISR}, which shows that imputing multiplicity to common knowledge assumptions is legitimate in dynamic games as well.\textsuperscript{4} Furthermore, very weak epistemic conditions (generically) suffice for agents to achieve coordination of expectations, with no need to invoke sophisticated backward or forward induction reasoning.

The rest of the paper is organized as follows: Section 3.2 contains several examples to illustrate the main concepts and results. Section 3.3 introduces the game theoretic framework; \textit{ISR} is presented in Section 3.4. Section 3.5 provides some robustness results: \textit{upper hemicontinuity}, \textit{type space-invariance} and \textit{model-invariance}. Section 3.6 explores the structure of \textit{ISR} on the universal type space, and proves that under a suitable \textit{richness condition} it is generically unique and that any refinement of \textit{ISR} is not robust. Section 3.7 concludes and discusses the relation with the literature. Proofs are in the appendix.

\textsuperscript{3}For example, Angeletos, Hellwig and Pavan (2007) show how imposing the global games information structure on dynamic environments may not deliver the familiar uniqueness results.

\textsuperscript{4}Thus, despite the important differences discussed above, the impact of higher order beliefs in dynamic environments is analogous to the static case. This finding contrasts with previous results for dynamic environments (e.g. Chassang, 2009 and Angeletos et al., 2007).
3.2 Relaxing CK-assumptions and Robustness in Dynamic Games

3.2.1 Preliminaries and Examples

The next example illustrates the logic of Interim Sequential Rationalizability (ISR) in the context of a game with complete information, and compares it with Pearce’s (1984) Extensive Form Rationalizability (EFR). In the example, the predictions made by EFR seem very compelling. Yet, when common knowledge assumptions are relaxed, these predictions are not robust to perturbations of higher order beliefs. Example 3.2 will illustrate the non-robustness of EFR.

Example 3.1. Consider the game in figure 3.1, and suppose that it is common knowledge that \( \theta = 0 \) (hence, the game has complete information. Denote this model by \( T_{CK} = \{t_{CK}\} \)). Then, strategy \( a_3 \) is dominated by \( a_1 \). Thus, if at the beginning of the game player 2 thinks that 1 is rational, he assigns zero probability to \( a_3 \) being played. For example, 2 could assign probability one to \( a_1 \), so that the next information set is unexpected. We can consider two different hypothesis on the agents’ “reasoning” in the game:
[H.1] 2 believes that 1 is rational even after an unexpected move; or

[H.2] 2 believes that 1 is rational as long as he is not surprised, but he is willing to consider that 1 is not rational if he observes an unexpected move.

If [H.1] is true, in the subgame player 2 would still assign zero probability to $a_3$, and play $b_1$ if rational. If 1 believes [H.1] and that 2 is rational, he would expect $b_1$ to be played. Then, if 1 is also rational, he would play $a_2$. This is the logic of Pearce’s (1984) Extensive Form Rationalizability (EFR), which delivers $(a_2, b_1)$ as the unique outcome in this game.

Now, let’s maintain that 2 is rational, but assume [H.2] instead: once surprised, player 2 is willing to consider that 1 is not rational. Hence, in the subgame, he may assign probability one to $a_3$ being played, which would justify $b_2$. If at the beginning 2 assigned positive probability to $a_2$, then the subgame would not be unexpected, and player 2 would still assign zero probability to $a_3$, making $b_1$ the unique best response. Thus, if [H.2] is true, either $b_1$ or $b_2$ may be played by a rational player 2. If 1 believes that 2 is rational and that [H.2] is true, he cannot rule out either $b_1$ or $b_2$, and so both $a_1$ and $a_2$ may be played by a rational player 1. Interim Sequential Rationalizability ($\mathcal{ISR}$) corresponds to the latter form of reasoning, and selects $\{a_1, a_2\} \times \{b_1, b_2\}$ for $t^{CK}$, written $\mathcal{ISR}(t^{CK}) = \{a_1, a_2\} \times \{b_1, b_2\}$. □

The logic of EFR in this example seems compelling. Yet, as the next example shows, its predictions are not robust. To illustrate the point we will construct a sequence of hierarchies of beliefs, converging the hierarchies in example 1, in which $(a_1b_2)$ is the unique $\mathcal{ISR}$-outcome (hence, also the unique EFR outcome). Since $(a_1, b_2)$ is ruled out by EFR in the limit, but uniquely selected along the converging
sequence, EFR is not “robust”.

**Example 3.2.** In the game of figure 3.1, let the space of uncertainty be \( \Theta = \{0, 3\} \). Suppose that player 1 knows the true state, while 2 doesn’t (and this is common knowledge). Let \( t^* = (t_1^*, t_2^*) \) represent the situation in which there is *common certainty* that \( \theta = 0 \): “type” \( t_1^* \) knows that \( \theta = 0 \), and puts probability one on 2 being type \( t_2^* \); type \( t_2^* \) puts probability one on \( \theta = 0 \), and player 1 being type \( t_1^* \). A reasoning similar to that in example 1 implies that \( \{a_1, a_2\} \) and \( \{b_1, b_2\} \) are the sets of ISR strategies for \( t_1^* \) and \( t_2^* \), written \( \text{ISR}(t^*) = \{a_1, a_2\} \times \{b_1, b_2\} \).

Now, a sequence of types \( \{t^m\} \) will be constructed, converging to \( t^* \), such that \( (a_1, b_2) \) is the unique ISR-outcome for each \( t^m \): Since it is the unique ISR-outcome along the sequence, any (strict) refinement of ISR that rules out \( (a_1, b_2) \) for type \( t^* \) would not be upper hemicontinuous, hence not “robust”. The type spaces used for the construction of the sequence \( \{t^m\} \) can be viewed as a “perturbed” version of Rubinstein’s (1989) *e-mail game*, in which there is a small probability that the original e-mail is sent in the wrong state of nature.

Fix \( \varepsilon \in (0, \frac{1}{6}) \) and let \( p \in \left(0, \frac{\varepsilon}{1 - 2\varepsilon}\right) \). Consider the set of type profiles \( T_1^\varepsilon \times T_2^\varepsilon \subseteq T_{0^*} \), where \( T_1^\varepsilon = \{-1^3, 1^0, 1^3, 3^3, 5^0, 5^3, \ldots\} \) and \( T_2^\varepsilon = \{0, 2, 4, \ldots\} \). Types \( k^\theta \) \((k = -1, 1, 3, \ldots, \theta = 0, 3)\) are player 1’s types who know that the true state is \( \theta \); 2’s types only know their own payoffs, which are constant across states, but don’t know the opponent’s type. Suppose that beliefs are described as follows. Type \(-1^3 \) puts probability one on facing type 0; type 0 assigns probability \( \frac{1}{1+p} \) to type \(-1^3 \), and complementary probability to types \( 1^0 \) and \( 1^3 \), with weights \((1 - \varepsilon)\) and \( \varepsilon \), respectively. Similarly, for all \( k = 2, 4, \ldots \) player 2’s type \( k \) puts probability \( \frac{1}{1+p} \) on 1’s types \((k - 1)^0 \) and \((k - 1)^3 \), with weights \((1 - \varepsilon)\) and \( \varepsilon \) respectively, and complementary probability \( \frac{p}{1+p} \) on the \((k + 1)\)-types, with weight \((1 - \varepsilon)\) on \((k + 1)^0 \) and \( \varepsilon \) on \((k + 1)^3 \).
other types of player 1, with \( k = 1, 3, \ldots \), and \( \theta = 0, 3 \), type \( k^0 \) puts probability \( \frac{1}{1+p} \) on 2’s type \( k - 1 \), and complementary probability on 2’s type \( k + 1 \). (The type space is represented in figure 3.2.) Notice that the increasing sequence of even \( k \)’s and odd \( k^0 \)’s converges to \( t^* \) as we let \( \varepsilon \) approach 0.\(^5\) It will be shown that player 2’s types 0, 2, 4, ... only play \( b_1 \), while 1’s types 1\(^0\), 3\(^0\), ... only play \( a_1 \): All types \( k^3 \) \((k = -1, 1, 3, \ldots)\) would play \( a_3 \), for they know it is dominant. Type 0 puts probability \( \frac{1}{1+p} \) on type \(-1\), who plays \( a_3 \); given these initial beliefs, type 0’s conditional conjectures after \( In \) must put probability at least \( \frac{1}{1+p} \) on \( a_3 \) being played, which makes \( b_2 \) optimal for him. Type 1\(^0\) also puts probability \( \frac{1}{1+p} \) on type 0, who plays \( b_2 \), thus \( a_1 \) is the unique best response. Type 2’s initial beliefs are such that type 1\(^0\) plays \( a_1 \) and types 1\(^3\) and 3\(^3\) play \( a_3 \). Hence, the probability of \( a_3 \) being played, conditional on \( In \) being observed, must be no smaller than

\[
\Pr (\theta = 3|\text{not } 1^0) = \frac{\varepsilon}{1 - \left(\frac{1}{1+p}\right)(1 - \varepsilon)} = \frac{(1 + p) \varepsilon}{p + \varepsilon}
\]

\(^5\) This convergence of the hierarchies of beliefs is in the product topology.
Given that \( p < \frac{x}{1-2x} \), this probability is greater than \( \frac{1}{2} \). Hence, playing \( b_2 \) is the unique best response, irrespective of type 2’s conjectures about \( 3^0 \)’s behavior. Given this, type \( 3^0 \) also plays \( a_1 \). The reasoning can be iterated, so that for all types \( 1^0, 3^0, 5^0, \ldots \), \( a_1 \) is the unique \( \mathcal{ISR} \) strategy, while for all types \( 0, 2, 4, \ldots \) of player 2, strategy \( b_2 \) is. □

The main result of the paper generalizes the insights of the example. In fact, it will be shown that when all common knowledge assumptions are relaxed, the strongest “robust” predictions are those based on \( \mathcal{ISR} \) alone: Any (strict) refinement delivers non-robust predictions. But any game theoretic model implicitly makes common knowledge assumptions. We thus need to be precise about what we mean when we say that “all common knowledge assumptions are being relaxed”. For instance, suppose that the parameter space for the game in figure 3.1 is \( \Theta = \{0, 3\} \), so that strategy \( a_3 \) is dominant in some state. If only common knowledge of \( \Theta \) is assumed, then it is not common knowledge that \( a_3 \) is not dominant. Thus, relaxing all common knowledge assumptions essentially means to consider the set of all possible hierarchies of beliefs (or “types”) that players may have about a sufficiently rich space of uncertainty, \( \Theta^* \). Assuming common knowledge of \( (\Theta^*, T^*_{\Theta}) \), where \( T^*_{\Theta} \) denotes the \( \Theta^* \)-based universal type space, entails essentially no loss of generality. Then, “robustness” may be formulated as a property of continuity of the solution concept on \( T^*_{\Theta} \).

In a dynamic game, the details of the modeling assumptions on agents’ information about \( \Theta^* \) are crucial. The next example shows how the distinction between knowledge and certainty may affect the predictions of a solution concept, raising novel questions of robustness.

**Example 3.3.** Consider the game in figure 3.3, and assume that \( \Theta^* = \{-3, 3\} \).
Let’s consider two alternative setups.

[A]: Player 2 observes the realization of $\theta$, and this is common knowledge. Suppose that agents are commonly certain that $\theta = 3$, denoted by $t^A = (t^A_1, t^A_2) \in T^*_A$: if type $t^A_2$, player 2 knows that $\theta = 3$, and puts probability one on player 1 being type $t^A_1$; if type $t^A_1$, agent 1 puts probability one on $\theta = 3$ and agent 2 being type $t^A_2$. (Notice that $t^A$ is not “common knowledge of $\theta = 3$” because $t^A_1$ does not know that $\theta = 3$.) Since 1 knows that 2 knows the true state of $\theta$, if 1 is rational and believes that 2 is rational, he must play Out: if 2 is rational and knows $\theta$, 1 obtains $-3$ in the subgame irrespective of the realization of $\theta$.

[B]: Players observe nothing about $\theta$, and this is common knowledge. Now, let’s maintain that agents share common certainty of $\theta = 3$, denoted by $t^B = (t^B_1, t^B_2) \in T^*_B$: each type $t^B_i$ ($i = 1, 2$) puts probability one on $\theta = 3$ and the opponent being type $t^B_j$ ($j \neq i$). Hence, agents in this setup share the same hierarchies as types $t^A$. If player 1 believes that $\theta = 3$, and believes that player 2 plays $D$, then 1’s optimal response is to play Out. If 2 believes this, then his information set is unexpected: If called to move, player 2 would be surprised and, if he revises his beliefs in favor of $\theta = -3$, he may play $U$. If 1 anticipates this, then playing $In$ is optimal even if 1 believes that $\theta = 3$. That is because 1 knows that 2, although certain of $\theta = 3$, does not know that $\theta = 3$. So, types $t^B$ share the same hierarchies as types $t^A$. 

---

Figure 3.3: Examples 3.3 and 3.4
but under this specification of agents’ information, \( \text{ISR}(t^B) = \{\text{In}, \text{Out}\} \times \{U, D\} \) (while \( \text{ISR}(t^A) = \{(\text{Out}, D)\} \)). □

Hence, in a dynamic game, the details of agents’ information about \( \Theta^* \) may crucially affect our predictions. So, for instance, types \( t^* \) in example 3.2 share the same hierarchies of beliefs as types \( t^{CK} \) in example 3.1, but agents have different information about \( \theta \) in the two examples. Yet, in this case \( \text{ISR}(t^*) = \text{ISR}(t^{CK}) \). Then, why is it the case that \( \text{ISR}(t^{CK}) = \text{ISR}(t^*) \) while \( \text{ISR}(t^A) \neq \text{ISR}(t^B) \)? What if we change agents’ information in example 3.2? How do differences in agents’ information affect the properties of “robustness” with respect to perturbations of agents’ hierarchies of beliefs? The purpose of the paper is to provide systematic answers to these questions.

The rest of the section provides a non-technical exposition of the formalization and the main results.

3.2.2 Non-technical Presentation of the Approach and Results

Basic Setup

Modeling Incomplete Information. A dynamic game is defined by an extensive form \( \langle N, \mathcal{H}, \mathcal{Z} \rangle \) \((N = \{1, ..., n\}) \) is the set of players; \( \mathcal{H} \) and \( \mathcal{Z} \) the sets of partial and terminal histories, respectively) and players’ payoffs, defined over the terminal histories. Incomplete information is modelled parametrizing the payoff functions on a rich space of uncertainty \( \Theta^* \), letting \( u_i : \mathcal{Z} \times \Theta^* \rightarrow \mathbb{R} \). In general, let \( \Theta^* \) be written as

\[
\Theta^* = \Theta^*_0 \times \Theta^*_1 \times ... \times \Theta^*_n.
\]

For each \( i = 1, ..., n, \Theta^*_i \) is the set of player \( i \)’s payoff types, i.e. possible pieces
of information that player $i$ may have about the payoff state; $\Theta_0^*$ instead represents any residual uncertainty that is left after pooling all players’ information. The interpretation is that when the payoff state is $\theta = (\theta_0, \theta_1, \ldots, \theta_n)$, player $i$ knows that $\theta \in \Theta_0^* \times \{\theta_i\} \times \Theta_{-i}^*$, where $\Theta_{-i}^* = \times_{j \in N \setminus \{i\}} \Theta_j^*$. The tuple $\langle \Theta_0^*, (\Theta_i^*, u_i)_{i \in N} \rangle$, where $u_i : Z \times \Theta^* \to \mathbb{R}$ for each $i \in N$, represents players’ information about everyone’s preferences, and is referred to as information structure.\(^6\) (Case A in example 3 can be formalized assuming that $\Theta_0^*$ and $\Theta_1^*$ are singletons, or (w.l.o.g.) that $\Theta_2^* = \Theta^*$)

Special cases of interest are those of private values, in which each $u_i$ depends only on $\Theta_i^*$; and the case in which the $u_i$’s depend on $\Theta_0^*$ only, so that players have no information about payoffs (i.e., without loss of generality, $\Theta^* = \Theta_0^*$). In a private values-environment, each player’s payoffs only depend on what he knows. Hence, in private values-environments (PV), the fact that everybody knows his own payoffs is common knowledge: Uncertainty only concerns the opponents’ payoffs and higher order beliefs. (In example 2, player 2’s payoffs are constant across states, and player 1 observes $\theta$. Hence, $\Theta_1^* = \Theta^*$, and the environment has private-values.) In contrast, in no information-environments (NI) players have no knowledge of the payoff state. In particular, each player does not know his own preferences over the terminal histories: He merely holds beliefs about that. (Case B in example 3 is a NI-environment.)

The Interim Approach. An information structure and an extensive form define a dynamic game with payoff uncertainty, but do not complete the description of the strategic situation: Players’ beliefs about what they don’t know must be specified, i.e. beliefs about $\Theta_0^* \times \Theta_{-i}^*$ (first order beliefs), beliefs about $\Theta_0^* \times \Theta_{-i}^*$ and the opponents’ beliefs about $\Theta_0^* \times \Theta_{-i}^*$ and the opponents’

\(^6\)The standard definition of an information structure specifies players’ partitions and priors over the set of states. Here players’ beliefs, i.e. their priors, are not specified. Appending players’ beliefs to an information structure, in the terminology adopted here, delivers a model (see below).
first order beliefs (second order beliefs), and so on. The $\Theta^*$-based universal type space, $T^*_{\Theta^*}$, can be thought of as the set of all such hierarchies of beliefs (Mertens and Zamir, 1985). Each element $t_i = (\theta_i, e_i) \in T^*_{\Theta^*}$ is a complete description of player $i$: His information $\theta_i$ (what he knows), and his epistemic type $e_i$ (his beliefs about what he doesn’t know, $\Theta^*_0 \times \Theta^*_{-i} \times E^*_{-i}$).

It is important to stress one point: Players’ hierarchies of beliefs (or types) are purely subjective states describing a player’s view of the strategic situation he is facing. As such, they enter the analysis as a datum and should be regarded in isolation (i.e. player by player and type by type). Nothing prevents players’ views of the world to be inconsistent with each other (i.e. to assign positive probability to opponents’ types other than the actual ones); they are part of the environment (exogenous variables), and as such game theoretic reasoning cannot impose restrictions on them; it is given such beliefs that we can apply game theoretic reasoning to make predictions about players’ behavior (the endogenous variables). The name “Interim” Sequential Rationalizability is meant to emphasize this point.

Robustness(-es)

Continuity in the Universal Model. To explore what predictions retain their validity when all common knowledge assumptions are relaxed, we specify a rich space of uncertainty $\Theta^*$ and look at players’ types in the universal space $T^*_{\Theta^*}$: Assuming common knowledge of $\langle \Theta^*, T^*_{\Theta^*} \rangle$ entails no loss of generality; $\langle \Theta^*, T^*_{\Theta^*} \rangle$ will thus be referred to as the universal model. A solution concept assigns to each player’s hierarchy of beliefs (the exogenous variables) a set of strategies (the endogenous variables).

In modeling a strategic situation, applied theorists typically select a subset of the possible hierarchies to focus on. To the extent that the “true” hierarchies are
understood to be only close to the ones considered in a specific model, the concern for robustness of the theory’s predictions translates into a continuity property of the solution concept correspondence. In this paper a solution concept is “robust” if it never rules out strategies that are not ruled out for arbitrarily close hierarchies of beliefs: This is equivalent to requiring upper hemicontinuity (u.h.c.) of the solution concept correspondence on \( T_{\Theta^*} \).

Clearly, a solution concept that never rules out anything is robust, but not interesting. One way to solve this trade-off is to look for a “strongest robust” solution concept. It will be shown (Proposition 3.4) that for any type \( t \in T_{\Theta^*} \) and any \( s \in ISR(t) \), there exists a sequence \( t^m \to t \) such that \( \{s\} = ISR(t^m) \) for any \( m \). Furthermore, \( ISR \) is u.h.c. on \( T_{\Theta^*} \) (Proposition 3.1). Hence, \( ISR \) is a strongest robust solution concept.

Type Spaces, Models and Invariance. \textit{Upper hemicontinuity in the universal model} addresses a specific robustness question: Robustness, with respect to small “mistakes” in the modeling choice of which subset of players’ hierarchies to consider. When applied theorists choose a subset of \( \Theta^* \)-hierarchies to focus on, they typically represent them by means of (non universal) \( \Theta^* \)-based type spaces, rather than elements of \( T_{\Theta^*} \).

**Definition 3.1.** A \( \Theta^* \)-based type space is a tuple

\[
T_{\Theta^*} = \langle \Theta^*_{0}, (\Theta^*_i, E_i, T_{i,\Theta^*}, \tau_i)_{i \in N} \rangle
\]

such that \( T_{i,\Theta^*} = \Theta^*_i \times E_i \) and \( \tau_i : T_{i,\Theta^*} \to \Delta (\Theta^*_0 \times T_{-i,\Theta^*}) \)

\( ^7 \)Any single-valued and constant solution concept is a strongest u.h.c. solution concept, but not interesting. \( ISR \) is the strongest among the class of solution concepts that satisfy \textit{initial common certainty of sequential rationality}.
Each type \( t_i \) in \( T_{i,\Theta^*} \) corresponds to a \( \Theta^* \)-hierarchy for player \( i \) (an element of \( T_{i,\Theta^*} \)). Representing a hierarchy as a type in a (non-universal) type space \( T_{\Theta^*} \) rather than an element of \( T_{\Theta^*} \) does not change the common knowledge assumptions on the information structure (\( \Theta^* \)), but it does impose common knowledge assumptions on players’ hierarchies of beliefs, and their correlation with the states of nature \( \theta_0 \). A solution concept is type space-invariant if the behavior prescribed for a given hierarchy does not depend on whether it is represented as an element of \( T_{\Theta^*} \) or of a different \( T_{\Theta^*} \). Thus, type space-invariance is also a robustness property: Robustness, with respect to the introduction of the extra common knowledge assumptions on players’ beliefs (and their correlations with \( \theta_0 \)) imposed by non-universal type spaces. Proposition 3.2 shows that \( ISR \) is type space-invariant under all information structures.

In writing down a game, as analysts we typically make common knowledge assumptions not only on players’ beliefs, but also on payoffs. For instance, suppose that all types in \( T_{\Theta^*} \) have beliefs concentrated on some strict subset \( \Theta \subset \Theta^* \), i.e. there is common certainty of \( \Theta \) in \( T_{\Theta^*} \). In applied work, states that receive zero probability \( (\theta \in \Theta^* \setminus \Theta) \) are usually excluded from the model. That is: common certainty of \( \Theta \) is made into common knowledge of \( \Theta \).

**Definition 3.2.** A model of the environment is a \( \Theta \)-based type space, where \( \Theta \) is such that \( \Theta_k \subseteq \Theta_k^* \) for each \( k = 0, ..., n \).

Each type in a model induces a \( \Theta \)-hierarchy, and hence a \( \Theta^* \)-hierarchy. A solution concept is model invariant if the behavior is completely determined by the \( \Theta^* \)-hierarchies irrespective of the model they are represented in. Model invariance is a stronger robustness property than type space-invariance, as it also requires robust-

\(^8\)Type space-dependence in static settings has been studied by Ely and Peski (2006) and Dekel, Fudenberg and Morris (2007).
ness to the introduction of extra common knowledge assumptions on the information structure.

**Example 3.4 (Model Dependence)** Consider example 3 again, and suppose that the “true” situation is the one described by case B, in which agents share *common certainty* that $\theta = 3$, but they have *no information* about $\theta$. As argued, for these types the solution concept delivers $\mathcal{ISR} (t^B) = \{In, Out\} \times \{U, D\}$. If that situation is modelled as there being *common knowledge* of $\theta = 3$, then the prediction of $\mathcal{ISR}$ is the backward induction outcome $\mathcal{ISR} (\bar{t}^{CK}) = (Out, D)$. Hence, despite $\bar{t}^{CK}$ and $t^B$ share the same hierarchies of beliefs, $\mathcal{ISR}$ delivers different predictions depending on the model in which those hierarchies are represented. Hence, in *NI-environments*, $\mathcal{ISR}$ is not *model-invariant*. □

Proposition 3.3 in section 3.5.3 shows that $\mathcal{ISR}$ is *model-invariant* in *environments* with *private-values*. (That is why $\mathcal{ISR} (\bar{t}^{CK}) = \mathcal{ISR} (t^*)$ in example 2.)

### 3.3 Game Theoretic Framework

The analysis that follows applies to dynamic games, defined by an *extensive form* and a *model*, i.e. a specification of players’ information and beliefs about everyone’s preferences and beliefs. The present work is concerned with robustness of solution concepts to different specifications of *players’ model*, i.e. the *extensive form* will be maintained fixed throughout, and the *model* varied. These concepts are formally introduced next:

**Extensive Forms.** An extensive form is defined by a tuple

$$\Gamma = \langle N, \mathcal{H}, \mathcal{Z}, (A_i)_{i \in N} \rangle$$
where $N = \{1, ..., n\}$ is the set of players; for each player $i$, $A_i$ is the (finite) set of his possible actions; a finite collection of histories (concatenations of action profiles), partitioned into the set of terminal histories $\mathcal{Z}$ and the set of partial histories $\mathcal{H}$ (which includes the empty history $\emptyset$). As the game unfolds, the partial history $h$ that has just occurred becomes public information and perfectly recalled by all players. At some stages there may be simultaneous moves. For each partial history $h \in \mathcal{H}$ and player $i \in N$, let $A_i(h)$ denote the (finite) set of actions available to player $i$ at history $h$, and let $A(h) = \times_{i \in N} A_i(h)$ and $A_{-i}(h) = \times_{j \in N \setminus \{i\}} A_j(h)$. Without loss of generality, $A_i(h)$ is assumed non-empty for each $h$: player $i$ is inactive at $h$ if $|A_i(h)| = 1$; he is active otherwise. If there is only one active player at each $h$, the game has perfect information.

Pure strategies of player $i$ assign to each partial history $h \in \mathcal{H}$ an action in the set $A_i(h)$. Let $S_i$ denote the set of reduced strategies (plans of actions) of player $i$. Two strategies correspond to the same reduced strategy $s_i \in S_i$ if and only if they are realization-equivalent to $s_i$, that is, they preclude the same collection of histories and for every non precluded history $h$ they select the same action, $s_i(h)$. Each profile of reduced strategies $s$ induces a unique terminal history $z(s) \in \mathcal{Z}$. For each $h \in \mathcal{H}$, let $S_i(h)$ be the set of $s_i$ allowing history $h$ (meaning that there is some $s_{-i}$ such that $h$ is a prefix of $z(s_i,s_{-i})$). Hence, $S_i(\emptyset) = S_i$. $\mathcal{H}(s_i)$ is the set of histories not precluded by $s_i$: $\mathcal{H}(s_i) = \{h \in \mathcal{H} : s_i \in S_i(h)\}$.

---

Formally: $A_i(h) = \{a_i \in A_i : \exists a_{-i} \in A_{-i} \text{ s.t. } (h, (a_i, a_{-i})) \in \mathcal{H} \cup \mathcal{Z}\}$

In the classical equilibrium approach, strategies specify a player’s behavior also at histories precluded by the strategy itself: in the equilibrium, this counterfactual behavior represents the opponents’ beliefs about $i$’s behavior in case he has deviated from the strategy (see Rubinstein, 1991). In this paper a non-equilibrium approach is considered, and players’ beliefs about the opponents’ behavior are modelled explicitly. We can thus restrict attention to players’ plans of actions (or reduced strategies).
Information Structures. Players have preferences over the terminal histories, represented by payoff functions \( u_i : Z \rightarrow \mathbb{R} \) for each \( i \). To model incomplete information, payoff functions are parametrized on a fundamental space of uncertainty \( \Theta^* \) such that

\[
\Theta^* = \Theta^*_0 \times \Theta^*_1 \times \ldots \times \Theta^*_n,
\]

and \( u_i : Z \times \Theta^* \rightarrow \mathbb{R} \). Elements of \( \Theta^* \) are referred to as payoff states. For each \( i = 1, \ldots, n \), \( \Theta^*_i \) is the set of player \( i \)'s payoff types, i.e. possible information that player \( i \) may have about the payoff state. \( \Theta^*_0 \) instead represents the states of nature (any residual uncertainty that is left after pooling all players’ information). Each \( \Theta^*_k \) \((k = 0, 1, \ldots, n)\) is assumed non-empty, Polish and convex, and each \( u_i : Z \times \Theta^* \rightarrow \mathbb{R} \) continuous.\(^{12}\) (For each \( i \), \( \Theta^*_{-i} = \times_{j \in N \setminus \{i\}} \Theta^*_j \), so that \( \Theta^* = \Theta^*_0 \times \Theta^*_i \times \Theta^*_{-i} \).) The tuple \( \langle \Theta^*, (u_i)_{i \in N} \rangle \) represents the fundamental information structure, and describes players’ information about everyone’s payoffs. For ease of reference, two special cases are defined:

**Definition 3.3.** Information structure \( \langle \Theta^*, (u_i)_{i \in N} \rangle \) has Private Values (PV) if for each \( i \in N \), \( u_i : Z \times \Theta^*_i \rightarrow \mathbb{R} \).

**Definition 3.4.** Information structure \( \langle \Theta^*, (u_i)_{i \in N} \rangle \) has No Information (NI) if for each \( i \in N \), \( u_i : Z \times \Theta^*_0 \rightarrow \mathbb{R} \).

In PV-structures (or environments) it is common knowledge that every agent knows his own preferences over the terminal nodes; in NI-environments instead it is common knowledge that players have no information about payoffs.

---

\(^{11}\) This representation is without loss of generality: For example, taking the underlying space of uncertainty \( \Theta^* \equiv ([0, 1]^n)^Z \) imposes no restrictions on agents’ preferences.

\(^{12}\) Convexity of \( \Theta^*_i \) \((i = 1, \ldots, n)\) is only required for the results in section 3.6. Convexity of \( \Theta^*_0 \) can be dropped without affecting any result.
The set $\Theta^*$ should be thought of as being a “rich” space of uncertainty, so that imposing common knowledge of $\Theta^*$ entails sufficiently little loss of generality (see Section 3.2): $\Theta^*$ represents the benchmark in which “no common knowledge assumptions are imposed”, hence the term “fundamental” information structure. Stronger common knowledge assumptions can be imposed considering smaller information structures: tuples $\langle \Theta, (u_i)_{i \in N} \rangle$ such that $\Theta = \times_{k=0}^n \Theta_k$ and $\Theta_k \subseteq \Theta^*_k$ is Polish for each $k$ are referred to as $(\Theta^*$-based) information structures.

**Type Spaces and Hierarchies of Beliefs.** Given an information structure

$$\langle \Theta, (u_i)_{i \in N} \rangle,$$

the description of the game is completed by a description of agents’ hierarchies of beliefs. Players’ hierarchies are defined as follows: for each $i \in N$ let $Z^0_i = \Theta_0 \times \Theta_{-i}$ and for $k \geq 1$ define $Z^k_i = Z^{k-1}_i \times \Delta(Z^{k-1}_{-i})$. An element of $\Delta(Z^{k-1}_i)$ is a $\Theta$-based $k$-order belief, one of $\Theta_i \times (\times_{k\geq 1} \Delta(Z^{k-1}_i))$ a $\Theta$-hierarchy: the first component (in $\Theta_i$) represents player $i$’s information, and the second (in $\times_{k\geq 1} \Delta(Z^{k-1}_i)$) represents his higher order beliefs. For each $i$, let $T^*_i$ denote the set of $i$’s collectively coherent $\Theta$-hierarchies (Mertens and Zamir, 1985). Players’ $\Theta$-hierarchies are represented by means of type spaces:

**Definition 3.5 (Theta-based Type Space).** A $(\Theta$-based) type space is a tuple $T = \langle \Theta, (T_i, \theta_i, \tau_i)_{i \in N} \rangle$ such that for each $i \in N$, $T_i$ is a compact set of types, $\theta_i : T_i \to \Theta_i$ (onto and continuous) assigns to each type a payoff type and $\tau_i : T_i \to \Delta(\Theta_0 \times T_{-i})$ (continuous) assigns to each type a belief about the states of nature and the opponents’ types.

Each type in a type space induces a $\Theta$-hierarchy: For each $t_i \in T_i$, let $\hat{\pi}^0_i (t_i) = \theta_i (t_i)$; construct mappings $\hat{\pi}^k_i : T_i \to \Delta(Z^{k-1}_i)$ recursively for all $i \in I$ and $k \geq 1$, s.t.
\( \hat{\pi}_i(t_i) \) is the pushforward of \( \tau_i(t_i) \) given by the map from \( \Theta \times T_{-i} \) to \( Z_i^0 \) such that
\[
(\theta_0, t_{-i}) \mapsto (\theta_0, \hat{\pi}_{-i}^0(t_{-i})) ,
\]
and \( \hat{\pi}_i^k(t_i) \) is the pushforward of \( \tau_i(t_i) \) given by the map from \( \Theta_0 \times T_{-i} \) to \( Z_i^{k-1} \) such that
\[
(\theta_0, t_{-i}) \mapsto (\theta_0, \hat{\pi}_{-i}^0(t_{-i}), \hat{\pi}_{-i}^1(t_{-i}), \ldots, \hat{\pi}_{-i}^{k-1}(t_{-i})). \tag{13}
\]
The maps \( \hat{\pi}_i^* : T_i \to T_i^* \) defined as
\[
t_i \mapsto \hat{\pi}_i^*(t_i) = (\hat{\pi}_i^0(t_i), \hat{\pi}_i^1(t_i), \hat{\pi}_i^2(t_i), \ldots) ,
\]
assign to each type in a \( \Theta \)-based type space, the corresponding \( \Theta \)-hierarchy of beliefs.

From Mertens and Zamir (1985) and Brandenburger and Dekel (1993) we know that when the sets \( T_i^* \; (i = 1, \ldots, n) \) are endowed with the product topology, there is a homeomorphism
\[
\phi_i : T_i^* \Theta \to \Delta \left( \Theta_{-i} \times T_{-i}^* \Theta \right)
\]
that preserves beliefs of all orders: for all \( t_i^* = (\theta_i, \pi_i^1, \pi_i^2, \ldots) \in T_i^* \Theta, \)
\[
marg_{Z_i^{k-1}} \phi_i(t_i^*) = \pi_i^k \quad \forall k \geq 1.
\]
Hence, the tuple \( T^*_\Theta = \left( \Theta, (T_i^* \Theta, \theta_i^*, \tau_i^*) \right)_{i \in N} \) with \( \tau_i^* = \phi_i \) is a type space. It will be referred to as the \( \Theta \)-based Universal Type Space. The maps \( \hat{\pi}_i^* : T_i \to T_i^* \) constitute the canonical belief morphism from \( T_\Theta \) to \( T^*_\Theta \).

A (\( \Theta \)-based) finite type is any element \( t_i \in T_i^* \Theta \) such that \( \left| \text{supp} \; \tau_i^*(t_i) \right| < \infty \). The set of finite types is denoted by \( \hat{T}_{\Theta} \subseteq \hat{T}_{\Theta}^* \).

\[\text{For measurable spaces } X \text{ and } Y, \text{ measure } \mu \in \Delta(X) \text{ and measurable map } Q : X \to Y, \text{ the pushforward of } \mu \text{ under } Q \text{ is } \mu' \in \Delta(Y) \text{ such that for every measurable } E \subseteq Y, \mu'[E] = \mu\left[ Q^{-1}(E) \right]. \]
Models. An information structure and a type space complete the description of a model:

**Definition 3.6.** A model is a tuple $\mathcal{M} = \langle \Theta, \mathcal{T}_\Theta, (\hat{u}_i)_{i \in N} \rangle$ such that: (i) $\langle \Theta, (u_i)_{i \in N} \rangle$ is a $\Theta^*$-based information structure; (ii) $\mathcal{T}_\Theta$ is a $\Theta$-based type space, and (iii) for each $i$, $\hat{u}_i : \mathcal{Z} \times \Theta_0 \times T \to \mathbb{R}$ are such that for each $(z, \theta_0, t) \in \mathcal{Z} \times \Theta_0 \times T$, $\hat{u}_i (z, \theta_0, t) = u_i (z, \theta_0, \theta (t))$. A model is finite if $|\Theta_0 \times T|$ is finite. The pair $\langle \Theta^*, \mathcal{T}^*_{\Theta^*}, (\hat{u}_i)_{i \in N} \rangle$ is referred to as the Universal Model.

As discussed in Section 3.2, the universal model represents the benchmark in which all common knowledge assumptions are being relaxed.\footnote{As discussed in Section 3.3, the notion of richness that is adopted qualifies what common knowledge assumptions are being relaxed. If $\Theta^* = \left( [0,1]^Z \right)^n$, with $u(\theta, z) = \theta (z)$, the only common knowledge restrictions is that agents are von Neumann-Morgenstern expected utility maximizers. (This specification of $\Theta^*$ satisfies the richness condition introduced in Section 3.6.)} Any smaller model imposes common knowledge assumptions: adopting a $\Theta^*$-based non universal type space (a model $\langle \Theta^*, \mathcal{T}^*_{\Theta^*}, (\hat{u}_i)_{i \in N} \rangle$) imposes common knowledge assumptions on agents’ hierarchies of beliefs; adopting a smaller information structure (a model $\langle \Theta, \mathcal{T}_\Theta, (\hat{u}_i)_{i \in N} \rangle$, $\Theta_k \subset \Theta^*_k$ for each $k$) entails extra common knowledge assumptions on payoffs.

Attaching a model $\mathcal{M} = \langle \Theta, \mathcal{T}_\Theta, (\hat{u}_i)_{i \in N} \rangle$ to the extensive form $\Gamma$ delivers a multistage game with observable actions:

$$\Gamma^\mathcal{M} = \langle N, \mathcal{H}, \mathcal{Z}, \Theta, (T_i, \theta_i, \tau_i, \hat{u}_i)_{i \in N} \rangle.$$  

(For simplicity of notation, the reference to the payoff functions will be omitted in the following, and models written as $\langle \Theta, \mathcal{T}_\Theta \rangle$.)

98
3.4 Interim Sequential Rationalizability

Fix a model $\mathcal{M} = \langle \Theta, T, (\hat{u}_i)_{i \in N} \rangle$, and consider the induced game $\Gamma^\mathcal{M}$. As the game unfolds, players form conjectures about their opponents’ behavior, their types and the state of nature, represented by conditional probability systems (CPS), i.e. arrays of conditional beliefs, one for each history, (denoted by $\mu^i = (\mu^i(h))_{h \in \mathcal{H}} \in \Delta^\mathcal{H} (\Theta_0 \times T_{-i} \times S_{-i})$) such that: (i) $\mu^i$ is consistent with Bayes’ rule whenever possible, and (ii) for each $h \in \mathcal{H}$, $\mu^i(h) \in \Delta (\Theta_0 \times T_{-i} \times S_{-i}(h))$.\(^{15}\) To avoid confusion, we refer to this kind of beliefs as “conjectures”, retaining the term “beliefs” only for those represented in the models introduced in Section 3.3.

**Conjectures.** Agents entertain conjectures about the space $\Theta_0^* \times T_{-i} \times S_{-i}$. As the game unfolds, and agents observe public histories, their conjectures change. For each history $h \in \mathcal{H}$, define the event $[h] \subseteq \Theta_0^* \times T_{-i} \times S$ as:

$$[h] = \Theta_0^* \times T_{-i} \times S_{-i}(h).$$

(Notice that, by definition, $[h] \subseteq [h']$ whenever $h'$ is a predecessor of $h$.)

**Definition 3.7.** A conjecture for agent $i$ is a conditional probability system (CPS hereafter), that is a collection $\mu^i = (\mu^i(h))_{h \in \mathcal{H}}$ of conditional distributions $\mu^i(h) \in \Delta (\Theta_0^* \times T_{-i} \times S_{-i})$ that satisfy the following conditions:

- **C.1** For all $h \in \mathcal{H}$, $\mu^i(h) \in \Delta ([h])$;

- **C.2** For every measurable $A \subseteq [h] \subseteq [h']$, $\mu^i(h) \cdot [A] \cdot \mu^i(h') \cdot [h] = \mu^i(h') \cdot [A]$.

For each type $t_i \in T_i$, his consistent conjectures are

$$\Phi^T_i (t_i) = \{ \mu^i \in \Delta^\mathcal{H} (\Theta_0 \times T_{-i} \times S_{-i}) : \text{marg}_{\Theta_0 \times T_{-i}} \mu^i(\phi) = \tau_i(t_i) \}.$$  

\(^{15}\)See Battigalli and Siniscalchi (2007).
Condition C.1 states that agents' are always certain of what they know, i.e. the observed public history; condition C.2 states that agents’ conjectures are consistent with Bayesian updating whenever possible. Type \( t_i \)'s consistent conjectures agree with his beliefs on the environment at the beginning of the game.

**Sequential Rationality.** The set of Sequential Best Responses for type \( t_i \) to conjectures \( \mu^i \in \Delta^H (\Theta_0 \times T_{-i} \times S_{-i}) \), denoted by \( r_i (\mu^i|t_i) \), is defined as:

\[
\begin{align*}
  s_i \in r_i (\mu^i|t_i) & \text{ if and only if } \forall h \in H (s_i) \\
  s_i & \in \arg \max_{s' \in S(h)} \int_{\Theta_0 \times T_{-i} \times S_{-i}} \hat{u}_i (\mathbf{z} (s_i, s_{-i}), \theta_0, t_{-i}, t_i) \, d\mu^i (h)
\end{align*}
\]  

(3.1)

**Definition 3.8.** A strategy \( s_i \in S_i \) is sequentially rational for type \( t_i \), written \( s_i \in r_i (t_i) \), if there exists \( \mu^i \in \Phi^T_i (t_i) \) such that \( s_i \in r_i (\mu^i|t_i) \).

The notion of sequential rationality is stronger than (normal-form) rationality, which only requires that a player optimizes with respect to his initial conjectures, hence putting no restrictions on behavior at zero-probability histories. Notice also that working with reduced strategies, the restrictions in equation (3.1) only concern histories that are reachable by \( s_i \).

**Interim Sequential Rationalizability.**

**Definition 3.9 (ISR).** Fix a \( \Theta \)-based model \( T \). For each \( i \in N \), let \( \mathcal{ISR}^{T,0}_i = T_i \times S_i \).

For each \( k = 0, 1, \ldots \), and \( t_i \in T_i \), let \( \mathcal{ISR}^{T,k}_i (t_i) = \left\{ s_i \in S_i : (t_i, s_i) \in \mathcal{ISR}^{T,k}_i \right\} \), \( \mathcal{ISR}^{T,k}_i = \times_{i=1,\ldots,n} \mathcal{ISR}^{T,k}_i \) and \( \mathcal{ISR}^{T,k}_{-i} = \times_{j \neq i,0} \mathcal{ISR}^{T,k}_j \). Define recursively, for \( k = 1, 2, \ldots \), and \( t_i \in T_i \)

\[
\mathcal{ISR}^{T,k}_i (t_i) = \left\{ \hat{s}_i \in \mathcal{ISR}^{T,k-1}_i (t_i) : \begin{array}{ll}
  \exists \mu^i \in \Phi^T_i (t_i) & \text{s.t.} \\
  (1) & \hat{s}_i \in r_i (\mu^i|t_i) \\
  (2) & \text{supp} (\mu^i (\phi)) \subseteq \Theta_0 \times \mathcal{ISR}^{T,k-1}_{-i}
\end{array} \right\}
\]
Finally: \( \mathcal{ISR}^T := \bigcap_{k \geq 0} \mathcal{ISR}^{T,k} \)

\( \mathcal{ISR} \) consists of an iterated deletion procedure: for each type \( t_i \), reduced strategy \( s_i \) survives the \( k \)-th round of deletion if and only if \( s_i \) is sequentially rational for type \( t_i \), with respect to conjectures \( \mu^i \) that, at the beginning of the game, are concentrated on pairs \((t_{-i}, s_{-i})\) consistent with the previous rounds of deletion. If history \( h \) is given zero probability by the conditional conjectures held the preceding node, \( i \)'s conjectures at \( h \) may be concentrated anywhere in \( \Theta_0 \times T_{-i} \times S_{-i}(h) \). The lack of restrictions on the conjectures held at unexpected histories rules out elements of forward induction reasoning (see example 3.1).\(^{16}\) Notice also that \( \mathcal{ISR} \) considers players’ conjectures that allow for correlation in the opponents’ strategies and \( \theta_0 \) in the Bayesian game.\(^{17}\) A more thorough discussion of the solution concept and its relation with the literature is postponed to section 3.7.

**Example 3.5:** Example 3.1 illustrated the basic logic of \( \mathcal{ISR} \). We repeat the argument here to familiarize with the procedure and the notation: \( a_3 \) is dominated by \( a_1 \), hence it is deleted at the first round. Given this, \( \mathcal{ISR} \) restricts 2’s initial conjectures to put zero probability on \( a_3 \), that is \( \mu^2(\phi)[a_3] = 0 \). No further restrictions are imposed: In particular, conjectures \( \hat{\mu}^2 \) can be such that \( \hat{\mu}^2(\phi)[a_1] = 1 \) and \( \hat{\mu}^2(In)[a_3] = 1 \), which makes \( b_2 \) the unique sequential best response to \( \hat{\mu}^2 \). Given that neither \( b_1 \) nor \( b_2 \) is deleted at the second round, also \( a_1 \) cannot be deleted, and the procedure stops at \( \mathcal{ISR} = \{a_1, a_2\} \times \{b_1, b_2\} \). □

\(^{16}\)An epistemic characterization for \( \mathcal{ISR} \) in finite models is provided in Penta (2009b). The epistemic conditions are “sequential rationality” and “initial common certainty of sequential rationality”.

\(^{17}\)Similarly to the distinction between interim independent and correlated rationalizability (Dekel et al., 2007), one could think of refining \( \mathcal{ISR} \) so that players’ conjectures on the opponents’ behavior are measurable with respect to their types. Given the results in section 3.5, such refinement would not be robust.
3.4.1 Example: Finitely Repeated Prisoner’s Dilemma

In games with complete information, it can be shown that $\mathcal{ISR}$ can be computed applying Dekel and Fudenberg’s (1990) $S^\infty W$-procedure to the normal form of the game: That is, one round of deletion of weakly dominated strategies, followed by iterated strict dominance.\(^{18}\) As in Dekel and Fudenberg (1990, Section 5), it is instructive to discuss how the procedure applies to the finitely repeated prisoners dilemma: Let stage payoffs be as in the following table, and the game be repeated a finite number $T$ of times (players sum payoffs over periods, with no discounting):

<table>
<thead>
<tr>
<th></th>
<th>$C$</th>
<th>$S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td>10,10</td>
<td>0,11</td>
</tr>
<tr>
<td>$S$</td>
<td>11,0</td>
<td>1,1</td>
</tr>
</tbody>
</table>

For any $T < \infty$, all the strategies that prescribe $C$ at the last stage are deleted at the first round, because they are not sequentially rational: for any conditional beliefs held at any history of length $T$, action $C$ is dominated, hence all strategies prescribing that behavior at some of these node are not sequentially rational. (In terms of the normal form, these are weakly dominated strategies.) If $T = 2$, after the first round of deletion, only two strategies survive for each player:

$$\mathcal{ISR}_i^1 = \{\langle S;SSSS \rangle, \langle C;SSSS \rangle\}$$

It is immediate to see that at this point strategy $\langle C;SSSS \rangle$ cannot be a sequential best response to any conjecture $\mu^i$ such that $\mu^i(\phi) \in \Delta (\mathcal{ISR}_i^1)$. (In terms of the normal form, strategies $\langle C;SSSS \rangle$ are strictly dominated in $\mathcal{ISR}_i^1$.) Hence, at the second round, $\langle C;SSSS \rangle$ is deleted, and $\mathcal{ISR}_i = \{\langle S;SSSS \rangle\}$: In the unique $\mathcal{ISR}$-outcome both players always Shirk.

\(^{18}\)Proposition A.1 in Appendix A.3 shows that in private-values environments with payoffs in generic position, $\mathcal{ISR}$ can be computed applying the $S^\infty W$-procedure to the interim (reduced) normal form of the game.
Things change if $T \geq 3$: In this case, the deletion procedure stops before delivering the subgame perfect solution. Suppose that $T = 3$. Let $(a_i^1 | a_i^{CC} a_i^{CS} a_i^{SC} a_i^{SS} | a_i^3)$ denote player $i$’s strategy that plays $a_i^1 \in \{C, S\}$ in the first period, $a_i^{a_1 a_2} \in \{C, S\}$ in the second round if $(a_i^1 a_i^2)$ was played in the first period, and action $a_i^3 \in \{C, S\}$ in the last period, irrespective of the history. As argued above, after the first round of deletion only strategies such that $a_i^3 = S$ survive. Hence, after the first round, we have
\[
\mathcal{ISR}^1 = \{ (a_i^1 | a_i^{CC} a_i^{CS} a_i^{SC} a_i^{SS} | S) : (a_i^1, a_i^{CC}, a_i^{CS}, a_i^{SC}, a_i^{SS}) \in \{C, S\}^5 \}
\]
Applying the $S^\infty W$-procedure, at the second round strategies that in the second period “cooperate no-matter-what” (i.e. strategies $(a_i^1 | C^{CC} C^{CS} C^{SC} C^{SS} | S)$) are deleted, because they are strictly dominated by the strategy that only replaces the second period’s behavior with “shirk no-matter-what” (i.e. $(a_i^1 | S^{CC} S^{CS} S^{SC} S^{SS} | S)$). Hence, the reduced strategies that survive the second round of deletion are:
\[
\mathcal{ISR}^2 = \{ s_1 = C^1 | C^{CC} S^{CS} - | S \}, s_2 = C^1 | S^{CC} C^{CS} - | S \}, \\
\quad s_3 = (C^1 | S^{CC} S^{CS} - | S), s_4 = (S^1 | C^{CC} S^{SS} | S) \}, \\
\quad s_5 = (S^1 | -S^{SC} C^{SS} | S), s_6 = (S^1 | -S^{SC} S^{SS} | S) \}
\]
\[
\mathcal{ISR}^2 = \{ s_1 = C^1 | C^{CC} - S^{SC} - | S \}, s_2 = C^1 | S^{CC} - C^{SC} - | S \}, \\
\quad s_3 = (C^1 | S^{CC} - S^{SC} - | S), s_4 = (S^1 | C^{CC} - S^{SS} | S) \}, \\
\quad s_5 = (S^1 | S^{CS} - C^{SS} | S), s_6 = (S^1 | S^{CS} - S^{SS} | S) \}
\]
The submatrix of the reduced normal form at this point of the procedure is the following $6 \times 6$ matrix (payoffs in bold corresponds to best responses):
It is easy to see that no further strategies are deleted at this point, because none of them is strictly dominated. Hence, \( ISR_i = \{s^1_i, s^2_i, s^3_i, s^4_i, s^5_i, s^6_i\} \). Notice that that if \( s^4_i \) were deleted at this round, iterated deletion of strictly dominated strategies would delete, in order, strategies \( s^1_i \), then \( s^2_i \) and \( s^3_i \), and finally \( s^5_i \), uniquely selecting the subgame perfect outcome \((s^6_1, s^6_2)\). But \( s^4_i \) is not deleted, because it is only weakly dominated, not strictly. To see how \( s^4_i \) can be a sequential best response to conjectures consistent with initial certainty of \( ISR^2 \), consider 1’s conjectures \( \mu^1 \) such that \( \mu^1(\phi)[s^6_2] = 1 \). Clearly, \( s^4_1 \) is a best response to \( \mu^1(h) \) at histories \( h = \phi, SS \) and at any \( h \) in the last stage: In particular, since \( SS \) is consistent with \( s^6_2 \), and 1’s initial beliefs are concentrated on \( s^6_2 \), also at \( SS \) player 1 is certain of \( s^6_2 \). Now, to see how \( s^4_1 \) is a best response at history \( SC \) too (we don’t need to check history \( CS \) because \( CS \notin \mathcal{H}(s^4_1) \)), notice that \( SC \notin \mathcal{H}(s^6_2) \), so receives zero probability under \( \mu^i(\phi) \). Hence, \( ISR \) imposes no restrictions on \( \mu^i(SC) \) (in particular, \( \mu^i(SC) \) need not be concentrated on \( ISR^2 \)). So, set \( \mu^i(SC)[s^6_2] = 1 \) where \( s^*_2 \) is a strategy that prescribes that 2 plays \( C \) in the last stage if and only if 1 plays \( C \) in the second:

<table>
<thead>
<tr>
<th></th>
<th>( s^1_2 )</th>
<th>( s^2_2 )</th>
<th>( s^3_2 )</th>
<th>( s^4_2 )</th>
<th>( s^5_2 )</th>
<th>( s^6_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s^1_1 )</td>
<td>21, 21</td>
<td>11, 22</td>
<td>11, 22</td>
<td>12, 12</td>
<td>2, 13</td>
<td>2, 13</td>
</tr>
<tr>
<td>( s^2_1 )</td>
<td>22, 11</td>
<td>12, 12</td>
<td>12, 12</td>
<td>11, 22</td>
<td>1, 23</td>
<td>1, 23</td>
</tr>
<tr>
<td>( s^3_1 )</td>
<td>22, 11</td>
<td>12, 12</td>
<td>12, 12</td>
<td>12, 12</td>
<td>2, 13</td>
<td>2, 13</td>
</tr>
<tr>
<td>( s^4_1 )</td>
<td>12, 12</td>
<td>22, 11</td>
<td>12, 12</td>
<td>3, 3</td>
<td>13, 2</td>
<td>3, 3</td>
</tr>
<tr>
<td>( s^5_1 )</td>
<td>13, 2</td>
<td>23, 1</td>
<td>13, 2</td>
<td>2, 13</td>
<td>12, 12</td>
<td>2, 13</td>
</tr>
<tr>
<td>( s^6_1 )</td>
<td>13, 2</td>
<td>23, 1</td>
<td>13, 2</td>
<td>3, 3</td>
<td>13, 2</td>
<td>3, 3</td>
</tr>
</tbody>
</table>

Given these conditional beliefs at \( h = SC \), action \( C \) at the second period followed by \( S \) in the last period is indeed a best response at \( h = SC \) for player 1. Notice though that \( s^*_2 \) is a dominated strategy, that had been deleted in the first round of the procedure. But, as discussed above, once players are surprised (i.e. at unexpected histories), \( ISR \) allows them to believe that the opponents may play anything, even
dominated strategies.

### 3.5 Robustness(-es)

This section gathers properties of $\mathcal{ISR}$ that can be interpreted as robustness properties (see section 3.2). These are, respectively: upper hemicontinuity, type space-invariance and model-invariance. The first two hold for any information structure $(\Theta, (u_i)_{i \in N})$, while the latter holds in PV-environments, but not in NI-environments (see example 4).

#### 3.5.1 Upper Hemicontinuity

As discussed in section 3.2, the upper hemicontinuity of $\mathcal{ISR}$ on the universal type space addresses a specific robustness question: the fact that $\mathcal{ISR}$ is u.h.c. means that any behavior that $\mathcal{ISR}$ rules out for a given $\Theta$-hierarchy is also ruled out for all nearby hierarchies. Specifically, in the product topology, suppose that as analysts we know players hierarchies only up to a finite order $k$: if a solution concept is not u.h.c. it means that we can never rule out that by refining our model of the beliefs of order higher than $k$, the solution concept allows behavior that is ruled out in the original model.$^{19}$

**Proposition 3.1.** For each $t \in T^*_\Theta$ and sequence $\{t^m\} : t^m \rightarrow t$ and for $\{s^m\} \subseteq S$ s.t. $s^m \rightarrow \hat{s}$ and $s^m \in \mathcal{ISR}^*(t^m)$ for all $m$, $\hat{s} \in \mathcal{ISR}^*(t)$.

**Proof.** (See appendix A.1) \hfill $\Box$

$^{19}$Weinstein and Yildiz (2007) extensively discuss the interpretation of the product topology.
3.5.2 Type Space Invariance

The problem of type space-dependence (Dekel et al., 2007) originates in the different possibility of correlation between types and payoff states $\theta_0$ that different type spaces, representing the same set of hierarchies of beliefs, may allow. If a solution concept imposes condition of conditional independence on the agents’ conjectures about the opponents’ strategies, these differences in the possibility of correlation built in the type space may affect the predictions of a solution concept, and the strategies selected for a type may not depend on the hierarchy of beliefs only.\footnote{The refinement of $\cal I S R$ mentioned in footnote 17 would not be type space-invariant.} The intuition behind the type space-invariance of $\cal I S R$ is the same as for that proved by Dekel et al. (2007) for \textit{interim correlated rationalizability (ICR)}: Solution concepts, such as $ICR$ and $\cal I S R$, that do not impose any condition of independence on players’ conjectures about the opponents’ strategies, already allow for all the possible correlation and therefore are not affected by the differences across type spaces discussed above. Hence, they are type space-invariant:

**Proposition 3.2.** Let $\cal T$ and $\tilde{\cal T}$ be two $\Theta$-based type spaces. If $t_i \in T_i, \tilde{t}_i \in \tilde{T}_i$ are s.t. $\hat{\pi}^*(t_i) = \hat{\pi}^*(\tilde{t}_i) \in T_i^* \Theta$, then $ISR^T(t_i) = ISR^{\tilde{T}}(\tilde{t}_i)$. Indeed, for each $k$, if for all $l \leq k$, $\hat{\pi}^l(t_i) = \hat{\pi}^l(\tilde{t}_i)$ then $ISR^T_{i,k-1}(t_i) = ISR^{\tilde{T},k-1}_{i}(\tilde{t}_i)$.

**Proof.** (See appendix A.1)

3.5.3 Model Invariance

The failure of model-invariance for $\cal I S R$ in NI-environments was shown in example 4. The reason behind that failure is the following: The only restrictions that $\cal I S R$ puts on the conjectures held at zero-probability histories come from the definition of the
model (that they be concentrated on $\Theta_0 \times T_{-i}$). Once surprised, type $t_i$ may assign positive probability to pairs $(\theta_0, t_{-i})$ that were initially given zero probability by the beliefs $\tau_i(t_i)$. Moving from smaller to larger models (e.g. from a $\Theta$-based type space, to a $\Theta^*$-based type space) puts less and less restrictions on such possible conjectures.

In $NI$-environments, larger $\Theta_0$ also means more freedom to specify a player’s beliefs about his own payoffs, thereby changing the set of sequential best responses. In contrast, in $PV$-environments players know their own payoffs: Even if their beliefs are completely upset, they don’t alter a type’s preferences over the terminal nodes. This provides the intuition for the model-invariance result in PV-settings.

Let $(\Theta, T_{\Theta}, (\hat{u}_i)_{i \in N})$ be a model. Any type $t_i \in T_{i, \Theta}$ induces a $\Theta$-hierarchy $\hat{\pi}_i^\ast (t_i) = (\theta_i(t_i), \hat{\pi}_i^1(t_i), ...) \in T_{i, \Theta}^\ast$. Since $\Theta = \times_{k=0}^n \Theta_k$ is such that $\Theta_k \subseteq \Theta_k^*$ for all $k$, $\hat{\pi}_i^* (t_i)$ can be naturally embedded in the $\Theta^*$-based universal type space, and be seen as a $\Theta^*$-based hierarchy. Let $\beta_i : T_{i, \Theta}^\ast \rightarrow T_{i, \Theta^*}^\ast$ denote such embedding and let $\kappa_i^* \equiv \beta_i \circ \hat{\pi}_i^*$. 

**Proposition 3.3.** Assume private values: For any finite model $(\Theta, T_{\Theta}, (\hat{u}_i)_{i \in N})$ and any type $t_i \in T_{i, \Theta}$, \( I\mathcal{S}\mathcal{R}_{t_i}^{T_{\Theta}} (t_i) = I\mathcal{S}\mathcal{R}_{t_i}^{T_{\Theta^*}} (\kappa_i^* (t_i)) \).

**Proof.** (See appendix A.1)

### 3.6 The structure of $I\mathcal{S}\mathcal{R}$ in the Universal Model

In Section 3.3 we informally introduced the universal model $(\Theta^*, T_{\Theta^*}^\ast)$ as representing the benchmark in which all common knowledge assumptions are being relaxed: if $\Theta^*$ is “sufficiently rich”, imposing common knowledge of $(\Theta^*, T_{\Theta^*}^\ast)$ entails essentially no loss of generality. The notion of “richness” therefore qualifies the common knowledge assumptions that are being relaxed: the richer $\Theta^*$, the less restrictive the assumption that $\Theta^*$ is common knowledge. The richness condition warrants that $\Theta^*$ is sufficiently

107
rich that, as $\theta$ varies, agents’ preferences vary enough.

**Definition 3.10.** Strategy $s_i$ is conditionally dominant at $\theta \in \Theta^*$ if $\forall h \in \mathcal{H}(s_i)$, $\forall s'_i \in S_i(h)$, $\forall s_{-i} \in S_{-i}(h)$

$$s_i(h) \neq s'_i(h) \Rightarrow u_i(z(s_i,s_{-i}),\theta) > u_i(z(s'_i,s_{-i}),\theta)$$

**Richness Condition (RC):** $\forall s \in S$, $\exists \theta^* = (\theta^*_0, \theta^*_i, \theta^*_{-i}) \in \Theta^*$: $\forall i \in N$, $s_i$ is conditionally dominant at $\theta^*$.

The main result in this section states that whenever a type profile $\hat{\tau}$ has multiple ISR-outcomes, any of these is uniquely ISR for a sequence of players’ types converging to $\hat{\tau}$ (Proposition 3.4). An immediate implication is that any refinement of ISR (e.g. extensive form rationalizability, or sequential equilibrium) is not robust (See example 3.2 in Section 3.2).

**Remark 3.1.** In the finitely repeated prisoner’s dilemma (Section 3.4.1), Proposition 3.4 implies that there exists a model of beliefs, arbitrarily close to the complete information benchmark, in which players uniquely play the grim-trigger strategy $s^1_i$: players cooperate in the first period, and shirk in the last, but they cooperate in the second period if and only if both players behaved cooperatively in the first period. The induced outcome is mutual cooperation in the first two periods, followed by mutual shirk in the last period.

Proposition 3.4 is analogous to Proposition 1 in Weinstein and Yildiz (2007), who proved a similar structure for interim correlated rationalizability. Weinstein and Yildiz’s analysis is based on a condition that assumes the existence of strict dominance states for each of the players’ strategies. This condition cannot be satisfied by the reduced normal form of a dynamic game, in which payoffs are defined over the terminal
histories. As Weinstein and Yildiz point out, an obvious candidate to overcome the problem is to introduce trembles in the solution concept, maintaining a normal form approach. But introducing trembles changes the game, and it is not clear whether the results on the structure of such “modified” rationalizability may be used to address questions of robustness for refinements of rationalizability (without trembles). Besides that, a normal form approach has the further drawback of overshadowing important issues raised by dynamic environments, such as the implications of the modeling assumptions on the agents’ information discussed above. The extensive form approach of this paper overcomes both these shortcomings.

3.6.1 Sensitivity of Multiplicity to higher order beliefs

Given the result of type space-invariance (Proposition 3.2), the reference to the type space in the notation $\mathcal{ISR}^{\Theta^*}$ is omitted in the following.

**Proposition 3.4.** Under the richness condition, for any finite type profile $\hat{t} \in \hat{T}_\Theta^*$ and any $s \in \mathcal{ISR}(\hat{t})$, there exists a sequence of finite type profiles $\{\hat{t}^m\} \subseteq \hat{T}_\Theta^*$ s.t. $\hat{t}^m \rightarrow \hat{t}$ as $m \rightarrow \infty$ and $\mathcal{ISR}(\hat{t}^m) = \{s\}$ for each $m$.

Furthermore, for each $m$, $\hat{t}^m$ belongs to a finite belief-closed subset of types, $T^m \subseteq T^*_{\Theta^*}$, such that for each $m$ and each $t \in T^m$, $|\mathcal{ISR}(t)| = 1$.

Proposition 3.4 implies that any refinement of $\mathcal{ISR}$ is not u.h.c.; since $\mathcal{ISR}$ is u.h.c. by Proposition 3.1, the following is true:

**Corollary 3.1.** $\mathcal{ISR}$ is a strongest upper hemicontinuous solution concept.

The last part of the proposition, stating that the types in the sequence belong to a finite, belief-closed subset of types, is of particular interest in the context of PV-environments, in which $\mathcal{ISR}$ is also model-invariant, which means that such belief-closed sets of types can be considered as models, in the sense of definition 3.6.
The proof of proposition 3.4 requires a substantial investment in additional concepts and notation. To facilitate the reader that is not interested in the technicalities, all these are confined to the next subsection. The argument is only sketched here.

The proof exploits a refinement of $\mathcal{ISR}$, $\mathcal{SSR}$, in which strategies that are never \textit{strict} sequential best responses are deleted at each round. The proof is articulated in two main steps: in the first (lemma 3.2), it is shown that if $s_i \in \mathcal{ISR}_i(t_i)$ for finite type $t_i$, then $s_i$ is also $\mathcal{SSR}$ for some type close to $t_i$; in the second (lemma 3.3), it is shown that by perturbing beliefs further, any $s_i \in \mathcal{SSR}_i(t'_i)$ can be made uniquely $\mathcal{ISR}$ for a type close to $t'_i$. The main points of departure from the analysis of Weinstein and Yildiz (2007) are due to the necessity of breaking the ties between strategies at unreached information sets: this is necessary to obtain uniqueness in the converging sequence. Although notationally involved, the idea is very simple. Consider the sequence constructed in example 2: to obtain $b_2$ as the unique $\mathcal{ISR}$ for player 2, given that 1 would play $a_1$, it was necessary to perturb player 2’s beliefs introducing, with arbitrarily small probability, the possibility that he is facing types $1^3, 3^3, \ldots$, i.e. types who are certain that $a_3$ is dominant. These “dominance types” play the role of trembles, and allow to break the tie between $b_1$ and $b_2$; but they do this in an indirect way, through the perturbation of the belief structure, thus avoiding the shortcomings of the tremble-based approach discussed above.

\textbf{Proof of Proposition 3.4}

\textbf{Definition 3.11.} Let $\mathcal{SSR}_i^0 = T_{i, \varnothing^*} \times S_i$. For each $k = 0, 1, \ldots$, and $t_i \in T_{i, \varnothing^*}$; let $\mathcal{SSR}_i^k (t_i) = \{ s_i \in S_i : (t_i, s_i) \in \mathcal{SSR}_i^k \}$, $\mathcal{SSR}^k = \times_{i=1, \ldots, n} \mathcal{SSR}_i^k$ and $\mathcal{SSR}_{-i}^k =$
Define recursively, for \( k = 1, 2, \ldots \), and \( t_i \in T_i \)

\[
\mathcal{SSR}_i^k (t_i) = \begin{cases} 
    \hat{s}_i \in \mathcal{SSR}_i^k (t_i) : \\
    \exists \mu^i \in \Phi_i (t_i) \text{ s.t. } \\
    (1). r_i (\mu^i | t_i) = \{ \hat{s}_i \} \\
    (2). \supp(\mu^i (\phi)) \subseteq \Theta_0^* \times \mathcal{SSR}_i^{k-1} (t_i) \\
    (3). \text{if } t_{-i} \in \supp(\text{marg}_{T_{-i}^*}, \mu^i (\phi)) \text{ and } s_{-i} \in \mathcal{SSR}_i^{k-1} (t_{-i}), \text{ then: } s_{-i} \in \supp(\text{marg}_{S_{-i}} \mu^i (\phi)) 
\end{cases}
\]

Finally: \( \mathcal{SSR} = \bigcap_{k \geq 0} \mathcal{SSR}^k \)

The following lemma states the standard fixed-point property for \( \mathcal{SSR} \).

**Lemma 3.1.** Let \( \{V_j\}_{j \in \mathbb{N}} \) be s.t. for each \( i \in N \), \( V_i \subseteq S_i \times T_i \) and \( \forall s_i \in V_i (t_i), \exists \mu^i \in \Phi_i (t_i) : \)

- (i) \( \supp(\mu^i (\phi)) \subseteq \times_{j \neq i} V_j \)
- (ii) \( \{ s_i \} = r_i (\mu^i | t_i) \)

Then: \( V_i (t_i) \subseteq \mathcal{SSR}_i (t_i) \)

Exploiting the richness condition, let \( \bar{\Theta} \subset \Theta^* \) be a finite set of dominance states, s.t. \( \forall s \in S, \exists ! \theta^s \in \bar{\Theta} \) at which \( s \) is conditionally dominant. For each \( s \in S \), let \( \bar{t}^s \in T_{\bar{\Theta}}^s \) be s.t. for each \( i \), \( \theta_i (\bar{t}^s) = \theta_i^s \) and \( \tau_i (\bar{t}^s) [\theta_0^s, \bar{t}^s_{-i}] = 1 \). Let \( \bar{T} = \{ \bar{t}^s : s \in S \} \), and let \( \bar{T}_j \) and \( \bar{T}_{-j} \) denote the corresponding projections: \( \bar{T} \) is a finite set of types representing common belief of \( \theta^s \), for each \( s \in S \). Elements of \( \bar{T}_i \) will be referred to as *dominance-types*, and will play the role of the \( k^\alpha \)-types in example 3.

For each \( i \) and \( s_i \in S_i \), let \( \bar{T}_{-i} (s_i) \) be s.t. \( \forall s_{-i} \in S_{-i}, \exists ! \bar{t}_{-i} \in \bar{T}_{-i} (s_i) \) s.t. \( \bar{t}_{-i} = \bar{t}_{-i}^{(s_i, s_{-i})} \). Notice that for each \( \bar{t}^s_i \in \bar{T}_i \), \( \{ s_i \} = \mathcal{SSR}_i^1 (\bar{t}^s_i) \), because \( s_i \) is the unique sequential best reply to any conjecture consistent with condition 3 for \( \mathcal{SSR} \) in definition 3.11.
Lemma 3.2. Under the richness condition, for any finite type $t_i \in \hat{T}_i, \Theta^*$, for any $s_i \in ISR_i (t_i)$, there exists a sequence of finite types $\{t^m (t_i, s_i)\}_{m \in \mathbb{N}}$, such that:

- (i) $t^m (t_i, s_i) \rightarrow t_i$ as $m \rightarrow \infty$
- (ii) $\forall m, s_i \in SSR_i (t^m (t_i, s_i))$ and $t^m (t_i, s_i) \in \hat{T}_i$
- (iii) $\forall m$, conjectures $\mu^{s_i, m} \in \Phi (t^m (t_i, s_i))$ s.t. $\{s_i\} = r_i (\mu^{s_i, m} | t^m (t_i, s_i))$ satisfy

$$T_{-i} (s_i) \subseteq \text{supp} \left( \text{marg}_{T_{-i}, \Theta^*} \mu^{s_i, m} (\phi) \right).$$

Proof. (See Appendix A.2.1)

Lemma 3.3. Under the richness condition, for each finite type $\hat{t}_i \in \hat{T}_i, \Theta^*$, for each $k$, for each $s_i \in SSR^k_i (\hat{t}_i)$ such that the conjectures $\mu^{s_i} \in \Phi (\hat{t}_i)$: $\{s_i\} = r_i (\mu^{s_i} | \hat{t}_i)$ satisfy

$$T_{-i} (s_i) \subseteq \text{supp} \left( \text{marg}_{T_{-i}, \Theta^*} \mu^{s_i} (\phi) \right),$$

there exists $\hat{t}_i \in \hat{T}_i$ s.t.

1. For each $k' \leq k$, $\hat{\pi}^{k'} (\hat{t}_i) = \hat{\pi}^{k'} (\hat{t}_i)$
2. $ISR_i^{k+1} (\hat{t}_i) = \{s_i\}$
3. $\hat{t}_i \in \hat{T}^{\hat{t}_i}_i$ for some finite belief closed set of types $\hat{T}^{\hat{t}_i}_i = \times_{j \in \mathbb{N}} \hat{T}^{\hat{t}_i}_j$ such that $|ISR_i^{k+1} (t)| = 1$ for each $t \in \hat{T}^{\hat{t}_i}_i$.

Hence, for any such $s_i \in SSR_i (\hat{t}_i)$ there exists a sequence of finite types $t_{i,m} \rightarrow \hat{t}_i$ s.t. $ISR_i (t_{i,m}) = \{s_i\}$.

Proof. (See Appendix A.2.2.)

112
Given the lemmata above, the proof of proposition 3.4 is immediate:

**Proof of Proposition 3.4:** Take any \( \hat{t} \in \hat{T} \) and any \( s \in ISR(\hat{t}) \). For each \( i \), from lemma 3.2 there exists a sequence \( \{t_i^m\} \subseteq \hat{T}_{i, \Theta^*} \) of finite types s.t. \( t_i^m \rightarrow \hat{t}_i \) and for each \( i \), \( s_i \in SSR_i(t_i^m) \) for each \( m \), for conjectures \( \mu^{s_i} \) as in the thesis of lemma 3.2 and in the hypothesis of lemma 3.3. Then we can apply lemma 3.3 to the types \( t_i^m \) for each \( m \): for \( s_i \in SSR_i(t_i^m) \), for each \( k \), there exists a sequence \( \{\tilde{t}_i^{m,k}\}_{k \in \mathbb{N}} \) s.t. \( \tilde{t}_i^{m,k} \rightarrow t_i^m \) for \( k \rightarrow \infty \) s.t. \( ISR_i(\tilde{t}_i^{m,k}) = \{s_i\} \). Because the universal type-space is metrizable, there exists a sequence \( k_m \rightarrow \infty \) with \( t_i^{m,k_m} \rightarrow \hat{t}_i \). Set \( \hat{t}_i^m = t_i^{m,k_m} \), so that \( \hat{t}_i^m \rightarrow \hat{t}_i \) as \( m \rightarrow \infty \) and \( ISR(\hat{t}_i^m) = \{s\} \) for each \( m \).

### 3.6.2 Genericity of Uniqueness

In this section it is proved that uniqueness holds for an open and dense set of types in the universal type space. The proof exploits the following known result:

**Lemma 3.4** (Mertens and Zamir (1985)). The set \( \hat{T}_{\Theta^*} \) of finite types is dense in \( T_{\Theta^*}^* \), i.e.

\[
T_{\Theta^*}^* = cl\left(\hat{T}_{\Theta^*}\right)
\]

**Proposition 3.5.** Under the richness assumption, the set

\[
\mathcal{U} = \{t \in T^* : |ISR(t)| = 1\}
\]

is open and dense in \( T_{\Theta^*}^* \). Moreover, the unique ISR outcome is locally constant, in the sense that \( \forall t \in \mathcal{U} \) such that \( ISR(t) = \{s\} \), there exists an open neighborhood of types, \( N_\delta(t) \), such that \( ISR(t') = \{s\} \) for all \( t' \in N_\delta(t) \).

**Proof:** \((\mathcal{U} \text{ is dense})\) To show that \( \mathcal{U} \) is dense, notice that by proposition 2, for any for any \( \hat{t} \in \hat{T} \) there exists a sequence \( \{\hat{t}^m\} \subseteq \hat{T}^m \) s.t. \( \hat{t}^m \rightarrow \hat{t} \) and \( ISR(\hat{t}^m) = \{s\} \)
for some \( s \in \mathcal{ISR}(\hat{t}) \). By definition, \( \hat{t}^m \in U \) for each \( m \). Hence, \( \hat{t} \in \text{cl}(U) \), thus \( \hat{T} \subseteq \text{cl}(U) \). But we know that \( \text{cl}(\hat{T}) = T^* \), therefore \( \text{cl}(U) \supseteq \text{cl}(\hat{T}) = T^* \). Hence \( U \) is dense.

\textbf{(U is open and \( \mathcal{ISR} \) locally constant in \( U \))} Since (proposition 2) \( \mathcal{ISR} \) is u.h.c., for each \( t \in U \), there exists a neighborhood \( N_\delta(t) \) s.t. for each \( t' \in N_\delta(t) \), \( \mathcal{ISR}(t') \subseteq \mathcal{ISR}(t) \). Since \( \text{ISR}(t) = \{s\} \) for some \( s \), and \( \text{ISR}(t') \neq \emptyset \), it follows trivially that \( \mathcal{ISR}(t') = \{s\} \), hence \( N_\delta(t) \subseteq U \). Therefore \( U \) is open. By the same token, we also have that \( \mathcal{ISR}(t') = \{s\} \) for all \( t' \in N_\delta(t) \), i.e. the unique \( \mathcal{ISR} \) outcome is locally constant.

\textbf{Corollary 3.2.} Generic uniqueness of \( \mathcal{ISR} \) implies generic uniqueness of any equilibrium refinement. In particular, of any Perfect-Bayesian Equilibrium outcome.

For each \( s \in S \), let \( U^s = \{ t \in \hat{T}_{x^*} : \mathcal{ISR}(t) = \{s\} \} \). From proposition 3.5 we know that these sets are open. Let the boundary be \( \text{bd}(U^s) = \text{cl}(U^s) \setminus U^s \).

\textbf{Corollary 3.3.} Under the richness condition, for each \( t \in \hat{T}_{x^*} : |\mathcal{ISR}(t)| > 1 \) if and only if there exist \( s, s' \in \mathcal{ISR}(t) : s \neq s' \) such that \( t \in \text{bd}(U^s) \cap \text{bd}(U^{s'}) \)

Summing up, the results of this section conclude that \( \mathcal{ISR} \) is a generically unique and locally constant solution concept, that yields multiple solutions at, and only at, the boundaries where the concept changes its prescribed behavior. The structure of \( \mathcal{ISR} \) is therefore analogous to that proved by Weinstein and Yildiz (2007) for ICR.

Given the results from this and the previous section, the analogues of the remaining results in Weinstein and Yildiz (2007) can be obtained in a straightforward manner for \( \mathcal{ISR} \): in particular, proposition 3.4 also holds if one imposes the common prior assumption.
3.7 Related Literature and Concluding Remarks

On NI- and PV-settings. In NI-settings, all common knowledge-assumptions are relaxed. In particular, the assumption that players know their own preferences. Under an equivalent richness condition, independent work by Chen (2009) studied the structure of ICR for dynamic NI-environments, extending Weinstein and Yildiz’s (2007) results. Together with the upper hemicontinuity of ISR, Chen’s results imply that the two solution concepts coincide on the universal model in these settings.\(^{21}\) Outside of the realm of NI-environments, ISR generally refines ICR (imposing sequential rationality restrictions). The fact that, under the richness condition, ISR coincides with ICR in NI-environments, implies that sequential rationality has no bite in these settings. The intuition is simple: In NI-environments players don’t know their own payoffs, they merely have beliefs about them. Once an unexpected information set is reached, Bayes’ rule does not restrict players’ beliefs, which can be set arbitrarily. Under the richness conditions, there are essentially no restrictions on players’ beliefs about their own preferences, so that any behavior can be justified. Hence, the only restrictions that retain their bite are those imposed by (normal form) rationality alone. This also provides the main intuition for ISR’s failure of model invariance in NI-settings (example 4).

To the extent that the interest in studying extensive form games comes from the notion of sequential rationality, PV-settings, in which the assumption that players know their own payoffs is maintained, seem most meaningful for dynamic environments. As shown by proposition 3.3, ISR is model invariant in these settings. In appendix A.3 it is also shown that, in PV-environments, ISR is (generically) equiva-

\(^{21}\) I’m grateful to Eddie Dekel for this observation.
lent to Dekel and Fudenberg’s (1990) $S^\infty W$ procedure applied to the interim normal form.\footnote{The $S^\infty W$-procedure consists of one round of deletion of weakly dominated strategies followed by iterated deletion of strongly dominated strategies.}

**On the Related Solution Concepts.** $\mathcal{ISR}$ generalizes to games with incomplete and imperfect information the solution concept developed by Ben-Porath (1997) to characterize the behavioral implications of (initial) common certainty of rationality. Ben-Porath (1997) also proves that for games with complete and perfect information with payoffs in generic position, his solution concept is equivalent to Dekel and Fudenberg’s (1990) $S^\infty W$-procedure. The results in Penta (2009c) and appendix A.3 generalize Ben-Porath’s (1997) to incomplete and imperfect information.

Battigalli and Siniscalchi (2007) notion of \textit{weak} $\Delta$-rationalizability is also closely related to $\mathcal{ISR}$: their solution concept is not defined for Bayesian games, but for games with payoff uncertainty. For Bayesian games with information types $\mathcal{ISR}$ can be shown to be equivalent to \textit{weak} $\Delta$-rationalizability, where the $\Delta$-restrictions on first order beliefs are those derived from the type space.\footnote{An analogous result in Battigalli et al. (2009) relates $\Delta$-rationalizability and \textit{interim correlated rationalizability} in games with information types.}

**On the Robustness(-es).** A closely related paper is Dekel and Fudenberg (1990): In that paper, solution concept $S^\infty W$ is shown to be “robust” to the possibility that players entertain small doubts about their opponents’ payoff functions. The robustness result for $\mathcal{ISR}$ is in the same spirit. Dekel and Fudenberg (1990) maintain the assumption that players \textit{know} their own payoffs: This corresponds to the PV-case in this paper, for which it is shown (appendix A.3) that $\mathcal{ISR}$ coincides with $S^\infty W$ applied to the interim normal form.
Type space invariance, in the context of normal form games, has been addressed by the literature: Ely and Peski (2006) and Dekel et al. (2007) pointed out the type space-dependence of Interim Independent Rationalizability (IIR). Based on this observation, Ely and Peski (2006) showed that the relevant measurability condition for IIR is not in terms of the $\Theta$-hierarchies, but in terms of $\Delta(\Theta)$-hierarchies. Dekel et al. (2007) instead introduced the concept of Interim Correlated Rationalizability (ICR) and showed that it is type space invariant. To the best of my knowledge, the problem of model invariance was not addressed by the literature.

On the Impact of higher order beliefs on multiplicity. The result of proposition 3.5 implies that, generically, initial common certainty of sequential rationality is sufficient to achieve coordination of expectations. Generically, multiplicity is driven by the common knowledge-assumptions of our models. As discussed in the introduction, this parallels what is known from Weinstein and Yildiz (2007) and the literature on static global games. A growing literature is exploring to what extent the main insights from the theory of global games can be extended to dynamic environments. These contributions are mainly from an applied perspective, and do not pursue a systematic analysis of these problems. Consequently, the conclusions are variegated: for instance, Chamley (1999) and Frankel and Pauzner (2000) obtain the familiar uniqueness result in different setups, under different sets of assumptions. On the other hand, few recent papers seem to question the general validity of these results: For instance, Angeletos, Hellwig and Pavan (2007) and Angeletos and Werning (2006) apply the global games’ information structure to dynamic environments, and obtain non-uniqueness results that contrast with the familiar ones in static settings. The origin of such multiplicity lies in a tension between the global games’ information
structure and the dynamic structure: by relaxing common knowledge-assumptions, the former favors uniqueness; in dynamic games, some information endogenously becomes common knowledge (e.g. public histories), thus mitigating the impact of the information structure.

Many aspects of the papers mentioned above make the comparison with the present framework difficult, and a careful analysis and understanding of the relations between the two approaches is an interesting open question for future research. One important difference underlying the contrasting results in term of uniqueness is certainly the fact that, in those papers, not all common knowledge assumptions are relaxed.\textsuperscript{24} For instance, in the paper by Angeletos et al. (2007), the assumption of common knowledge that the stage game does not depend on the previous history is maintained throughout. Important work for future research is to investigate the robustness and uniqueness questions when only some common knowledge assumptions are relaxed.\textsuperscript{25} This line of research may help shed some light on the relations between the present work and the literature on dynamic global games.

\textsuperscript{24} Another important difference is that the game in Angeletos et al. (2007) has infinite horizon. Extending the analysis of this paper to multistage games with infinite horizon is an important question, left to future research.

\textsuperscript{25} Chapter 2 explored these questions in the context of static games.
Chapter 4

Robust Dynamic Mechanism Design

Abstract: In situations in which the social planner has to make several decisions over time, before agents have observed all the relevant information, static mechanisms may not suffice to induce agents to reveal their information truthfully. This paper focuses on questions of partial and full implementation in dynamic mechanisms, when agents’ beliefs are unknown to the designer (hence the term “robust”). It is shown that a social choice function (SCF) is (partially) implementable for all models of beliefs if and only if it is ex-post incentive compatible. Furthermore, in environments with single crossing preferences, strict ex-post incentive compatibility and a “contraction property” are sufficient to guarantee full robust implementation. This property limits the interdependence in agents’ valuations, the limit being tighter the stronger the “intertemporal effects”.

Full robust implementation requires that, for all models of agents beliefs, all the perfect Bayesian equilibria of a mechanism induce outcomes consistent with the SCF. This paper shows that, for a weaker notion of equilibrium and for a general class of dynamic games, the set of all such
equilibria can be computed by means of a “backwards procedure” which combines the logic of \textit{rationalizability} and \textit{backward induction} reasoning. It thus provides foundation to a tractable approach to the implementation question, allowing at the same time stronger implementation results.

\textbf{Keywords:} backwards induction – dynamic mechanism – implementation – social choice function – rationalizability – robust implementation.

\textbf{JEL Codes:} C72; C73; D82.

\section{4.1 Introduction}

Several situations of economic interest present problems of mechanism design that are inherently dynamic. Consider the problem of a public authority (or “social planner”) who wants to assign yearly licenses for the provision of a public good to the most productive firm in each period. Firms’ productivity is private information and may change over time; it may be correlated over time, and later productivity may depend on earlier allocative choices (for example, if there is learning-by-doing). Hence, the planner’s choice depends on private information of the firms, and the design problem is to provide firms with the incentives to reveal their information truthfully. But firms realize that the information revealed in earlier stages can be used by the planner in the future, affecting the allocative choices of later periods. Thus, in designing the mechanism (e.g. a sequence of auctions), the planner has to take into account “intertemporal effects” that may alter firms’ static incentives.

A rapidly growing literature has recently addressed similar problems of \textit{dynamic mechanism design}, in which the planner has to make several decisions over different periods, with the agents’ information changing over time. In the standard approach, some commonly known distribution over the stochastic process generates payoffs and
signals. Hence, it is implicitly assumed that the designer knows the agents’ beliefs about their opponents’ private information and their beliefs, conditional on all possible realizations of agents’ private information. In that approach, classical implementation questions can be addressed: For any given “model of beliefs”, we can ask under what conditions there exists a mechanism in which agents reveal their information truthfully in a Perfect Bayesian Equilibrium (PBE) of the game (partial implementation), or whether there exists a mechanism in which, all the PBE of the induced game induce outcomes consistent with the social choice function (full implementation).

It is commonly accepted that the assumption that the designer knows the agents’ entire hierarchies of beliefs is too strong. In dynamic settings in particular, the assumption of a commonly known prior entails the planner’s knowledge of significantly more complex objects, such as agents’ hierarchies of conditional beliefs: For instance, in the example above, it means that the designer knows the firms’ conditional beliefs (conditional on all possible realizations of private signals) over own future productivity and the other firms’ current and future productivities and their beliefs, conditional on all realization of their signals. Not only are these assumptions strong, but the sensitivity of game theoretic results to the fine details of agents’ higher order beliefs is also well documented. Weakening the reliance of game theoretic analysis on common knowledge assumptions seems thus crucial to enable us “to conduct useful analysis of practical problems” (Wilson, 1987, p.34).²

¹Among others, see Bergemann and Valimaki (2008), Athey and Segal (2007), Pavan, Segal and Toikka (2009). Gershov and Moldovanu (2009a,b) depart from the “standard” approach described above in that the designer does not know the “true” distribution, combining implementation problems with learning.

²In the context of mechanism design, this research agenda (sometimes referred to as Wilson’s doctrine) has been put forward in a series of papers by Bergemann and Morris, who developed a
This paper focuses on the question of whether partial and full implementation can be achieved, in dynamic environments, when agents’ beliefs are unknown to the designer (hence the term “robust”). For the partial implementation question, building on the existing literature on static robust mechanism design (particularly, Bergemann and Morris, 2005) it is not difficult to show that a Social Choice Function (SCF) is PBE-implementable for all models of beliefs if and only if it is ex-post incentive compatible. The analysis of the full implementation question instead raises novel problems: in order to achieve robust full implementation we need a mechanism such that, for any model of beliefs, all the PBE induce outcomes consistent with the SCF. The direct approach to the question is to compute the set of PBE for each model of beliefs; but the obvious difficulties that this task presents have been a major impediment to the development of a robust approach to dynamic mechanism design.

This paper introduces and provides foundations to a methodology that avoids the difficulties of the direct approach. The key ingredient is the notion of interim perfect equilibrium (IPE). IPE weakens Fudenberg and Tirole’s (1991) PBE allowing a larger set of beliefs off-the-equilibrium path. The advantage of weakening PBE in this context is twofold: on the one hand, full implementation results are stronger if obtained under a weaker solution concept (if all the IPE induce outcomes consistent with the SCF, then so do all the PBE, or any other refinement of IPE); on the other hand, the weakness of IPE is crucial to making the problem tractable. In particular, it is shown that the set of IPE-strategies across models of beliefs can be computed by means of a “backwards procedure” that combines the logic of rationalizability and backward induction reasoning: For each history, compute the set of rationalizable continuation-strategies, treating private histories as “types”, and proceed backwards belief-free approach to classical implementation questions, known as “robust” mechanism design.
from almost-terminal histories to the beginning of the game. (Refinements of IPE would either lack such a recursive structure, or require more complicated backwards procedures.)

The results are applied to study conditions for full implementation in environments with monotone aggregators of information: In these environments information is revealed dynamically, and while agents’ preferences may depend on their opponents’ information (interdependent values) or on the signals received in any period, in each period all the available information (across agents and current and previous periods) can be summarized by one-dimensional statistics. In environments with single-crossing preferences, sufficient conditions for full implementation in direct mechanisms are studied: these conditions bound the amount of interdependence in agents’ valuations, such bounds being more stringent the stronger the “intertemporal effects”.

The rest of the paper is organized as follows: Section 4.2 discusses an introductory example to illustrate the main concepts and insights. Section 4.3 introduces the notion of environments, which define agents’ preferences and information structure (allowing for information to be obtained over time). Section 4.4 introduces mechanisms. Models of beliefs, used to represent agents’ higher order uncertainty, are presented in Section 4.5. Section 4.6 is the core of the paper, and contains the main solution concepts and results for the proposed methodology. Section 4.7 focuses on the problem of partial implementation, while Section 4.8 analyzes the problem of full implementation in direct mechanisms. Proofs are in the Appendices.

4.2 A Dynamic Public Goods Problem

I discuss here an example introducing main ideas and results, abstracting from some technicalities. The section ends with a brief discussion of the suitable generalizations
of the example’s key features.

Consider an environment with two agents \((n = 2)\) and two periods \((T = 2)\). In each period \(t = 1, 2\), agents privately observe a signal \(\theta_{i,t} \in [0, 1]\), \(i = 1, 2\), and the planner chooses some quantity \(q_t\) of public good. The cost function for the production of the public good is \(c(q_t) = \frac{1}{2} q_t^2\) in each period, and for each realization \(\theta = (\theta_{i,1}, \theta_{i,2}, \theta_{j,1}, \theta_{j,2})\), \(i, j = 1, 2\) and \(i \neq j\), agent \(i\)’s valuation for the public goods \(q_1\) and \(q_2\) are, respectively,

\[
\alpha_{i,1} (\theta_1) = \theta_{i,1} + \gamma \theta_{j,1}
\]

and

\[
\alpha_{i,2} (\theta_1, \theta_2) = \varphi (\theta_{i,1}, \theta_{i,2}) + \gamma \varphi (\theta_{j,1}, \theta_{j,2})
\]

where \(\gamma \geq 0\) and \(\varphi : [0, 1]^2 \to \mathbb{R}\) is assumed continuously differentiable and strictly increasing in both arguments. Notice that if \(\gamma = 0\), we are in a private-values setting; for any \(\gamma > 0\), agents have interdependent values. Also, since \(\varphi\) is strictly increasing in both arguments, there are “intertemporal effects”: the first period signal affects the agents’ valuation in the second period.

The notation \(\alpha_{i,t}\) is mnemonic for “aggregator”: functions \(\alpha_{i,1}\) and \(\alpha_{i,2}\) “aggregate” all the information available up to period \(t = 1, 2\) into real numbers \(a_{i,1}, a_{i,2}\), which uniquely determine agent \(i\)’s preferences. Agent \(i\)’s utility function is

\[
u_i (q_1, q_2, \pi_{i,1}, \pi_{i,2}, \theta) = \alpha_{i,1} (\theta_1) \cdot q_1 + \pi_{i,1} + [\alpha_{i,2} (\theta_1, \theta_2) \cdot q_2 + \pi_{j,2}],
\]

where \(\pi_{i,1}\) and \(\pi_{i,2}\) represent the quantity of private good in period \(t = 1, 2\). The
optimal provision of public good in each period is therefore

\[ q_1^* (\theta_1) = \alpha_{i,1} (\theta_1) + \alpha_{j,1} (\theta_1) \quad \text{and} \]
\[ q_2^* (\theta) = \alpha_{i,2} (\theta_1, \theta_2) + \alpha_{j,2} (\theta_1, \theta_2), \]

(4.2)

(4.3)

Consider now the following direct mechanism: agents publicly report messages \( m_{i,t} \in [0,1] \) in each period, and for each profile of reports \( m = (m_{i,1}, m_{j,1}, m_{i,2}, m_{j,2}) \), agent \( i \) receives generalized Vickrey-Clarke-Groves transfers

\[
\pi_{i,1}^* (m_{i,1}, m_{j,1}) = - (1 + \gamma) \left[ \gamma \cdot m_{i,1} \cdot m_{j,1} + \frac{1}{2} m_{j,1}^2 \right] \quad \text{and} \\
\pi_{i,2}^* (m_{i,1}, m_{j,1}) = - (1 + \gamma) \left[ \gamma \cdot \varphi (m_{i,1}m_{i,2}) \cdot \varphi (m_{j,1}m_{j,2}) + \frac{1}{2} \varphi (m_{j,1}m_{j,2})^2 \right],
\]

(4.4)

(4.5)

and the allocation is chosen according to the optimal rule, \( (q_1^* (m_1), q_2^* (m_1, m_2)) \).

If we complete the description of the environment with a model of agents’ beliefs, then the mechanism above induces a dynamic game with incomplete information. The solution concept that will be used for this kind of environments is “interim perfect equilibrium” (IPE), a weaker version of Perfect Equilibrium in which agents’ beliefs at histories immediately following a deviation are unrestricted (they are otherwise obtained via Bayesian updating).

“Robust” implementation though is concerned with the possibility of implementing a social choice function (SCF) irrespective of the model of beliefs. So, consider the SCF \( f = (q_t^*, \pi_{i,t}^*, \pi_{j,t}^*)_{t=1,2} \) that we have just described: We say that \( f \) is partially robustly implemented by the direct mechanism if, for any model of beliefs, truthfully reporting the private signal in each period is an “interim perfect equilibrium” (IPE) of the induced game.
For each \( \theta = (\theta_{i,1}, \theta_{i,2}, \theta_{j,1}, \theta_{j,2}) \) and \( m = (m_{i,1}, m_{i,2}, m_{j,1}, m_{j,2}) \), define

\[
\Delta_i (\theta, m) = \varphi (m_{i,1}, m_{i,2}) - \varphi (\theta_{i,1}, \theta_{i,2}) - \gamma \cdot [\varphi (\theta_{j,1}, \theta_{j,2}) - \varphi (m_{j,1}, m_{j,2})]
\]

\[
= \alpha_{i,2} (m) - \alpha_{i,2} (\theta). \]

In words: given payoff state \( \theta \) and reports \( m \) (for all agents and periods), \( \Delta_i (\theta, m) \) is the difference between the value of the aggregator \( \alpha_{i,2} \) under the reports profile \( m \), and its “true” value if payoff-state is \( \theta \).

For given first period (public) reports \( \hat{m}_1 = (\hat{m}_{i,1}, \hat{m}_{j,1}) \) and private signals \( (\hat{\theta}_{i,1}, \hat{\theta}_{i,2}) \), and for point beliefs \( (\theta_{j,1}, \theta_{j,2}, m_{j,2}) \) about the opponent’s private information and report in the second period, if we ignore problems with corner solutions, then the best response \( m_{i,2}^* \) of agent \( i \) at the second period in the mechanism above satisfies:\footnote{We ignore here the possibility of corner solutions, which do not affect the fundamental insights. Corner solutions will be discussed in Section 4.8.}

\[
\Delta_i \left( \hat{\theta}_{i,1}, \hat{\theta}_{i,2}, \theta_{j,1}, \theta_{j,2}, \hat{m}_1, m_{i,2}^*, m_{j,2} \right) = 0. \tag{4.6}
\]

Also, given private signal \( \hat{\theta}_{i,1} \), and point beliefs about \((\theta_{j,2}, m_{j,1}, m_{i,2}, m_{j,2}) \equiv (\theta_{\backslash(i,1)}, m_{\backslash(i,1)})\), the first period best-response satisfies:

\[
m_{i,1}^* - \hat{\theta}_{i,1} = \gamma \left( \theta_{j,1} - m_{j,1} \right) + \frac{\partial \varphi (m_{i,1}^*, m_{j,2})}{\partial m_{i,1}} \cdot \Delta \left( \hat{\theta}_{i,1}, \theta_{\backslash(i,1)}, m_{i,1}^*, m_{\backslash(i,1)} \right) \tag{4.7}
\]

This mechanism satisfies ex-post incentive compatibility: For each possible realization of \( \theta \in [0, 1] \), conditional on the opponents reporting truthfully, if agent \( i \) has reported truthfully in the past (i.e. \( m_{i,1} = \theta_{i,1} \)), then equation (4.6) is satisfied if and only if \( m_{i,2} = \theta_{i,2} \). Similarly, given that \( \Delta (\theta, m) = 0 \) in the second period, the right-hand side of (4.7) is zero if the opponents report truthfully in the first period, and
so it is optimal to report \( m_{i,1} = \theta_{i,1} \) (independent of the realization of \( \theta \)). Notice that this is the case for any \( \gamma \geq 0 \). Since such incentive compatibility is realized \emph{ex-post}, conditioning to all information being revealed, incentive compatibility will also be realized with respect to any model of beliefs. Thus, for any such model of beliefs, there always exist an IPE that induces truthful revelation, that is, \( f \) is \emph{robustly partially implementable} if \( \gamma \geq 0 \).

Even with ex-post incentive compatibility, it is still possible that, for some model of beliefs, there exists an IPE which does not induce truthful revelation: To achieve \emph{full robust implementation} in this mechanism we must guarantee that \emph{all} the IPE for \emph{all} models of beliefs induce truthful revelation. We approach this problem indirectly, applying a “backwards procedure” to the “belief-free” dynamic game that will be shown to characterize the set of IPE-strategies across models of beliefs. In the procedure, for each public history \( \hat{m}_1 \) (profile of first-period reports), apply rationalizability in the continuation game, treating the private histories of signals as “types”; then, apply rationalizability at the first stage, maintaining that continuation strategies are rationalizable in the corresponding continuations.

Before illustrating the procedure, notice that equation (4.6) implies that, conditional on having reported truthfully in the first period \( (m_{i,1} = \theta_{i,1}) \), truthful revelation in the second period is a best-response to truthful revelation of the opponent irrespective of the realization of \( \theta \). Now, maintain that the opponent is revealing truthfully \( (m_{j,t} = \theta_{j,t} \text{ for } t = 1, 2) \); if \( m_{i,1} \neq \theta_{i,1} \), i.e. if \( i \) has misreported in the first period, the optimal report in the second period is a further misreport \( (m_{i,2} \neq \theta_{i,2}) \), such that the implied value of the aggregator \( \alpha_{i,2} \) is equal to its true value (i.e.: \( \Delta (\theta, \hat{m}_1, m_2) = 0 \)). This is the notion of \emph{self-correcting strategy}, \( s^*_i \): a strategy that reports truthfully at the beginning of the game and at every truthful history, but in which earlier mis-
reports (which do not arise if \( s^c \) is played) are followed by further misreports, to “correct” the impact of the previous misreports on the value of the aggregator \( \alpha_{i,2} \). It will be shown next that, if \( \gamma < 1 \), then the self-correcting strategy profile is the only profile surviving the “backward procedure” described above. Hence, given the results of Section 4.6, the self-correcting strategy is the only strategy played in all IPE for all “models of beliefs”. Since \( s^c \) induces truthful revelation, this implies that, if \( \gamma < 1 \), SCF \( f \) is fully robustly implemented.

For given \( \tilde{m}_1 \) and \( \theta_i = (\theta_{i,1}, \theta_{i,2}) \), let \( x_i(\theta_i) = [\varphi(\tilde{m}_{i,1}, m_{i,2}) - \varphi(\theta_{i,1}, \theta_{i,2})] \) denote type \( \theta_i \)'s “implied over-report” of the value of \( \varphi \). Then equation (4.6) can be interpreted as saying that “the optimal over-report of \( \varphi \) is equal to \(-\gamma \) times the (expected) opponent’s under-report of \( \varphi \)”. Let \( x^0_j \) and \( \bar{x}^0_j \) denote the minimum and maximum possible values of \( x_j \). Then, if \( i \) is rational, his over-reports are bounded by \( x_i(\theta_i) \leq \bar{x}_i \equiv \gamma \cdot x^0_j \) and \( x_i(\theta_i) \geq x^1_j \equiv -\gamma \cdot \bar{x}^0_j \). Recursively, define \( \bar{x}^k_i = -\gamma \cdot x^{k-1}_j \) and \( x^k_i = -\gamma \cdot \sum_{j \neq i} x^{k-1}_j \). Also, for each \( k \) and \( i \), let \( y^k_i \equiv [x^k_i - x^1_i] \) denote the distance between the maximum and lowest possible over-report at step \( k \). Then, substituting, we obtain the following system of difference equations:

\[
\begin{align*}
\mathbf{y}^k &= \Gamma \cdot \mathbf{y}^{k-1} \\
\mathbf{y}^k &= \begin{pmatrix}
  y^k_i \\
  y^k_j
\end{pmatrix} \quad \text{and} \quad \Gamma = \begin{bmatrix}
  0 & \gamma \\
  \gamma & 0
\end{bmatrix}
\end{align*}
\]

(4.8)

Notice that the continuation game from \( \tilde{m}_1 \) is dominance solvable if and only if \( \mathbf{y}^k \to 0 \) as \( k \to \infty \). In that case, for each \( \theta_i \), \( x_i(\theta_i) \to 0 \), and so truthtelling is the unique rationalizable strategy. Thus, it suffices to study conditions for the dynamic system above to converge to the steady state \( 0 \). In this example, \( 0 \) is an asymptotically stable steady state if and only if \( \gamma < 1 \). Hence, if \( \gamma < 1 \), the only rationalizable outcome in the continuation from \( \tilde{m}_1 \) guarantees that \( \Delta = 0 \). Given this, the first period best
response simplifies to
\[ m_{i,1}^* - \hat{\theta}_{i,1} = \gamma (\theta_{j,1} - m_{j,1}). \]

The same argument can be applied to show that truthful revelation is the only rationalizable strategy in the first period if and only if \( \gamma < 1 \) (cf. Bergemann and Morris, 2009). Then, if \( \gamma < 1 \), the self-correcting strategy is the only “backward rationalizable” strategy, hence the only strategy played as part of IPE for all models of beliefs.

**Key Properties and their Generalizations.** The analysis in Section 4.8 generalizes several features of this example: The notions of aggregator functions and of self-correcting strategy have a fairly straightforward generalization. An important feature is that, in each period, the marginal rate of substitutions between \( q_t \) and the other goods is increasing in \( \alpha_{i,t} \) for each \( i \). This property implies that, for given beliefs about the space of uncertainty and the opponents’ messages, higher types report higher messages: such monotonicity, allowing us to construct the recursive system (4.8). Several versions of single-crossing conditions generalize this aspect in Section 4.8. Finally, the generalization of the idea that \( \gamma < 1 \) takes the form of a “contraction property”.\(^4\) Consider the first period: for any \( \theta_1 \) and \( m_1 \),

\[ \alpha_{i,1}(m_{i,1}, m_{j,1}) - \alpha_{i,1}(\theta_{i,1}, \theta_{j,1}) = (m_{i,1} - \theta_{i,1}) + \gamma (m_{j,1} - \theta_{j,1}). \]

Thus, if \( \gamma < 1 \), for any set of possible “deceptions” \( D \) there exists at least one agent \( i \in \{1, 2\} \) which can unilaterally sign \( [\alpha_{i,1}(m'_{i,1}, m_{j,1}) - \alpha_{i,1}(\theta_{i,1}, \theta_{j,1})] \) by reporting some message \( m'_{i,1} \neq \theta_{j,1} \), irrespective of \( \theta_{j,1} \) and \( m_{j,1} \). That is, for all \( \theta_{j,1} \) and \( m_{j,1} \) in

---

\(^4\)The name, borrowed from Bergemann and Morris (2009), is evocative of the logic behind equation 4.8.
\[ D: \]
\[
\text{sign} \left[ \alpha_{i,1} \left( m'_i,1, m_{j,1} \right) - \alpha_{i,1} \left( \theta_{i,1}, \theta_{j,1} \right) \right] = \text{sign} \left[ m'_i,1 - \theta_{i,1} \right].
\]
Similarly, for public history \( \hat{m}_1 \) in the second period, \( \gamma < 1 \) guarantees that there exists at least one agent that can unilaterally sign \( \alpha_{i,2} \left( \hat{m}_1, m'_i,2, m_{j,2} \right) - \alpha_{i,2} \left( \theta_i, \theta_j \right) \), (uniformly over \( \theta_j \) and \( m_{j,2} \)), by reporting some message \( m'_i,2 \) other than the one implied by the self-correcting strategy, \( s^c_i,2 \):
\[
\text{sign} \left[ m'_i,2 - s^c_i,2 \right] = \text{sign} \left[ \alpha_{i,2} \left( \hat{m}_1, m'_i,2, m_{j,2} \right) - \alpha_{i,2} \left( \theta_i, \theta_j \right) \right].
\]
This property will be required to hold at all histories.\(^5\)

### 4.3 Environments

Consider an environment with \( n \) agents and \( T \) periods. In each period \( t = 1, \ldots, T \), each agent \( i = 1, \ldots, n \) observes a signal \( \theta_{i,t} \in \Theta_{i,t} \). For each \( t \), \( \Theta_t := \Theta_{1,t} \times \ldots \times \Theta_{n,t} \) denotes the set of period-\( t \) signals profiles. For each \( i \) and \( t \), the set \( \Theta_{i,t} \) is assumed non-empty and compact subset of a finitely dimensional Euclidean space. For each agent \( i \), \( \Theta_i^* := \times_{t=1}^T \Theta_{i,t} \) is the set of \( i \)'s payoff types: a payoff-type is a complete sequence of agent \( i \)'s signals in every period. A state of nature is a profile of agents’ payoff types, and the set of states of nature is defined as \( \Theta^* := \Theta_1^* \times \ldots \times \Theta_n^* \).

In each period \( t \), the social planner chooses an allocation from a non-empty subset of a finitely dimensional Euclidean space, \( \Xi_t \) (possibly a singleton). The set \( \Xi^* = \times_{t=1}^T \Xi_t \) denotes the set of feasible sequences of allocations. Agents have preferences over sequences of allocations that depend on the realization of \( \Theta^* \): for each \( i = 1, \ldots, n \), preferences are represented by utility functions \( u_i : \Xi^* \times \Theta^* \rightarrow \mathbb{R} \). Thus, the states of nature characterize everybody’s preferences over the sets of feasible allocations.

\(^5\)The general formulation (Section 4.8) allows to accommodate the case analogous to the possibility of corner solutions in the example above.
An *environment* is defined by a tuple

\[ \mathcal{E} = \langle N, \Xi^*, \Theta^*, (u_i)_{i \in N} \rangle, \]

assumed common knowledge.

Notice that an *environment* only represents agents’ *knowledge* and *preferences*: it does not encompass agents’ beliefs. Each agent’s *payoff-type* \( \theta_i \in \Theta^*_i \) represents his *knowledge* of the state of nature at the end of period \( T \). That is, his knowledge of everyone’s preferences about the feasible allocations.

For each \( t \), let \( Y^t_i := \times_{t=1}^T \Theta^*_i \) denote the set of possible histories of player \( i \)’s signals up to period \( t \). For each \( t \) and private signals \( y^t_i = (\theta_{1,1}, ..., \theta_{t,t}) \in Y^t_i \), agent \( i \) knows that the “true” state of nature \( \theta^* \in \Theta^* \) belongs to the set \( \{y^t_i\} \times (\times_{t=t+1}^T \Theta_{i,t}) \times \Theta^*_{i,t} \).

At any point in time, agent form *beliefs* about the features of the environment they don’t *known*. These beliefs should be interpreted here as purely subjective. Since *robust* mechanism design is concerned with problems of implementation as agents’ model of beliefs change, we maintain the description of the agents’ beliefs separate from the description of their information (which is part of the environment, and held constant). Models of beliefs are presented in Section 4.5.

**Social Choice Functions.** The description of the primitives of the problem is completed by the specification of a *social choice function* (SCF), \( f : \Theta^* \rightarrow \Xi^* \).

Given the constraints of the environment, a necessary condition for a SCF to be implementable is that period–\( t \) choices be measurable with respect to the information available in that period. That is:

**Remark 4.1.** A necessary condition for a SCF \( f : \Theta^* \rightarrow \Xi^* \) to be implementable is that there exist functions \( f_t : Y^t \rightarrow \Xi_t, t = 1, ..., T \), such that for each \( \theta = (\theta_1, ..., \theta_T) \),

\[ f(\theta) = (f_t(\theta_1, ..., \theta_t))_{t=1}^T. \]
In the following, we will only consider SCF that satisfy such necessary condition. We thus write $f = (f_t)_{t=1}^T$.

### 4.4 Mechanisms

A mechanism is a tuple

$$\mathcal{M} = \left( \left( (M_{i,t})_{i \in N} \right)_{t=1}^T, (g_t)_{t=1}^T \right)$$

where each $M_{i,t}$ is a non-empty set of messages available to agent $i$ at period $t$ ($i \in N$ and $t = 1, ..., T$); $g_t$ are “outcome functions”, assigning allocations to each history at each stage. As usual, for each $t$ we define $M_t = \times_{i \in N} M_{i,t}$.

Formally, let $H^0 := \{\emptyset\}$ ($\emptyset$ denotes the empty history). For each $t = 1, ..., T$, the period-$t$ outcome function is a mapping $g_t : H^{t-1} \times M_t \to \Xi_t$, where for each $t$, the set of public histories of length $t$ is defined as:

$$H^t = \left\{ (h^{t-1}, m_t, \xi_t) \in H^{t-1} \times M_t \times \Xi_t : \xi_t = g_t (h^{t-1}, m_t) \right\}.$$

The set of public histories is defined as $H = \bigcup_{\tau=0}^T H^\tau$. Throughout the paper we focus on compact mechanisms, in which the sets $M_{i,t}$ are compact subsets of finitely dimensional Euclidean spaces.

#### 4.4.1 Games with Payoff Uncertainty

An environment $\mathcal{E}$ and a mechanism $\mathcal{M}$ determine a dynamic game with payoff uncertainty, that is a tuple

$$(\mathcal{E}, \mathcal{M}) = \left\langle N, (\mathcal{H}_i, \Theta^*_i, u_i)_{i \in N} \right\rangle.$$ 

Sets $N$, $\Theta^*_i$ and payoff functions $u_i$ are as defined in $\mathcal{E}$, while sets $\mathcal{H}_i$ are defined as follows: the set of player $i$’s private signals is given by $Y^t_i = (\times_{\tau=1}^t \Theta_{i,\tau})$; sets $H^t$
$(t = 0, 1, \ldots, T)$ are defined as in $\mathcal{M}$; player $i$’s set of private histories of length $t$ $(t = 1, \ldots, T)$ is defined as $H_t^i := H_t^{i-1} \times Y_t^i$, and finally $\mathcal{H}_i := \{\emptyset\} \cup (\cup_{t=1}^{T} H_t^i)$ denotes the set of $i$’s private histories. Thus, each private history of length $t$ is made of two components: a public component, made of the previous messages of the agents and the allocations chosen by the mechanism in periods 1 through $t - 1$; and a private component, made of agent $i$’s private signals from period 1 through $t$.

It is convenient to introduce notation for the partial order representing the precedence relation on the sets $\mathcal{H}$ and $\mathcal{H}_i$: $h^r \prec h^t$ indicates that history $h^r$ is a predecessor of $h^t$ (similarly for private histories: $(h^{r-1}_t, y^r_i) \prec (h^{t-1}_t, y^t_i)$ if and only if $h^r \prec h^t$ and $y^r_i \prec y^t_i$).

**Remark 4.2.** The tuple $(\mathcal{E}, \mathcal{M})$ is not a standard incomplete information game (Harsanyi, 1967-68), because it does not encompass a specification of agents’ interactive beliefs. A standard incomplete information game is obtained by appending a model of beliefs $\mathcal{B}$, introduced in Section 4.5. Concepts and notation for structures $(\mathcal{E}, \mathcal{M}, \mathcal{B})$ will be introduced in Section 4.6.1.

**Strategic Forms.**

Agents’ strategies in the game $(\mathcal{E}, \mathcal{M})$ are measurable functions $s_i : \mathcal{H}_i \to M_i$ such that $s_i(h^t_i) \in M_{i,t}$ for each $h^t_i \in \mathcal{H}_i$. The set of player $i$’s strategies is denoted by $S_i$, and as usual we define the sets $S = \times_{i \in N} S_i$ and $S_{-i} = \times_{j \neq i} S_j$. Payoffs are defined as in $\mathcal{E}$, as functions $u_i : \Xi^* \times \Theta^* \to \mathbb{R}$. For any strategy profile $s \in S$, each realization of $\theta \in \Theta^*$ induces a terminal allocation $g^s(\theta) \in \Xi^*$. Hence, we can define strategic-form payoff functions $U_i : S \times \Theta^* \to \mathbb{R}$ as $U_i(s, \theta) = u_i(g^s(\theta), \theta)$ for each $s$ and $\theta$.

As the game unfolds, agents learn about the environment observing the private signals, but they also learn about the opponents’ behavior through the public histo-
ries: for each public history $h^t$ and player $i$, let $S_i(h^t)$ denote the set of player $i$’s strategies that are consistent with history $h^t$ being observed. Clearly, since $i$’s private histories are only informative about the opponents’ behavior through the public history, for each $i$, $h^t_i = (h^{t-1}_i, y^t_i) \in \mathcal{H}_i$ and $j \neq i$, $S_j(h^t_i) = S_j(h^{t-1})$.

For each $h^t$, $S_i^{ht}$ denotes the set of strategies in the subform starting from $h^t$, and for each $s_i \in S_i$, $s_i|h^t \in S_i^{ht}$ denotes the continuation $s_i$ starting from $h^t$. The notation $g^{s|ht}(\theta)$ refers to the terminal history induced by strategy profile $s$ from the public history $h^t$ if the realized state of nature is $\theta$. Strategic-form payoff functions can be defined for continuations from a given public history: for each $h^t \in \mathcal{H}$ and each $(s, \theta) \in S \times \Theta^*$, $U_i(s, \theta; h^t) = u_i(g^{s|ht}(\theta), \theta)$. For the initial history $\phi$, it will be written $U_i(s, \theta)$ instead of $U_i(s, \theta; \phi)$. Sets $\mathcal{H}_i$ and $S_i$ are endowed with the standard metrics derived from the Euclidean metric on $H^T \times \Theta^*$.

4.4.2 Direct Mechanisms

The notion of direct mechanism is based on the observation made in remark 4.1:

**Definition 4.1.** A mechanism is direct, denoted by $\mathcal{M}^*$, if for each $i$ and for each $t = 1, \ldots, T$, $M_{i,t} = \Theta_{i,t}$, and $g_t = f_t$.

In a direct mechanism, agents are asked to announce their signals at every period. Based on the reports, the mechanism chooses the period-$t$ allocation according to the function $f_t : Y^t \rightarrow \Xi_t$, as specified by the SCF. The truth-telling strategies are those that, conditionally on having reported truthfully in the past, report truthfully the period-$t$ signal, $\theta^t_i$. Truth-telling strategies may differ in the behavior they prescribe at histories following past misreports, but they all are outcome equivalent and induce

\[^6\text{See Appendix B.1.1 for details.}\]
truthful revelation in each period. The set of such strategies is denoted by $S_i^*$, with typical element $s_i^*$.

### 4.5 Models of Beliefs

A *model of beliefs* for an environment $\mathcal{E}$ is a tuple

$$\mathcal{B} = \langle N, \Theta^*, (B_i, \beta_i)_{i \in N} \rangle$$

such that for each $i$, $B_i$ is measurable space (the set of *types*), and $\beta_i : B_i \to \Delta (\Theta^* \times B_{-i})$ is a measurable function.$^7$

At period 0 agents have no information about the environment. Their (subjective) “priors” about the payoff state and the opponents’ beliefs is implicitly represented by means of types $b_i$, with beliefs given by $\beta_i(b_i) \in \Delta (\Theta^* \times B_{-i})$. In periods $t = 1, ..., T$, agents update their beliefs using their private information (the history of payoff signals), and other information possibly disclosed by the mechanism set in place. The main difference with respect to standard (static) type spaces with payoff types, as in Bergemann and Morris (2005) for example, is that players here don’t *know* their own *payoff-types* at the interim stage: payoff-types are disclosed over time, and known only at the end of period $T$. Thus, an agent’s type at the beginning of the game is completely described by a “prior” belief over the payoff states and the opponents’ types.

In standard models of dynamic mechanism design (e.g. Bergemann and Valimaki, 2008, and Athey and Segal, 2007, Pavan et al., 2009), the history of payoff types completely determines players’ beliefs about the payoff states and the opponents’

---

$^7$For a measurable space $X$, $\Delta (X)$ denotes the set of probability measures on $X$, endowed with the topology of weak convergence and the corresponding Borel sigma-algebra.
beliefs at each point of the process.\footnote{Classical mechanism design focuses almost exclusively on the case of payoff type spaces. Neeman (2004) shows how relaxing this assumption may crucially affect the results.} In the present setting this corresponds to the case where, for each \( i \), \( B_i \) is a singleton and \( \text{supp}(\text{marg}_{\Theta^*;\beta_i (b_i)}) = \Theta_i^* \): A unique “prior” describes the beliefs (of any order) for each player, so that conditional beliefs are uniquely determined for all possible realizations of the payoff types.

To summarize our terminology, in an environment with beliefs \((\mathcal{E}, \mathcal{B})\) we distinguish the following “stages”: in period 0 (the interim stage) agents have no information, their (subjective) prior is represented by types \( b_i \), with beliefs \( \beta_i (b_i) \in \Delta (\Theta^* \times B_{-i}) \); \( T \) different period-\( t \) interim stages, for each \( t = 1, ..., T \), when a type’s beliefs after a history of signals \( y^t_i \) would be concentrated on the set

\[
\{y^t_i\} \times (\times_{r=t+1}^T \Theta_{i,t}) \times \Theta^*_{-i} \times B_{-i}.
\]

The term “ex-post” refers to hypothetical situations in which interim profiles are revealed: “period-\( t \) ex-post stage” refers to a situation in which everybody’s signals up to period \( t \) are revealed. By “ex-post” stage we refer to the final realization, when payoff-states are fully revealed (or the period-\( T \) ex-post stage).

### 4.6 Solution Concepts

This section is organized in two parts: the first, introduces the main solution concept for environments with a model of beliefs; the second introduces the solution concept for environments without beliefs, which will be used in the analysis of the full-implementation problem in Section 4.8.
4.6.1 Mechanisms in Environments with Beliefs: \((\mathcal{E}, \mathcal{M}, \mathcal{B})\)

A tuple \((\mathcal{E}, \mathcal{M}, \mathcal{B})\) determines a dynamic incomplete information game in the sense of Harsanyi. Strategies are thus measurable mappings \(\sigma_i : B_i \rightarrow S_i\), and the set of strategies is denoted by \(\Sigma_i\). At period 0, agents only know their own type. Hence, the set of agent \(i\)'s private histories of length 0 coincides with the set of his types. It is therefore convenient to identify types with such histories, and write \(h^0_i \in B_i\).

A system of beliefs consists of collections \((p_i(h^t_i))_{h^t_i \in \mathcal{H}_i \setminus \{\phi\}}\) for each agent \(i\), such that \(p_i(h^t_i) \in \Delta(\Theta^* \times B_{-i})\): a belief system represents agents’ conditional beliefs about the payoff state and the opponents’ types at each private history. A strategy profile and a belief system \((\sigma, p)\) form an assessment. For each agent \(i\), a strategy profile \(\sigma\) and conditional beliefs \(p_i\) induce, at each private history \(h^{t-1}_i\), a probability measure \(P^{\sigma, p_i}(h^{t-1}_i)\) over the histories of length \(t\).

**Definition 4.2.** Fix a strategy profile \(\sigma \in \Sigma\). A beliefs system \(p\) is consistent with \(\sigma\) if for each \(i \in N\):

\[
\forall h^0_i \in B_i; p_i(h^0_i) = \beta_i(h^0_i) \quad (4.9)
\]

\[
\forall h^t_i = (y^t_i, h^{t-1}_i) \in \mathcal{H}_i \setminus \{\phi\}
\supp[p_i(h^t_i)] \subseteq \{y^t_i\} \times (\times_{\tau=t+1}^{T} \Theta_{i, \tau}) \times \Theta^* \times B_{-i} \quad (4.10)
\]

and for each \(h^t_i\) such that \(h^{t-1}_i \prec h^t_i\), \(p_i(h^t_i)\) is obtained from \(p_i(h^{t-1}_i)\) and \(P^{\sigma, p_i}(h^{t-1}_i)\) via Bayesian updating (whenever possible).

Condition (4.9) requires each agent’s beliefs conditional on observing type \(b_i\) to agree with that type’s beliefs as specified in the model \(B\); condition (4.10) requires conditional beliefs at each private history to be consistent with the information about
the payoff state contained in the history itself; finally, the belief system \( p_i \) is consistent with Bayesian updating whenever possible.

From the point of view of each \( i \), for each \( h^i_{t} \in \mathcal{H}_i \setminus \{\phi\} \) and strategy profile \( \sigma \), the induced terminal history is a random variable that depends on the realization of the state of nature and opponents’ type profile (agent \( i \)’s type \( h^0_i \) is known to agent \( i \) at \( h^i_{t} \), \( h^0_i \prec h^i_{t} \)). This is denoted by \( g^{\sigma|h^i_{t}}(\theta, b_{-i}) \). As done for games without a model of beliefs (Section 4.4.1), we can define strategic-form payoff functions as follows:

\[
U_i(\sigma, \theta, b_{-i}; h^i_{t}) = u_i \left( g^{\sigma|h^i_{t}}(\theta, b_{-i}), \theta \right).
\]

**Definition 4.3.** Fix a belief system \( p \). A strategy profile is sequentially rational with respect to \( p \) if for every \( i \in N \) and every \( h^i_{t} \in \mathcal{H}_i \setminus \{\phi\} \), the following inequality is satisfied for every \( \sigma^i_{t} \in \Sigma_i \):

\[
\int_{\Theta^i \times B_{-i}} U_i(\sigma, \theta, b_{-i}; h^i_{t}) \cdot dp_i(h^i_{t}) \geq \int_{\Theta^i \times B_{-i}} U_i(\sigma^i_{t}, \sigma_{-i}, \theta, b_{-i}; h^i_{t}) \cdot dp_i(h^i_{t}).
\]  

(4.11)

**Definition 4.4.** An assessment \((\sigma, p)\) is an Interim Perfect Equilibrium (IPE) if:

1. \( \sigma \) is sequentially rational with respect to \( p \); and

2. \( p \) is consistent with \( \sigma \).

If inequality (4.11) is only imposed at private histories of length zero, the solution concept coincides with interim equilibrium (Bergemann and Morris, 2005). IPE refines interim equilibrium imposing two natural conditions: first, sequential rationality; second, consistency of the belief system.

---

9I avoid the adjective “Bayesian” (preferring the terminology “interim” perfect equilibrium) because the models of beliefs under consideration are not necessarily consistent with a common prior. For the same reason, Bergemann and Morris (2005) preferred the terminology “interim” to that of Bayes-Nash equilibrium.
The notion of consistency adopted here imposes no restrictions on the beliefs held at histories that receive zero probability at the preceding node.\textsuperscript{10} Hence, even if agents’ initial beliefs admit a common prior, IPE is weaker than Fudenberg and Tirole’s (1991) perfect Bayesian equilibrium. Also, notice that any player’s deviation is a zero probability event, and treated the same way. In particular, if history $h^t_i$ is precluded by $\sigma_i(h^{t-1}_i)$ alone, $h^t_i \notin \text{ supp}^\alpha P^\sigma_{\text{ps}}(h^{t-1}_i)$, and agent $i$’s beliefs at $h^t_i$ are unrestricted the same way they would be after an unexpected move of the opponents. This feature of IPE is not standard, but it is key to the result that the set of IPE-strategies across models of beliefs can be computed by means of a convenient “backwards procedure”: Treating own deviations the same as the opponents’ is key to the possibility of considering continuation games “in isolation”, necessary for the result. In Penta (2009a) I consider a minimal strengthening of IPE, in which agents’ beliefs are not upset by unilateral own deviations, and I show how the analysis that follows adapts to that case: The “backwards procedure” to compute the set of equilibria across models of beliefs must be modified, so to keep track of the restrictions the extensive form imposes on the agents’ beliefs at unexpected nodes. Losing the possibility of envisioning continuation games “in isolation”, the modified procedure is more complicated, essentially undoing the advantages of the indirect approach.

Furthermore, for the sake of the full-implementation analysis, it can be shown that in the framework considered in Section 4.8, the set of IPE-strategies across models of beliefs coincides with the set of strong IPE-strategies across models of beliefs. Thus, from the point of view of the full-implementation results of Section 4.8, this point is

\textsuperscript{10}IPE is consistent with a “trembling-hand” view of unexpected moves, in which no restrictions on the possible correlations between “trembles” and other elements of uncertainty are imposed. Unlike other notions of weak perfect Bayesian equilibrium, in IPE agents’ beliefs are consistent with Bayesian updating also off-the-equilibrium path. In particular, in complete information games, IPE coincides with subgame-perfect equilibrium.
not critical.

4.6.2 Mechanisms in Environments without Beliefs: \((\mathcal{E}, \mathcal{M})\)

This section introduces a solution concept for dynamic games without a model of beliefs, \textit{backward rationalizability} \((\mathcal{BR})\), and shows that it characterizes the set of IPE-strategies across models of beliefs (proposition 4.1). It is also shown that \(\mathcal{BR}\) can be conveniently solved by a “backwards procedure” that extends the logic of backward induction to environments with incomplete information (proposition 4.2).

In environments without a model of beliefs we will not follow a classical equilibrium approach: no coordination of beliefs on some equilibrium strategy is imposed. Rather, agents form conjectures about everyone’s behavior, which may or may not be consistent with each other. To avoid confusion, we refer to this kind of beliefs as “conjectures”, retaining the term “beliefs” only for those represented in the models of Section 4.5.

**Conjectures.** Agents entertain conjectures about the space \(\Theta^* \times S\). As the game unfolds, and agents observe their private histories, their conjectures change. For each private history \(h_i^t = (h_i^{t-1}, y_i^t) \in \mathcal{H}_i\), define the event \([h_i^t] \subseteq \Theta^* \times S\) as:

\[
[h_i^t] = \{y_i^t\} \times (\times_{\tau=t+1}^T \Theta_{i,\tau}) \times \Theta_{i,\tau} \times S(h_i^{t-1}).
\]

(Notice that, by definition, \([h_i^t] \subseteq [h_i^{t-1}]\) whenever \(h_i^{t-1} \leq h_i^t\).)

**Definition 4.5.** A conjecture for agent \(i\) is a conditional probability system \((\text{CPS hereafter})\), that is a collection \(\mu^i = (\mu^i(h_i^t))_{h_i^t \in \mathcal{H}_i}\) of conditional distributions \(\mu^i(h_i^t) \in \Delta(\Theta^* \times S)\) that satisfy the following conditions:

\(C.1\) For all \(h_i^t \in \mathcal{H}_i\), \(\mu^i(h_i^t) \in \Delta([h_i^t]);\)
C.2 For every measurable $A \subseteq [h^i_t] \subseteq [h^{i-1}_t]$, $\mu^i (h^i_t) [A] \cdot \mu^i (h^{i-1}_{i-1}) [h^i_t] = \mu^i (h^{i-1}_{i-1}) [A]$.

The set of agent $i$’s conjectures is denoted by $\Delta^i (\Theta^* \times S)$.\textsuperscript{11}

Condition C.1 states that agents’ are always certain of what they know; condition C.2 states that agents’ conjectures are consistent with Bayesian updating whenever possible. Notice that in this specification agents entertain conjectures about the payoff state, the opponents’ and their own strategies. This point is discussed in Section 4.6.3.

**Sequential Rationality.** A strategy $s_i$ is sequentially rational with respect to conjectures $\mu^i$ if, at each history $h^i_t \in H_i$, it prescribes optimal behavior with respect to $\mu^i (\cdot ; h^i_t)$ in the continuation of the game. Formally: Given a CPS $\mu^i \in \Delta^i (\Theta^* \times S)$ and a history $h^i_t = (h^{i-1}_t, y^i_t)$, strategy $s_i$ expected payoff at $h^i_t$, given $\mu^i$, is defined as:

$$U_i (s_i, \mu^i; h^i_t) = \int_{\Theta^* \times S_{-i}} U_i (s_i, s_{-i}, \theta; h^{i-1}_t) \cdot \text{d} \nu_{\Theta^* \times S_{-i}} \mu^i (h^i_t). \quad (4.12)$$

**Definition 4.6.** A strategy $s_i$ is sequentially rational with respect to $\mu^i \in \Delta^i (\Theta^* \times S)$, written $s_i \in r_i (\mu^i)$, if and only if for each $h^i_t \in H_i$ and each $s'_i \in S_i$ the following inequality is satisfied:

$$U_i (s_i, \mu^i; h^i_t) \geq U_i (s'_i, \mu^i; h^i_t). \quad (4.13)$$

If $s_i \in r_i (\mu^i)$, we say that conjectures $\mu^i$ “justify” strategy $s_i$.

**Backward Rationalizability.**

We introduce now the solution concept that will be shown (proposition 4.1) to characterize the set of IPE-strategies across models of beliefs, Backwards Rationalizability (BR). The name is justified by proposition 4.2, which shows that BR can be computed by means of a “backwards procedure” that combines the logic of rationalizability and backwards induction.

\textsuperscript{11}The general definition of a CPS is in Appendix B.1.2.
Definition 4.7. For each $i \in N$, let $\mathcal{BR}_i^0 = S_i$. Define recursively, for $k = 1, 2, ...$

$$\mathcal{BR}_i^k = \left\{ \hat{s}_i \in \mathcal{BR}_i^{k-1} : \begin{array}{l} \exists \mu^i \in \Delta^{{\mathcal{H}}_i} (\Theta^* \times S) \text{ s.t.} \\
(1) \hat{s}_i \in r_i (\mu^i) \\
(2) \text{supp} (\mu^i (\phi)) \subseteq \Theta^* \times \{\hat{s}_i\} \times \mathcal{BR}_i^{k-1} \\
(3) \text{for each } h^t_i = (h^{t-1}_i, y^t_i) \in \mathcal{H}_i: \\
s \in \text{supp} (\text{marg}_{s} \mu^i (h^t_i)) \text{ implies:} \\
(3.1) s_i | h^t_i = \hat{s}_i | h^t_i, \text{ and} \\
(3.2) \exists s'_{-i} \in \mathcal{BR}_{-i}^{k-1} : s'_{-i} | h^{t-1} = s_{-i} | h^{t-1} \end{array} \right\}$$

Finally, $\mathcal{BR} := \bigcap_{k \geq 0} \mathcal{BR}_i^k$.

$\mathcal{BR}$ consists of an iterated deletion procedure. At each round, strategy $s_i$ survives if it is justified by conjectures $\mu^i$ that satisfy two conditions: condition (2) states that at the beginning of the game, the agent must be certain of his own strategy $s_i$ and have conjectures concentrated on opponents’ strategies that survived the previous rounds of deletion; condition (3) restricts the agent’s conjectures at unexpected histories: condition (3.1) states that agent $i$ is always certain of his own continuation strategy; condition (3.2) requires conjectures to be concentrated on opponents’ continuation strategies that are consistent with the previous rounds of deletion. However, at unexpected histories, agents’ conjectures about $\Theta^*$ are essentially unrestricted. Thus, condition (3) embeds two conceptually distinct kinds of assumptions: the first concerning agents’ conjectures about $\Theta^*$; the second concerning their conjectures about the continuation behavior. For ease of reference, they are summarized as follows:

- **Unrestricted-Inference Assumption (UIA):** At unexpected histories, agents’ conjectures about $\Theta^*$ are unrestricted. In particular, agents are free to infer anything about the opponents’ private information (or their own future signals) from the public history.

\footnote{It goes without saying that whenever we write a condition like $\mu^i (X|h^t_i) \geq \kappa$ and $X$ is not measurable, the condition is not satisfied.}
For example, conditional conjectures may be such that \( \text{marg}_{\Theta_i} \mu^t(\cdot | h^t_i) \) is concentrated on a “type” \( y^t_{-i} \) for which some of the previous moves in \( h^{t-1} \) are irrational. Nonetheless, condition (3.2) implies that it is believed that \( y^t_{-i} \) will behave rationally in the future. From an epistemic viewpoint, it can be shown that \( BR \) can be interpreted as common certainty of future rationality at every history.

- **Common Certainty in Future Rationality (CCFR):** at every history (expected or not), agents share common certainty in future rationality.

Thus, CCFR can be interpreted as a condition of belief persistence on the continuation strategies.\(^{13}\)

### 4.6.3 Results

We discuss now the two main results which are useful to tackle the problem of full implementation in Section 4.8. The first result shows that \( BR \) characterizes the set of IPE-strategies across models of beliefs; the second result shows that this set can be computed by means of a convenient “backwards procedure”.

**Characterization of the set of IPE.** As emphasized above, in \( BR \) agents hold conjectures about both the opponents’ and their own strategies. First, notice that conditions (2) and (3.2) in the definition of \( BR \) maintain that agents are always certain of their own strategy; furthermore, the definition of sequential best response (def. 4.6) depends only on the marginals of the conditional conjectures over \( \Theta^* \times S_{-i} \). Hence, this particular feature of \( BR \) does not affect the standard notion of rationality. The fact

\(^{13}\)In games of complete information, an instance of the same principle is provided by *subgame perfection*, where agents believe in the equilibrium continuation strategies both on- and off-the-equilibrium path. The belief persistence hypothesis goes hand in hand with the logic of *backward induction*, allowing to envision each subgame “in isolation”. (cf. Perea, 2010, and discussion below, Section 4.9.)
that conjectures are elements of $\Delta^i (\Theta^* \times S)$ rather than $\Delta^i (\Theta^* \times S_{-i})$ corresponds to the assumption, discussed in Section 4.6.1, that IPE treats all deviations the same; its implication is that both histories arising from unexpected moves of the opponents and from one’s own deviations represent zero-probability events, allowing the same set of conditional beliefs about $\Theta^* \times S_{-i}$, with essentially the same freedom that IPE allows after anyone’s deviation. This is the main insight behind the following result (the proof is in Appendix B.2.1):

**Proposition 4.1** (Characterization). Fix a game $(\mathcal{E}, \mathcal{M})$. For each $i$: $s_i \in \mathcal{BR}_i$ if and only if $\exists \mathcal{B}, b_i \in B_i$ and $(\hat{\sigma}, \hat{p})$ such that: (i) $(\hat{\sigma}, \hat{p})$ is an IPE of $(\mathcal{E}, \mathcal{M}, \mathcal{B})$ and (ii) $s_i \in \text{supp} \, \hat{\sigma}_i (b_i)$.

An analogous result can be obtained for the more standard refinement of IPE, in which unilateral own deviations leave an agents’ beliefs unchanged, applying to a modified version of $\mathcal{BR}$: such modification entails assuming that agents only form conjectures about $\Theta^* \times S_{-i}$ (that is: $\mu^i \in \Delta^i (\Theta^* \times S_{-i})$) and by consequently adapting conditions (2) and (3) in the definition of $\mathcal{BR}$ (see Penta, 2009a). Hence, the assumption that IPE treats anyone’s deviation the same (and, correspondingly, that in $\mathcal{BR}$ agents hold conjectures about their own strategy as well) is not crucial to characterize the set of equilibrium strategies across models of beliefs. As already discussed in Section 4.6.1, it is crucial instead for the next result, which shows that such set can be computed applying a procedure that extends the logic of backward induction to environments with incomplete information (proposition 4.2 below).

**The “Backwards Procedure”**. The backwards procedure is described as follows: Fix a public history $h^{T-1}$ of length $T - 1$. For each payoff-type $y^T_i \in \Theta^*_i$ of each agent, the continuation game is a static game, to which we can apply the standard notion of
Δ-rationalizability (Battigalli and Siniscalchi, 2003). For each $h_{T-1}$, let $\mathcal{R}_i^{h_{T-1}}$ denote the set of pairs $(y_i^T, s_i|h_{T-1})$ such that continuation strategy $s_i|h_{T-1}$ is rationalizable in the continuation game from $h_{T-1}$ for type $y_i^T$. We now proceed backwards: for each public history $h_{T-2}$ of length $T - 2$, we apply again Δ-rationalizability to the continuation game from $h_{T-2}$ (in normal form), restricting continuation strategies $s_i|h_{T-2} \in S_i^{h_{T-2}}$ to be Δ-rationalizable in the continuation games from histories of length $h_{T-1}$. $\mathcal{R}_i^{h_{T-2}}$ denotes the set of pairs $(y_i^{T-1}, s_i|h_{T-2})$ such that continuation strategy $s_i|h_{T-2}$ is rationalizable in the continuation game from $h_{T-2}$ for “type” $y_i^{T-1}$. Inductively, this is done for each $h_{t-1}$, until the initial node $\phi$ is reached, for which the set $\mathcal{R}_i^\phi$ is computed. We can now introduce the “backwards procedure” result (the proof and the formal definition of $\mathcal{R}_i^\phi$ are in appendix B.2.2):

**Proposition 4.2** (Computation). $\mathcal{BR}_i = \mathcal{R}_i^\phi$ for each $i$.

Properties UIA and CCFR provide the basic insight behind this result. First, notice that under UIA, the set of beliefs agents are allowed to entertain about the opponents’ payoff types (i.e. the support of their marginal beliefs over $\Theta^*_i$) is the same at every history (equal to $\Theta^*_i$). Hence, in this respect, their information about the opponents’ types in the subform starting from (public) history $h_{t-1}$ is the same as if the game started from $h_{t-1}$. Also, CCFR implies that agents’ epistemic assumptions about everyone’s behavior in the continuation is also the same at every history. Thus, UIA and CCFR combined imply that, from the point of view of $\mathcal{BR}$, a continuation from history $h_{t-1}$ is equivalent to a game with the same space of uncertainty and strategy spaces equal to the continuation strategies, which justifies the possibility of analyzing continuation games “in isolation”.\(^{14}\)

\(^{14}\)Hence, $\mathcal{BR}$ satisfies a property that generalizes the notion of “subgame consistency”, according
4.7 Partial Implementation

Under our assumption that the designer can commit to the mechanism, it is easy to show that a revelation principle holds for dynamic environments, so that restricting attention to direct mechanisms (definition 4.1) entails no loss of generality for the analysis of the partial implementation problem.

The notion of implementation adopted by the classical literature on static mechanism design is that of interim incentive compatibility:

**Definition 4.8.** A SCF is interim implementable (or interim incentive compatible) on $B = \langle N, \Theta^*, (B_i, \beta_i)_{i \in N} \rangle$ if truthful revelation is an interim equilibrium of $(\mathcal{E}, \mathcal{M}^*, B)$. That is, $\exists \sigma^* \in \Sigma^*$ such that for each $i \in N$ and $b_i \in B_i$, for all $\sigma_i \in \Sigma_i$,

$$
\int_{\Theta^* \times B_{-i}} U_i(\sigma^*, \theta, b_{-i}; b_i) \cdot d\beta(b_i) \\
\geq \int_{\Theta^* \times B_{-i}} U_i(\sigma_i, \sigma^*_{-i}, \theta, b_{-i}; b_i) \cdot d\beta(b_i).
$$

(Recall that $\Sigma^*$ denotes the set of truth-telling strategies.) Bergemann and Morris (2005) showed that a SCF is interim incentive compatible on all type spaces, if and only if it is ex-post incentive compatible, that is:

**Definition 4.9.** A SCF $f$ is ex post implementable (or ex post incentive compatible) if for each $i$, for each $\theta \in \Theta^*$ and $s'_i \in S_i$,

$$
U_i(s^*, \theta) \geq U_i(s'_i, \sigma^*_{-i}, \theta).
$$

We say that a SCF is Strictly Ex-Post Incentive Compatible if for any $s'_i \notin S_i^*$, the inequality holds strictly.

to which ‘the behavior prescribed on a subgame is nothing else than the solution of the subgame itself’ (Harsanyi and Selten, 1988, p.90).
Interim incentive compatibility imposes no requirement of perfection: If players cannot commit to their strategies, more stringent incentive compatibility requirements must be introduced, to account for the dynamic structure of the problem. We thus apply the solution concept introduced in Section 4.6.1, IPE: A mechanism is interim perfect implementable if the truth-telling strategy is an IPE of the direct mechanism.

**Definition 4.10.** A SCF is interim perfect implementable (or interim perfect incentive compatible) on \( B = \langle N, \Theta^*, (B_i, \beta_i)_{i \in N} \rangle \) if there exist beliefs \( (p^i)_{i \in N} \) and \( \sigma^* \in \Sigma^* \) such that, \( (\sigma^*, p) \) is an IPE of \( (\mathcal{E}, \mathcal{M}^*, B) \).

For a given model of beliefs, interim perfect incentive compatibility is clearly more demanding than interim incentive compatibility. But, as the next result shows, the requirement of “perfection” is no more demanding than the “ex-ante” incentive compatibility if it is required for all models of beliefs:

**Proposition 4.3** (Partial Implementation). A SCF is perfect implementable on all models of beliefs if and only if it is ex post implementable.

Hence, as far as “robust” partial implementation is concerned, assuming that agents can commit to their strategies is without loss of generality: The dynamic mechanism can be analyzed in its normal form.

On the other hand, in environments with dynamic revelation of information, agents’ signals are intrinsically multidimensional. Hence, given proposition 4.3, the negative result on ex-post implementation by Jehiel et al. (2006) can be interpreted as setting tight limits for the Wilson’s doctrine applied to dynamic mechanism design problems. However, the literature provides examples of environments of economic interest where ex-post implementation with multidimensional signals is possible (e.g. Picketty, 1999; Bikhchandani, 2006; Eso and Maskin, 2002).
4.8 Full Implementation in Direct Mechanisms

We begin by focusing on direct mechanisms. Unlike the static case of Bergemann and Morris (2009), in environments with dynamic revelation of information direct mechanisms may not suffice to achieve full robust implementation: Section 4.8.4 shows how simple “enlarged” mechanisms, can improve on the direct ones, yet avoiding the intricacies of the “augmented mechanisms” required for classical Bayesian Implementation.\(^{15}\)

**Definition 4.11.** SCF \(f\) is fully perfectly implementable in the direct mechanism if for every \(\mathcal{B}\), the set of IPE-strategies of \((\mathcal{E}, \mathcal{M}^*, \mathcal{B})\) is included in \(\Sigma^*\).

The following proposition follows immediately from proposition 4.1.

**Proposition 4.4.** SCF \(f\) is (fully) robustly perfect-implementable in the direct mechanism if and only if \(\mathcal{BR} \subseteq S^*\).

### 4.8.1 Environments with Monotone Aggregators of Information

In this Section it is maintained that each set \(\Theta_{i,t} = [\theta_{i,t}^l, \theta_{i,t}^h] \subseteq \mathbb{R}\), so that, for each \(t = 1, ..., T\), \(Y^t \subseteq \mathbb{R}^{nt}\). Environments with monotone aggregators are characterized by the property that for each agent, in each period, all the available information (across time and agents) can be summarized by a one-dimensional statistic. Furthermore, such \(T\) statistics uniquely determine an agent’s preferences. (This notion generalizes properties of preferences discussed in the example in Section 4.2).

\(^{15}\)Classical references are Postlewaite and Schmeidler (1988), Palfrey and Srivastava (1989) and Jackson (1991).
Definition 4.12. An Environment admits monotone aggregators (EMA) if, for each $i$, and for each $t = 1, \ldots, T$, there exists an aggregator function $\alpha^t_i : Y^t \to \mathbb{R}$ and a valuation function $v_i : \Xi^* \times \mathbb{R}^T \to \mathbb{R}$ such that $\alpha^t_i$ and $v_i$ are continuous, $\alpha^t_i$ is strictly increasing in $\theta_{i,t}$ and for each $(\xi^*, \theta^*) \in \Xi^* \times \Theta^*$,

$$u_i (\xi^*, \theta^*) = v_i \left( \xi^*, (\alpha^t_i (y^* (\theta^*)))_{\tau=1}^T \right).$$

Assuming the existence of the aggregators and the valuation functions, per se, entails no loss of generality: the bite of the representation derives from the continuity assumptions and the further restrictions on the aggregator functions that will be imposed in the following.

The self-correcting strategy. The analysis is based on the notion of self-correcting strategy, $s^c$, which generalizes what we have already described in the leading example of Section 4.2: at each period-$t$ history, $s^c_i$ reports a message such that the implied period-$t$ valuation is “as correct as it can be”, given the previous reports. That is: conditional on past truthful revelation, $s^c_i$ truthfully reports $i$’s period-$t$ signal; at histories that come after previous misreports of agent $i$, $s^c_i$ entails a further misreport, to offset the impact on the period-$t$ aggregator of the previous misreports.\footnote{An earlier formulation of the idea of self-correcting strategy can be found in Pavan (2007). I thank Alessandro Pavan for pointing this out.} Formally:

Definition 4.13. The self-correcting strategy, $s^c_i \in S_i$, is such that for each $t = 1, \ldots, T$ and public history $h^{t-1} = (\tilde{y}^{t-1}, x^{t-1})$, and for each private history $h^t_i = (h^{t-1}_i, y^t_i)$,

$$s^c_i (h^t_i) = \arg \min_{m_{i,t} \in \Theta_{i,t}} \left\{ \max_{y^t_{i,-i} \in Y^t_{i,-i}} \left| \alpha^t_i (y^t_i, y^t_{i,-i}) - \alpha^t_i (\tilde{y}^{t-1}_i, m_{i,t}, y^t_{i,-i}) \right| \right\}. \quad (4.14)$$
Clearly, $s^c$ induces truthful reporting (that is: $s^c \in S^*$): if $h^{t-1} = (\tilde{y}^{t-1}, x^{t-1})$ and $y^t_i = (\tilde{y}^{t-1}_i, \theta_{i,t})$, then $s^c_i(h^t, y^t_i) = \theta_{i,t}$. Also, notice that $s^c_i(h^t)$ only depends on the component of the public history made of $i$’s own reports, $\tilde{y}^{t-1}_i$. Let $\tilde{y}^{t-1}_{-i}$ be such that: 

$$\tilde{y}^{t-1}_{-i} \in \arg\max_{y^{t-1}_{-i} \in Y^{t-1}_{-i}} \left| \alpha^t_i(y^t_i, y^{t-1}_{-i}) - \alpha^t_i(\tilde{y}^{t-1}_i, s^c_i(h^t), y^{t-1}_{-i}) \right| .$$

Then, by definition of $s^c_i$ and the fact that $\alpha^t_i$ is strictly increasing in $\theta_{i,t}$, we may have three cases:

$$\alpha^t_i(y^t_i, y^{t-1}_{-i}) = \alpha^t_i(\tilde{y}^{t-1}_i, s^c_i(h^t), y^{t-1}_{-i}) \text{ for all } y^{t-1}_{-i} \in Y^{t-1}_{-i}, \quad (4.15)$$

$$\alpha^t_i(y^t_i, \tilde{y}^{t-1}_{-i}) > \alpha^t_i(\tilde{y}^{t-1}_i, s^c_i(h^t), \tilde{y}^{t-1}_{-i}) \text{ and } s^c_i(h^t) = \theta^+_{i,t}, \quad (4.16)$$

$$\alpha^t_i(y^t_i, \tilde{y}^{t-1}_{-i}) < \alpha^t_i(\tilde{y}^{t-1}_i, s^c_i(h^t), \tilde{y}^{t-1}_{-i}) \text{ and } s^c_i(h^t) = \theta^-_{i,t}. \quad (4.17)$$

Equation (4.15) corresponds to the case in which strategy $s^c_i$ can completely offset the previous misreports. But there may exist histories at which no current report can offset the previous misreports. In the example of Section 4.2, suppose that the first period under- (resp. over-) report is so low (resp. high), that even reporting the highest (lowest) possible message in the second period is not enough to “correct” the implied value of $\varphi$. This was the case corresponding to the possibility of corner solutions, and corresponds cases to (4.16) and (4.17) respectively.

**The Contraction Property.** The results on full implementation are based on a *contraction property* that limits the dependence of agents’ aggregator functions on the private signals of the opponents. Before formally introducing the contraction property, some extra notation is needed: for each set of strategy profiles $D = \times_{j \in I} D_j \subseteq S$ and

150
for each private history $h^t_i$, let
\[
D_i(h^t_i) := \{ m_{i,t} : \exists s_i \in D_i, \text{s.t. } s_i(h^t_i) = m_{i,t} \}
\]
and
\[
D_i(h^{t-1}) := \bigcup_{y^t_i \in Y^t_i} D_i(h^{t-1}, y^t_i).
\]
Define also:
\[
s_i[D(h^{t-1})] := \{ (m_{i,t}, y^t_i) \in M_{i,t} \times Y^t_i : m_{i,t} \in D_i(h^{t-1}, y^t_i) \}
\]
and
\[
s^c_i[h^{t-1}] := \{ (m_{i,t}, y^t_i) \in M_{i,t} \times Y^t_i : m_{i,t} = s^c_i(h^{t-1}, y^t_i) \}
\]

**Definition 4.14 (Contraction Property).** An environment with monotone aggregators of information satisfies the Contraction Property if, for each $D \subseteq S$ such that $D \neq \{s^c\}$ and for each $h^{t-1} = (\tilde{y}^{t-1}, x^{t-1})$ such that $s[D(h^{t-1})] \neq s^c[h^{t-1}]$, there exists $y^t_i$ and $m'_{i,t} \in D_i(h^{t-1}, y^t_i)$, $m'_{i,t} \neq s^c_i(h^{t-1}, y^t_i)$, such that:
\[
\text{sign} \left[ s^c_i(h^{t-1}, y^t_i) - \theta'_{i,t} \right] = \text{sign} \left[ \alpha^t_i(y^t_i, y^t_{-i}) - \alpha^t_i(\tilde{y}^{t-1}, \theta'_{i,t}, \theta'_{-i,t}) \right] \tag{4.18}
\]
for all $y^t_{-i} = (y_{-i}^{t-1}, \theta_{-i,t}) \in Y^t_{-i}$ and $m'_{-i,t} \in D_{-i}(h^{t-1}, y^t_{-i})$.

To interpret the condition, rewrite the argument of the right-hand side of (4.18) as follows:
\[
\alpha^t_i(y^t_i, y^t_{-i}) - \alpha^t_i(\tilde{y}^{t-1}, m'_{i,t}, m'_{-i,t})
\]
\[
= \left[ \alpha^t_i(\tilde{y}^{T-1}, s^c_i(h^{T-1}, y^t_i), s^c_{-i}(h^{T-1}, y^t_{-i})) - \alpha^t_i(\tilde{y}^{t-1}, m'_{i,t}, m'_{-i,t}) \right]
\]
\[
+ \kappa(h^{t-1}, y^t_i, y^t_{-i}) \tag{4.19}
\]
where
\[
\kappa(h^{t-1}, y^t_i, y^t_{-i}) = \alpha^t_i(y^t_i, y^t_{-i}) - \alpha^t_i(\tilde{y}^{t-1}, s^c(h^{t-1}, y^t_i, y^t_{-i})) \tag{4.20}
\]
The term in the first square bracket in (4.19) represents the impact, on the period-
t aggregator, of a deviation (in the set $D$) from the self-correcting profile at history $h^{t-1}$;
the term $\kappa \left( h^{t-1}, y^t_i, y^t_{-i} \right)$ represents the extent by which the self-correcting profile is
incapable of offsetting the previous misreports. Suppose that $\kappa \left( h^{t-1}, y^t_i, y^t_{-i} \right) = 0$, i.e.
strategy profile $s^c_i$ fully offsets the previous misreports (in particular, this is the case
if $h^{t-1}$ is a truthful history: $y^{t-1} = y^{t-1}$), then, the contraction property boils down
to the following:


t(Simple CP) For each public history at which the behavior allowed
by the set of deviations $D$ is different from $s^c$, there exists at least one
player’s “type” $y^t_i$ of some agent $i$, for which for some $m^t_{i,t} \in D_i \left( h^{t-1}, y^t_i \right)$,
$\alpha^t_i \left( y^t_i, y^t_{-i} \right) - \alpha^t_i \left( \tilde{y}^{t-1}, m^t_{i,t}, m^t_{-i,t} \right)$ is unilaterally signed by $\left[ s^c_i \left( h^{t-1}, y^t_i \right) - m^t_{i,t} \right]$, 
uniformly over the opponents private information and current reports.

From equations (4.15)-(4.17) it is easy to see that $\kappa \left( h^{t-1}, y^t_i, y^t_{-i} \right) = 0$ whenever
$s^c_i \left( h^t_i \right) \in \left( \theta^t_{i,t} \right.$, $\theta^t_{i,t} \left. \right)$. Hence, this corresponds precisely to the case considered in the
example of Section 4.2. For histories such that the self-correcting strategy is not
sufficient to offset the previous misreports, then the simple CP must be strength-
ened so that the sign of the impact of deviations from $s^c$ at $h^{t-1}$ on the aggregator
$\alpha^t_i$ is not upset by the previous misreports, $\kappa$. So, in principle, the bound on the
interdependence in agents’ valuations may depend on the histories of payoff signals.
Section 4.8.4 though will show how simple “enlarged” mechanisms, in which agents’
sets of messages are extended at every period so that any possible past misreport can
be “corrected”, eliminate this problem, inducing $\kappa \left( h^{t-1}, y^t_i, y^t_{-i} \right) = 0$ at all histories.
Given the simplicity of their structure, such mechanisms will be called “quasi-direct”.
4.8.2 Aggregator-Based SCF

Consider the SCF in the example of Section 4.2 (equations 4.2-4.5): the allocation chosen by the SCF in period $t$, is only a function of the values of the aggregators in period $t$. The notion of aggregator-based SCF generalizes this idea:

**Definition 4.15.** The SCF $f = (f_t)_{t=1}^T$ is aggregator-based if for each $t$, $\alpha^t_i (y^t) = \alpha^t_i (\tilde{y}^t)$ for all $i$ implies $f_t (y^t) = f_t (\tilde{y}^t)$.

The next proposition shows that, if the contraction property is satisfied, an aggregator-based SCF is fully implementable in environments that satisfy a single-crossing condition:

**Definition 4.16 (SCC-1).** An environment with monotone aggregators of information satisfies SCC-1 if, for each $i$, valuation function $v_i$, is such that: for each $t$, and $\xi, \xi' \in \Xi^*: \xi_{\tau} = \xi'_{\tau}$ for all $\tau \neq t$, then for each $a^*_{i_{t-1}} \in \mathbb{R}^{T-1}$ and for each $\alpha_i < \alpha'_i < \alpha''_i$,

$$v_i (\xi, \alpha_i, a^*_{i_{t-1}}) > v_i (\xi', \alpha_i, a^*_{i_{t-1}}) \text{ and } v_i (\xi, \alpha'_i, a^*_{i_{t-1}}) = v_i (\xi', \alpha'_i, a^*_{i_{t-1}})$$

implies: $v_i (\xi, \alpha''_i, a^*_{i_{t-1}}) < v_i (\xi', \alpha''_i, a^*_{i_{t-1}})$

In words: For any two allocations $\xi$ and $\xi'$ that only differ in their period-$t$ component, for any $a^*_{i_{t-1}} \in \mathbb{R}^{T-1}$, the difference $\delta_{i,t} (\xi, \xi', \alpha_i) = v_i (\xi, \alpha_i, a^*_{i_{t-1}}) - v_i (\xi', \alpha_i, a^*_{i_{t-1}})$ as a function of $\alpha_i$ crosses zero (at most) once (see figure 4.1.a, p. 156). We are now in the position to present the first full-implementation result:

**Proposition 4.5.** In an environment with monotone aggregators (def. 4.12) satisfying SCC-1 (def. 4.16) and the contraction property (def. 4.14), if an aggregator-based social choice function satisfies Strict EPIC (definition 4.9), then $\mathcal{BR} = \{s^c\}$.

The argument of the proof is analogous to the argument presented in Section 4.2: For each history of length $T - 1$, it is proved that the contraction property and SCC-1
imply that agents play according to $s^c$ in the last stage; then the argument proceeds by induction: given that in periods $t+1, \ldots, T$ agents follow $s^c$, a misreport at period $t$ only affects the period-$t$ aggregator (because the SCF is “aggregator-based”). Then, SCC-1 and the contraction property imply that the self-correcting strategy is followed at stage $t$.

**An Appraisal of the “aggregator-based” assumption.** Consider the important special case of *time-separable preferences*: Suppose that, for each $i$ and $t = 1, \ldots, T$, there exist an “aggregator” function $\alpha_i^t : Y^t \to \mathbb{R}$ and a valuation function $u_i^t : \Xi^t \times \mathbb{R} \to \mathbb{R}$ such that for each $(\xi^*, \theta^*) \in \Xi^* \times \Theta^*$,

$$
u_i (\xi^*, \theta^*) = \sum_{t=1}^T u_i^t (\xi_i^*, \alpha_i^t (y_i^t (\theta^*))) .$$

In this case, the condition that the SCF is *aggregator-based* (def. 4.15) can be interpreted as saying that the SCF only responds to changes in preferences: If two distinct payoff states $\theta$ and $\theta'$ induce the same preferences over the period-$t$ allocations, then the SCF chooses the same allocation under $\theta$ and $\theta'$ in period $t$. This is the case of the example in Section 4.2. These preferences though cannot accommodate phenomena of “path-dependence” such as “learning-by-doing”. For instance, in the context of the example of Section 4.2, suppose that agents’ preferences are the following:

$$u_i (q_1, q_2, \pi_{i,1}, \pi_{i,2}, \theta) = \alpha_{i,1} (\theta_1) \cdot q_1 + \pi_{i,1} + [\alpha_{i,2} (\theta_1, \theta_2) \cdot F (q_1) \cdot q_2 + \pi_{j,2}] .$$

\footnote{In that example the set of allocations includes the transfers, hence for each $t$ the social choice function is: $f_i (\theta) = (q_i^t (\theta), \pi_{i,t}^* (\theta), \pi_{j,t}^* (\theta))$. The first component is clearly aggregator-based (see equations 4.2 and 4.3); furthermore, if $\gamma \in [0, 1)$, the values of the aggregators uniquely determine the size of the transfers (equations 4.4 and 4.5). The social choice function is thus “aggregator-based”.}
The marginal utility of $q_2$ now also depends on the amount of public good provided in the first period. Then, the optimal policy for the second period is to set $q_2^* (\theta) = [\alpha_{i,t} (\theta) + \alpha_{j,t} (\theta)] \cdot F (q_t)$. This rule is not aggregator-based, as the period-2 allocation choice depends on both the period-2 aggregators and the previous period allocation. Thus, to allow the SCF to respond to possible “path-dependencies” in agents’ preferences (such as “learning-by-doing” effects) it is necessary to relax the “aggregator-based” assumption.

In environments with transferable utility (such as the example above) our notion of SCF includes the specification of the transfers scheme: $f_t (\theta) = (q_t (\theta), \pi_{i,t} (\theta), \pi_{j,t} (\theta))$ for each $t$. Since the requirement that the SCF is aggregator-based applies to all its components, it also applies to transfers. In general, it is desirable to allow for arbitrary transfers, not necessarily aggregator-based. The general results of the next section can be easily adapted to accommodate the possibility of arbitrary transfers in environments with transferable utility (Section 4.8.3).

### 4.8.3 Relaxing “Aggregator-Based”

In the proof of proposition 4.5, the problem with relaxing the assumption that the SCF is “aggregator-based” is that a one-shot deviation from $s^c$ at period-$t$ may induce different allocations in period-$t$ and in subsequent periods. Hence, the “within period” single-crossing condition (SCC-1) may not suffice to conclude the inductive step, and guarantee that strategy $s^c$ is played at period-$t$: Some bound is needed on the impact that a one-shot deviation has on the outcome of the SCF. The next condition guarantees that the impact of a one-shot deviation is not too strong.

**Definition 4.17 (SCC-2).** An environment with monotone aggregators of information satisfies SCC-2 if, for each $i$: for each $\theta, \theta' \in \Theta^*$ such that $\exists t \in \{1, ..., T\} : y^* (\theta) =
for all $\tau < t$ and for all $j$, $\alpha_j^\tau (\theta) = \alpha_j^\tau (\theta')$ for all $\tau > t$, then for each $a_{i,-t}^* \in \mathbb{R}^{T-1}$ and for each $\alpha_{i,t} < \alpha'_{i,t} < \alpha''_{i,t}$,

\begin{align*}
&\text{if } v_i \left( f (\theta) , \alpha_{i,t}^* , a_{i,-t}^* \right) > v_i \left( f (\theta') , \alpha_{i,t} , a_{i,-t}^* \right) \\
&\text{and } v_i \left( f (\theta) , \alpha'_{i,t} , a_{i,-t}^* \right) = v_i \left( f (\theta') , \alpha'_{i,t} , a_{i,-t}^* \right) , \\
&\text{then } v_i \left( f (\theta') , \alpha''_{i,t} , a_{i,-t}^* \right) < v_i \left( f (\theta') , \alpha''_{i,t} , a_{i,-t}^* \right)
\end{align*}

SCC-2 compares the allocations chosen for any two “similar” states of nature: states $\theta$ and $\theta'$ are “similar” in the sense that they are identical up to period $t-1$, and imply the same value of the aggregators at all periods other than $t$. Since agents’ preferences are uniquely determined by the values of the aggregators (definition 4.12), the preferences induced by states $\theta$ and $\theta'$ only differ along the dimension of the period-$t$ aggregator. The condition requires a single-crossing condition for the corresponding outcomes to hold along this direction. The condition is easily interpretable from a graphical viewpoint: suppose that $\theta$ and $\theta'$ are as in definition 4.17. Then, if the SCF is “aggregator-based” and the environment satisfies SCC-1 (definition 4.16), the difference in payoffs for $f (\theta)$ and $f (\theta')$ as a function of the period-$t$ aggregator crosses zero (at most) once. (Figure 1.a). If $f$ is not “aggregator based”, allocations at periods $\tau > t$ may differ under $f (\theta)$ and $f (\theta')$, shifting (or changing the shape) of the curve $\delta_{i,t} (f (\theta) ; f (\theta'), \alpha_{i,t})$. SCC-2 guarantees that such shifting maintains the single-crossing property (figure 4.1.b).

(The “path-dependent” preferences in equation (4.21) satisfy SCC-2 for any choice of $F : \mathbb{R} \to \mathbb{R}$.)

**Proposition 4.6 (Full Implementation).** In an environment with monotone aggregators (def. 4.12) satisfying the contraction property (def. 4.14), if a SCF $f$ is Strictly EPIC (definition 4.9) and satisfies SCC-2 (def. 4.17), then $\mathcal{BR} = \{ \sigma^c \}$. 

156
Corollary 4.1. Since \( s^c \in S^* \), if the assumptions of propositions 4.5 or 4.6, then \( f \) is fully robustly implementable.

Transferable Utility.

A special case of interest is that of additively separable preferences with transferable utility: For each \( t = 1, \ldots, T \), the space of allocations is \( \Xi_t = Q_t \times (\times_{i=1}^n \Pi_{i,t}) \), where \( Q_t \) is the set of “common components” of the allocation and \( \Pi_{i,t} \subseteq \mathbb{R} \) is the set of transfers to agent \( i \) (\( i \)'s “private component”). Maintaining the restriction that the environment admits monotone aggregators, agent \( i \)'s preferences are as follows: For each \( \xi^* = (q_t, \pi_{1,t}, \ldots, \pi_{n,t})_{t=1}^T \in \Xi^* \) and \( \theta^* \in \Theta^* \),

\[
    u_i(\xi^*, \theta^*) = \sum_{t=1}^T v^t_i((q_t)_{\tau=1}^t, (y^t(\theta^*))) + \pi_{i,t},
\]

where for each \( t = 1, \ldots, T \), \( v^t_i : (\times_{\tau=1}^t Q_\tau) \times \mathbb{R} \to \mathbb{R} \) is the period-\( t \) valuation of the common component. Notice that functions \( v^t_i : (\times_{\tau=1}^t Q_\tau) \times \mathbb{R} \to \mathbb{R} \) are defined over the entire history \( (q_1, \ldots, q_t) \): this allows period-\( t \) valuation of the current allocation.
(q_t) to depend on the previous allocative decisions (q_1, ..., q_{t-1}). This allows us to accommodate the “path dependencies” in preferences discussed above.\(^{18}\)

In environments with transferable utility, it is common to define a social choice function only for the common component, \(\chi_t : Y^t \rightarrow Q_t\) \((t = 1, ..., T)\), while transfer schemes \(\pi_{i,t} : Y^t \rightarrow \mathbb{R}\) \((i = 1, ..., n\) and \(t = 1, ..., T)\) are specified as part of the mechanism. Not assuming transferable utility, social choice functions above were defined over the entire allocation space \((f_t : Y^t \rightarrow \Xi_t)\), they thus include transfers in the case of transferable utility. The transition from one approach to the other is straightforward. Any given pair of choice function and transfer scheme \((\chi_t, (\pi_{i,t})_{i=1}^n)_{t=1}^T\) trivially induces a social choice function \(f^{\chi,\pi}_t : Y^t \rightarrow \Xi_t\) \((t = 1, ..., T)\) in the setup above: for each \(t\) and \(y'_t \in Y^t\), \(f^{\chi,\pi}_t(y'_t) = (\chi_t(y'_t); (\pi_{i,t}(y'_t))_{i=1}^n)\).

It is easy to check that, in environments with transferable utility, if agents’ preferences over the common component \(Q^* = \times_{t=1}^T Q_t\) satisfy (SCC-1), and \(\chi : \Theta^* \rightarrow Q\) is aggregator-based, then for any transfer scheme \((\pi_{i,t}(y'_t))_{i=1}^n\), the “full” social choice function \(f^{\chi,\pi}\) satisfies (SCC-2). More generally, if \(\chi\) and agents’ preferences over \(Q^*\) satisfy (SCC-2), then \(f^{\chi,\pi}\) satisfies (SCC-2) for any transfer scheme \((\pi_{i,t}(y'_t))_{i=1}^n\).

Given this, the following corollary of proposition 4.6 is immediate:

**Corollary 4.2.** In environments with monotone aggregators of information and transferable utility, if agents’ preferences over \(Q^*\) and \(\chi : \Theta^* \rightarrow Q^*\) satisfy: (i) the contraction property; (ii) the single crossing condition (SCC-2); and (iii) there exist transfers \(\pi\) that make \(\chi\) strictly ex-post incentive compatible; then \(f^{\chi,\pi}\) is fully robustly implemented.

\(^{18}\)The special case of “path-independent” preferences corresponding to the example in section 4.2 is such that period-t valuation are functions \(v_t^i : Q_t \times \mathbb{R} \rightarrow \mathbb{R}\).
4.8.4 “Quasi-direct” Mechanisms

This section shows how simple “enlarged” mechanisms may avoid incurring into corner solutions, which allows us to relax the bite of the contraction property (definition 4.14) by guaranteeing that the sign condition holds with \( \kappa(h^{t-1}, y^t) = 0 \) at every history (equation 4.20), thus weakening the sufficient condition for full implementation.

Let \( \hat{\alpha}_{i,t} : \mathbb{R}^n \rightarrow \mathbb{R} \) be a continuous extension of \( \alpha_{i,t} : Y^t \rightarrow \mathbb{R} \) from \( Y^t \) to \( \mathbb{R} \), strictly increasing in the component that extends \( \theta_{i,t} \) and constant in all the others on \( \mathbb{R} \setminus Y^t \) (from definition 4.12 \( \alpha_{i,t} \) is only assumed strictly increasing in \( \theta_{i,t} \) on \( Y^t \)).

Set \( m_{i,1}^- = \theta_{i,1}^- \) and \( m_{i,1}^+ = \theta_{i,1}^+ \), and for each \( t = 1, \ldots, T \), let \( \hat{\Theta}_{i,t} = [m_{i,t}, m_{i,t}^+] \), and \( \hat{Y}_i^t = \times_{\tau=1}^t \hat{\Theta}_{i,\tau} \) where \( m_{i,t}^- \) and \( m_{i,t}^+ \) are recursively defined so to satisfy:

\[
\begin{align*}
    m_{i,t}^+ &= \max \left\{ m_i \in \mathbb{R} : \max_{(y_{i,t}, y_{i-1}) \in Y^t} \left| \hat{\alpha}_{i,t} (y_{i,t}, y_{i-1}) - \min_{\hat{g}_{t-1} \in \hat{Y}_{i,t-1}} \hat{\alpha}_{i,t} (\hat{y}_{t-1}^{t-1}, m_i, y_{i-1}) \right| = 0 \right\} \\
    m_{i,t}^- &= \min \left\{ m_i \in \mathbb{R} : \max_{(y_{i,t}, y_{i-1}) \in Y^t} \left| \hat{\alpha}_{i,t} (y_{i,t}, y_{i-1}) - \max_{\hat{g}_{t-1} \in \hat{Y}_{i,t-1}} \hat{\alpha}_{i,t} (\hat{y}_{t-1}^{t-1}, m_i, y_{i-1}) \right| = 0 \right\}
\end{align*}
\]

Set the message spaces in the mechanism such that \( M_{i,t} = \hat{\Theta}_{i,t} \) for each \( i \) and \( t \). By construction, for any private history \( h_i^t = (h_i^{t-1}, y_i^t) \), the self-correcting report \( s_i^c (h_i^t) \) satisfies equation (4.15), that is \( s^c \) is capable of fully offset previous misreports: messages in \( \hat{\Theta}_{i,t} \setminus \Theta_{i,t} \) are used whenever equations (4.16) or (4.17) would be the case in the direct mechanism. (Clearly, such messages never arise if \( s^c \) is played.) To complete the mechanism, we need to extend the domain of the outcome function to account for these “extra” messages. Such extension consists of treating these reports in terms of the implied value of the aggregator: For given sequence of reports \( \hat{y}_i^t \in \hat{Y}^t \) such that some message in \( \hat{\Theta}_{i,t} \setminus \Theta_{i,t} \) has been reported at some period \( \tau \leq t \), let \( g_t (\hat{y}_i^t) = f_t (\theta) \) for some \( \theta \) such that \( \alpha_{i,\tau} (\theta) = \alpha_{i,\tau} (\hat{y}_i^\tau) \) for all \( i \) and \( \tau \leq t \), \( f_i (\theta) = f_i (\theta') \).
4.9 Further Remarks on the Solution Concepts

Backwards procedure, Subgame-Perfect Equilibrium and IPE. In games with complete and perfect information, the “backwards procedure” $\mathcal{R}^\phi$ coincides with the backward induction solution, hence with subgame perfection.$^{19}$ The next example (borrowed from Perea, 2010) shows that if the game has complete but imperfect information, strategies played in Subgame-Perfect Equilibrium (SPE) may be a strict subset of $\mathcal{R}^\phi$:

Example 4.1.

Consider the game in the following figure:

```
\begin{center}
\begin{tikzpicture}
\node (1) at (0,0) {1};
\node (2) at (-1,-1) {c};
\node (3) at (1,-1) {d};
\node (4) at (2,-2) {e};
\node (5) at (-2,-2) {f};
\node (6) at (0,-2) {g};
\node (7) at (1,-2) {h};
\draw (1) -- (2);
\draw (1) -- (3);
\draw (1) -- (4);
\draw (1) -- (5);
\draw (1) -- (6);
\draw (1) -- (7);
\end{tikzpicture}
\end{center}
```

Backwards Procedure, SPE and IPE

$\mathcal{R}^\phi_1 = \{bc, bd, ac\}$ and $\mathcal{R}^\phi_2 = \{f, g\}$. The game though has only one SPE, in which player 1 chooses $b$: in the proper subgame, the only Nash equilibrium entails the mixed (continuation) strategies $\frac{1}{2}c + \frac{1}{2}d$ and $\frac{3}{4}f + \frac{1}{4}g$, yielding a continuation payoff of $\frac{11}{4}$ for player 1. Hence, player 1 chooses $b$ at the first node.$^\Box$

$^{19}$For the special case of games with complete information, Perea (2010) independently introduced a procedure that is equivalent to $\mathcal{R}^\phi$, and showed that $\mathcal{R}^\phi$ coincides with the backward induction solution if the game has perfect information.
In games with complete information, IPE coincides with SPE, but $\mathcal{R}^\phi$ in general is weaker than subgame perfection. At first glance, this may appear in contradiction with propositions 4.1 and 4.2, which imply that $\mathcal{R}^\phi$ characterizes the set of strategies played in IPE across models of beliefs. The reason is that even if the environment has no payoff uncertainty ($\Theta^*$ is a singleton), the complete information model in which $B_i$ is a singleton for every $i$ is not the only possible: models with redundant types may exist, for which IPE strategies differ from the SPE-strategies played in the complete information model. The source of the discrepancy is analogous to the one between Nash equilibrium and subjective correlated equilibrium (Aumann, 1974).

We illustrate the point constructing a model of beliefs and an IPE in which strategy $(ac)$ is played with positive probability by some type of player 1.20 Let payoffs be the same as in example 4.1, and consider the model $B$ such that $B_1 = \{b_{bc}^1, b_{ac}^1\}$ and $B_2 = \{b_{f2}^1, b_{g2}^1\}$, with the following beliefs:

$\beta_1(b_1)[b_{f2}^1] = \begin{cases} 1 & \text{if } b_1 = b_{bc}^1, b_{ac}^1 \\ 0 & \text{otherwise} \end{cases}$

and

$\beta_2(b_{g2}^1)[b_{ad}^1] = 1, \beta_2(b_{f2}^1)[b_{bc}^1] = 1$

The equilibrium strategy profile $\sigma$ is such that $\forall i, \forall b_i, \sigma_i(b_i^{s_i}) = s_i$. The system of beliefs agrees with the model’s beliefs at the initial history, hence the beliefs of types $b_{f2}^1$ and $b_{ad}^1$ are uniquely determined by Bayesian updating. For types $b_i^{s_i} \neq b_{g2}^1, b_{ad}^1$, it is sufficient to set $p_i(b_i^{s_i}, a_i) = \beta_i(b_i^{s_i})$ (i.e. maintain whatever the beliefs at the beginning of the game were) Then, it is easy to verify that $(\theta, p)$ is an IPE.

On the other hand, if $|\Theta^*| = 1$ and the game has perfect information (no stage

---

20It is easy to see that such difference is not merely due to the possibility of zero-probability types. Also the relaxation of the common prior assumption is not crucial.
with simultaneous moves), then \( R^\phi \) coincides with the set of SPE-strategies. Hence, in environments with no payoff uncertainty and with perfect information, only SPE-strategies are played in IPE for any model of beliefs.

4.10 Concluding Remarks

On the Solution Concepts. Proposition 4.1 can be seen as the dynamic counterpart of Brandenburger and Dekel’s (1987) characterization of correlated equilibrium. As discussed above, the adoption of IPE in this paper is motivated by the result in proposition 4.2, which makes the full implementation problem of Section 4.8 tractable. The weakness of IPE (relative to other notions of perfect Bayesian equilibrium) is key to that result: the heart of proposition 4.2 is \( BR \)’s property of “subgame consistency” (cf. footnote 14), which allows us to analyze continuation games “in isolation”, in analogy with the logic of backward induction. The CCFR and UIA assumptions (p. 143) provide the epistemic underpinnings of the argument. To understand this point, it is instructive to compare \( BR \) with Battigalli and Sinicalschi’s (2007) weak and strong versions of extensive form rationalizability (EFR), which correspond respectively to the epistemic assumptions of (initial) common certainty of rationality (CCR) and common strong belief in rationality (CSBR): \( BR \) is stronger than the first, and weaker than the latter. The strong version of EFR fails the property of “subgame consistency” because it is based on a forward induction logic, which inherently precludes the possibility of envisioning continuations “in isolation”: by taking into account the possibility of counterfactual moves, agents may draw inferences from their opponents’ past moves and refine their conjectures on the behavior in the continuation. The weak version of EFR fails “subgame consistency” for opposite reasons: an agent can make weaker predictions on the opponents’ behavior in the continuation...
than he would make if he envisioned the continuation game “in isolation”, because no restrictions on the agents’ beliefs about their opponents’ rationality are imposed after an unexpected history. Thus, the form of “backward induction reasoning” implicit in IPE (which generalizes the idea of subgame perfection) is based on stronger (respectively, weaker) epistemic assumptions than CCR (respectively, CSBR).

**Dynamic Mechanisms in Static Environments.** Consider an environment in which agents obtain all the relevant information before the planner has to make a decision. The designer may still have reasons to adopt a dynamic mechanism (e.g. an ascending auction).\(^{21}\) In the context of an environment with complete information, Bergemann and Morris (2007) recently argued that dynamic mechanisms may improve on static ones by reducing agents’ strategic uncertainty: They showed how the backward induction outcome of a second-price clock-auction guarantees full robust implementation of the efficient allocation for a larger set of parameters than the rationalizable outcomes of a second price sealed-bid auction. The approach of this paper allows us to extend the analysis to incomplete information settings: It can be shown that, with incomplete information, the ascending clock-auction does not improve on the static one. The reason is that the logic of backward induction loses its bite when the assumption of complete information is relaxed.\(^{22}\) In incomplete information environments, the case for the role of dynamic mechanisms in reducing strategic uncertainty must rely on stronger solution concepts (e.g. based on forward induction reasoning), that allow agents to draw stronger inferences on their opponents’ private information from their past moves (see Mueller, 2009).

\(^{21}\)**In the formal setup of the paper, this amounts to a situation in which \(|\Theta_t| = 1\) for all \(t > 1\) and \(|\Xi_t| = 1\) for all \(t < T\).

\(^{22}\)**See Kunamoto and Tercieux (2009) for a related negative result.
Appendices
Appendix A

Appendix to Chapter 3

A.1 Proofs of results from Section 3.5

A.1.1 Proof of proposition 3.1

The proof is by induction. The initial step is trivial, for $\mathcal{ISR}^0$ is vacuously u.h.c.

For the inductive step, suppose that $\mathcal{ISR}^{k-1}$ is u.h.c., and let $\{t_i^m\}$ be s.t. $t_i^m \to t_i$, $\{s_i^m\}$ s.t. $s_i^m \to \hat{s}_i$ and $s_i^m \in \mathcal{ISR}_i^k(t_i^m)$ for all $m$. Then, for each $m$, there exists $\mu^{i,m} \in \Delta^\mathcal{H}(\Theta_0 \times T^*_{-i} \times \Theta \times S_{-i})$ s.t.

$(1)$ $s_i^m \in r_i(\mu^{i,m} | t_i^m)$

$(2)$ $\tau_{t_i^m}^{i,m} = \arg\min_{\Theta_0 \times T^*_{-i} \times \Theta \times S_{-i}} \mu^{i,m} (\cdot | \phi)$

$(3)$ $\text{supp}(\mu^{i,m} (\cdot | \phi)) \subseteq \Theta_0 \times \mathcal{ISR}_i^{k-1}$

We want to show that $\hat{\hat{s}}_i \in \mathcal{ISR}_i^k(t_i)$, i.e. that $\exists \hat{\mu}^i \in \Delta^\mathcal{H}(T^*_{-i} \times S_{-i})$ s.t.

$(1')$ $\hat{\hat{s}}_i \in r_i(\hat{\mu}^i | \hat{t}_i)$

$(2')$ $\tau_{\hat{t}_i}^* = \arg\min_{\Theta_0 \times T^*_{-i} \times \Theta \times S_{-i}} \hat{\mu}^i (\cdot | \phi)$

$(3')$ $\text{supp}(\hat{\mu}^i (\cdot | \phi)) \subseteq \Theta_0 \times \mathcal{ISR}_i^{k-1}$

Consider the sequence $\{\mu^{i,m}\} \subseteq \Delta^\mathcal{H}(\Theta_0 \times T^*_{-i} \times \Theta \times S_{-i})$. Since $\{\mu^{i,m}\}$ is in a compact, there exists a subsequential limit $\hat{\mu}^i$. By continuity of $\tau^*$, $\hat{\mu}^i$ satisfies condition $(2')$, since (2) holds for each $\mu^{i,m}$ and $t_i^m \to t_i$. Furthermore, since the best-response correspondence $r_i (\cdot)$ is u.h.c. and $s_i^m \in r_i(\mu^{i,m} | t_i^m)$ for each $m$, also condition $(1')$ is satisfied. Given that $\mu^{i,m} \to \hat{\mu}^i$, the upper hemicontinuity of $\mathcal{ISR}_i^{k-1}$ (from the inductive hypothesis) suffices for $(3')$.■
A.1.2 Proof of proposition 3.2

We prove that, for each \( k \), \( (\hat{\pi}^t (t_i))_{t=0}^k = (\hat{\pi}^t (\bar{t}_i))_{t=0}^k \) implies \( \mathcal{I} \mathcal{S} \mathcal{R}^{-k} (t_i) = \mathcal{I} \mathcal{S} \mathcal{R}^{-k} (\bar{t}_i) \).

This is proven by induction: suppose that it holds for \((k - 1)\) steps, that \( \hat{\pi}^k (t_i) = \hat{\pi}^k (\bar{t}_i) \) and that \( \hat{s}_i \in \mathcal{I} \mathcal{S} \mathcal{R}^{\bar{t}_i} (t_i) \). Then:

\[
\exists \mu^i \in \Delta^\mathcal{H} (\Theta_0 \times T_{-i} \times S_{-i}) \text{ s.t.} \\
(1). \hat{s}_i \in \tau_i (\mu^i | t_i) \\
(2). \tau_i = \operatorname{mrg} (\Theta_0 \times T_{-i} \mu^i (\phi)) \\
(3). \int_{\Theta_0 \times \mathcal{I} \mathcal{S} \mathcal{R}^{\bar{t}_i} (t_i)} \mu^i (\phi) = 1
\]

We want to show that \( \hat{s}_i \in \mathcal{I} \mathcal{S} \mathcal{R}^{\bar{t}_i} (\bar{t}_i) \). That is, we will show that:

\[
\exists \mu^i \in \Delta^\mathcal{H} (\Theta_0 \times \bar{T}_{-i} \times S_{-i}) \text{ s.t.} \\
(1)^\prime. \hat{s}_i \in \tau_i (\mu^i | \bar{t}_i) \\
(2)^\prime. \tau_i (\bar{t}_i) = \operatorname{mrg} (\Theta_0 \times \bar{T}_{-i} \mu^i (\phi)) \\
(3)^\prime. \int_{\Theta_0 \times \mathcal{I} \mathcal{S} \mathcal{R}^{\bar{t}_i} \mu^i (\phi) = 1
\]

Let \( D_{-i}^{k-1} = \left\{ (\hat{\pi}_{-i}^t (t_{-i}))_{t=0}^{k-1} : t_{-i} \in T_{-i} \right\} \), and \( \rho_{-i, k-1} : \operatorname{supp} (\tau_i (t_i)) \rightarrow \Theta_0 \times D_{-i}^{k-1} \) the map \( (\theta_0, t_{-i}) \mapsto (\theta_0, (\hat{\pi}_{-i}^t (t_{-i}))_{t=0}^{k-1}) \). Similarly define \( \tilde{\rho}_{-i, k-1} : \operatorname{supp} (\tau_i (\bar{t}_i)) \rightarrow \Theta_0 \times D_{-i}^{k-1} \). For each \( h \in \mathcal{H} \), let \( \varphi_i^{k-1} (h) \in \Delta (\Theta_0 \times D_{-i}^{k-1}) \) be the pushforward of \( \operatorname{mrg} (\Theta_0 \times T_{-i} \mu^i (h)) \) under \( \rho_{-i, k-1} \). Under the inductive hypothesis, for every measurable \( B \subseteq \Theta_0 \times D_{-i}^{k-1} \), \( \tau_i (\bar{t}_i) [\tilde{\rho}_{-i, k-1}^{-1} (0) [B]] = \varphi_i^{k-1} (\phi) [B] \).

Define \( \sigma_i^{k-1} : \Theta_0 \times D_{-i}^{k-1} \times S_{-i} \rightarrow [0, 1] \) be such that such that for every \( s_{-i} \in S_{-i} \) and measurable \( B \subseteq \Theta_0 \times D_{-i}^{k-1} \),

\[
\int_{(\theta_0, d_{-i}^{k-1}) \in B} \sigma_i^{k-1} (\theta_0, d_{-i}^{k-1}, s_{-i}) = \begin{cases} \\ \frac{\mu^i (\phi) [\rho_{-i, k-1}^{-1} (B)]}{\varphi_i^{k-1} (\phi) [B]} & \text{if } \varphi_i^{k-1} (\phi) [B] > 0 \\
0 & \text{otherwise} \\
\end{cases}
\]

Let \( \tilde{\mu}^i \in \Delta^\mathcal{H} (\Theta_0 \times \bar{T}_{-i} \times S_{-i}) \) be s.t. for each measurable \( B \subseteq \Theta_0 \times \bar{T}_{-i} \) and \( s_{-i} \in S_{-i} (h) \)

\[
\tilde{\mu}^i (\phi) [B \times \{ s_{-i} \}] = \tau_i (\bar{t}_i) [B] \int_{(\theta_0, d_{-i}^{k-1}) \in \rho_{-i, k-1} (B)} \sigma_i^{k-1} (\theta_0, d_{-i}^{k-1}, s_{-i})
\]

166
By definition of CPS, $\tilde{\mu}^i(\phi)$ pins down conditional beliefs at all histories reached with positive probability. At all other histories, let $\tilde{\mu}^i(h)$ be the pushforward of $\mu^i(h)$ under the map $\nu: supp\mu^i(h) \rightarrow \Theta_0 \times \tilde{T}_{-i} \times S_{-i}$ defined as

$$(\theta_0, t_{-i}, s_{-i}) \mapsto (\theta_0, \kappa_{-i}(t_{-i}), s_{-i})$$

where $\kappa_{-i}(t_{-i}) \in \tilde{T}_{-i}$ is such that $(\hat{\pi}_{-i}^i(t_{-i}))_{l=0}^{k-1} = (\hat{\pi}_{-i}^i(\kappa_{-i}(t_{-i})))_{l=0}^{k-1}$ (this is possible since $(\hat{\pi}^i(t_i))_{l=0}^k = (\hat{\pi}^i(\tilde{\tilde{t}}_i))_{l=0}^k$). First, by construction, $\tilde{\mu}^i \in \Phi^\tilde{T}(\tilde{\tilde{t}}_i)$ and (under the inductive hypothesis) $supp\tilde{\mu}^i(\phi) \subseteq \Theta_0 \times ISR^k_{-i}$. Also, for all histories $h$ that receive zero probability under $\tilde{\mu}^i(\phi)$, $\tilde{\mu}^i(h)$ induces the same distribution over $\Theta_0 \times \Theta_{-i} \times S_{-i}$ (the payoff relevant variables) as $\mu^i(h)$. If we show that this is true also at $\phi$, it follows that $\hat{s}_i \in r_i(\mu^i|t_i)$ implies $\hat{s}_i \in r_i(\tilde{\mu}^i|\tilde{t}_i)$ (since $\theta_i(t_i) = \theta_i(\tilde{\tilde{t}}_i)$ by assumption $\hat{\pi}_{-i}^0(t_{-i}) = \hat{\pi}_{-i}^0(\tilde{\tilde{t}}_{-i})$). In fact, for each $B \subseteq \Theta_0 \times D^0_{-i}$ (and $\Theta_0 \times \Theta_{-i}$)

$$\tilde{\mu}^i(\phi) [\varrho_{-i,0}^{-1}(B) \times \{s_{-i}\}]$$

$$= \tau_i(\tilde{\tilde{t}}_i) [\varrho_{-i,0}^{-1}(B)] \cdot \int_B d\sigma^0_i(\cdot, s_{-i})$$

$$= \tau_i(\tilde{\tilde{t}}_i) [\varrho_{-i,0}^{-1}(B)] \cdot \mu^i(\phi) [\varrho_{-i,0}^{-1}(B) \times \{s_{-i}\}] \frac{\varphi^0_i(\phi) [B]}{\varphi^0_i(\phi) [B]}$$

$$= \mu^i(\phi) [\varrho_{-i,0}^{-1}(B) \times \{s_{-i}\}]$$

This completes the proof. ■

A.1.3 Proof of proposition 3.3

The proof is by induction: Let $t^*_i \in T^*_i \Theta_\ast$. Clearly, if $\theta_i(t_i) = \hat{\pi}^0_i(t^*_i)$, $ISR^k_{\Theta^{\ast}}(t_i) = ISR^k_{\Theta^{\ast}}(t^*_i)$. As inductive hypothesis, assume that $(\hat{\pi}^n_i(t_i))_{n=0}^{k-1} = (\hat{\pi}^n_i(t^*_i))_{n=0}^{k-1}$ implies that $ISR^k_{\Theta^{\ast}}(t_i) = ISR^k_{\Theta^{\ast}}(t^*_i)$. Suppose that $\hat{\pi}^n_i(t_{-i})_{n=0}^k = (\hat{\pi}^n_i(t^*_i))_{n=0}^k$. It will be shown that $(\hat{\pi}^n_i(t_i))_{n=0}^k = (\hat{\pi}^n_i(t^*_i))_{n=0}^k$ implies that

$$ISR^k_{\Theta^{\ast}}(t_i) = ISR^k_{\Theta^{\ast}}(t^*_i)$$.
Under the inductive hypothesis, \( s_{-i} \in ISR_{T_{0}, k}^{t_{-i}}(t_{-i}) \) for some \( t_{-i} \in supp \tau_{i}(t_{i}) \) if and only if \( s_{-i} \in ISR_{T_{-i_{0}}, k}^{t_{-i}}(t_{*}) \) for some \( t_{*} \in supp \tau_{i}(t_{i}^{*}) \). In PV-environments only the conjectures about \( S_{-i} \) are payoff relevant for player \( i \) (\( \theta_{-i} \)'s don't affect \( i \)'s payoffs, and \( \Theta_{0} \) is a singleton.). Thus, under the inductive hypothesis, \( \exists \mu_{i} \in \Phi_{i}(t_{i}) \) s.t. \( supp(marg_{T_{-i_{0}, \Theta_{i}} \times s_{-i}, \mu_{i}}) \subseteq ISR_{T_{-i_{0}, \Theta_{i}}, k+1}^{t_{i}} \) and \( s_{i} \in r_{i}(\mu_{i}|t_{i}) \) if and only if \( \exists \hat{\mu}_{i} \in \Phi_{i}(t_{i}) \) s.t. \( supp(marg_{T_{-i_{0}, \Theta_{i}} \times s_{-i}, \hat{\mu}_{i}}) \subseteq ISR_{T_{-i_{0}, \Theta_{i}}, k+1}^{t_{i}} \) and \( s_{i} \in r_{i}(\hat{\mu}_{i}|t_{i}^{*}) \). (Remember the only restrictions on the conjectures over \( S_{-i} \) imposed by \( ISR \) are at the beginning of the game). Hence \( ISR_{T_{0}, k+1}^{t_{i}}(t_{i}) = ISR_{T_{0}, k+1}^{t_{i}^{*}}(t_{i}^{*}) \).\[ \]

A.2 Proofs of results from Section 3.6

A.2.1 Proof of lemma 3.2

(Part I:) Fix \( t_{i} \in \hat{T}_{i_{0}, \Theta_{i}} \). For each \( k \neq i \), let \( \Theta_{k}^{'} \) be the finite set of payoff states that receive positive probability by \( t_{i} \in T_{j}, j \neq k \), and let \( \Theta_{i} = \Theta_{i}^{'} \cup \bar{\Theta}_{i} \). (\( \Theta_{i} \) is finite because \( t_{i} \) is a finite type and \( \bar{\Theta}_{i} \) is finite). \( \forall s_{i} \in ISR(t_{i}), \exists \mu_{i}^{s_{i}} \in \Phi_{i}(t_{i}) \) s.t. (1) \( s_{i} \in r_{i}(\mu_{i}^{s_{i}|t_{i}}) \) and (2) \( supp(\mu_{i}^{s_{i}}(\phi)) \subseteq \Theta_{0} \times ISR_{-i} \). Given a probability space \( (\Omega, \mathcal{B}) \) and a set \( A \subseteq \mathcal{B} \), denote by \( v_{[A]} \) the uniform probability distribution concentrated on \( A \). For each \( \varepsilon \in [0, 1] \), consider the set of types profiles \( \times_{i \in N} T_{i}^{\varepsilon} \subseteq T_{0}^{e} \), s.t. each \( T_{i}^{\varepsilon} \) consists of all the types \( \tilde{t}_{i} \in \hat{T}_{i} \) and of types \( \bar{t}_{i} (t_{i}, s_{i}, \varepsilon) \) s.t.:

\[
\theta_{i} \left( \bar{t}_{i} (t_{i}, s_{i}, \varepsilon) \right) = \varepsilon \theta_{i} + (1 - \varepsilon) \theta_{i} (t_{i})
\]

and

\[
\tau_{i}^{\varepsilon} \left( \bar{t}_{i} (t_{i}, s_{i}, \varepsilon) \right) = \varepsilon v_{[\{\theta_{0}\} \times T_{-i}(s_{i})]} + (1 - \varepsilon) \left[ \mu_{i}^{s_{i}}(\phi) \circ \tilde{T}_{-i}^{-1,\varepsilon} \right]
\]

where \( T_{-i} \subseteq T_{-i}^{\varepsilon} \) is the subset of dominance-types profiles defined above, and

\[
\hat{t}_{-i, \varepsilon} : \Theta_{0} \times T_{-i} \times S_{-i} \to \Theta_{0} \times T_{-i}^{\varepsilon} \text{ is s.t.}
\]

\[
\hat{t}_{-i, \varepsilon} (\theta_{0}, s_{-i}, t_{-i}) = (\theta_{0}, \bar{t}_{-i} (t_{-i}, s_{-i}, \varepsilon))
\]

168
By construction, with probability $\varepsilon$, type $\bar{i}_i (t_i, s_i, \varepsilon)$ is certain that $s_i$ is conditionally dominant, and puts positive probability to all of the opponents’ dominance types in $T_{-i}$. Define $\gamma : \Theta_0 \times T^\varepsilon_{-i} \rightarrow \Theta_0 \times T^\varepsilon_{-i} \times S_{-i}$ s.t.:

$$\forall \bar{i}_{-i} (t_{-i}, s_{-i}, \varepsilon) \in T^\varepsilon_{-i} :$$

$$\gamma (\theta_0, \bar{i}_{-i} (t_{-i}, s_{-i}, \varepsilon)) = (\theta_0, \bar{i}_{-i} (t_{-i}, s_{-i}, \varepsilon), s_{-i})$$

and for every $\bar{t}^s_{-i} \in \overline{T}_{-i} \subseteq T^\varepsilon_{-i}$,

$$\gamma (\theta_0, \bar{t}^s_{-i}) = (\theta_0, \bar{t}^s_{-i}, s_{-i})$$

Consider the conjectures $\hat{\mu}^i \in \Delta^H (\Theta_0 \times T^\varepsilon_{-i} \times S_{-i})$ defined by:

$$\hat{\mu}^i (\phi) = (\tau^e_i (\bar{i}_i (t_i, s_i, \varepsilon)) \circ \gamma^{-1}) \in \Delta (\Theta_0 \times T^\varepsilon_{-i} \times S_{-i})$$

For any $\varepsilon > 0$, the conjectures $\hat{\mu}^i$ are such that $T_{-i} \subseteq \text{supp} \left(\text{marg}_{T^\varepsilon_{-i}} \hat{\mu}^i (\phi)\right)$. From the definition of $\gamma$, it follows that $\text{supp} \left(\text{marg}_{S_{-i}} \hat{\mu}^i (\phi)\right) = S_{-i}$, so that the entire CPS $(\hat{\mu}^i (h))_{h \in H}$ can be obtained via Bayes’ Rule. This also implies that $\hat{\mu}^i$ satisfies condition (3) in the definition of $SSR$. Furthermore, by construction, $\hat{\mu}^i \in \Phi_i (\bar{i}_i (t_i, s_i, \varepsilon))$, and $\forall \varepsilon > 0$, $\forall h \in H$, $\exists \eta^{\varepsilon,h} \in (0, 1)$ s.t. $\eta^{\varepsilon,h} \rightarrow 0$ as $\varepsilon \rightarrow 0$ and

$$\text{marg}_{\Theta_0 \times T^*_{-i} \times S_{-i}} \hat{\mu}^i (h) = \eta^{\varepsilon,h} \text{marg}_{\Theta_0 \times T^h_{-i} (s_i) \times S_{-i} (h)} \left[\left\{\theta^s_0\right\} \times \bar{t}^h_{-i} (s_i) \times S_{-i} (h)\right]$$

$$+ (1 - \eta^{\varepsilon,h}) \text{marg}_{\Theta_0 \times S_{-i}} \mu^{s_i} (h)$$

where $T^h_{-i} (s_i) = \left\{\bar{t}^{(s_i, s_{-i})}_{-i} : s_{-i} \in S_{-i} (h)\right\}$. Hence, the conditional conjectures $\hat{\mu}^i (h)$ of type $\bar{t}_{-i} (t_{-i}, s_{-i}, \varepsilon)$ are a mixture: with probability $(1 - \eta^{\varepsilon,h})$ they agree with $\mu^i (\phi)$, which made $s_i$ sequential best response; with probability $\eta^{\varepsilon,h}$ they are concentrated on payoff states $\left\{\theta^s_0\right\} \times \left\{\theta_{-i} (t_{-i}) : t_{-i} \in \bar{t}^h_{-i} (s_i)\right\}$, which together with the fact that $\theta_i (\bar{i}_i (t_i, s_i, \varepsilon)) = \varepsilon \theta^s_i + (1 - \varepsilon) \theta_i (t_i)$ breaks the all ties in favor of $s_i$ so that
\( r_i (\hat{\mu}_i^s (t_i, s_i, \varepsilon)) = \{s_i\} \). Thus, \( s_i \in SS\mathcal{R}_i (\bar{r}_i (t_i, s_i, \varepsilon)) \), so that (ii) and (iii) in the lemma are satisfied for all \( \varepsilon > 0 \).

The remainder of the proof guarantees that also part (i) in the lemma holds, and it is identical to WY’s counterpart.

\textbf{(Part II:) } it will be shown that \( \hat{\pi}_i^* (\bar{r}_i (t_i, s_i, \varepsilon)) \to \hat{\pi}_i^* (t_i) \) as \( \varepsilon \to 0 \). By construction, \( \tau_i (\bar{r}_i (t_i, s_i, \varepsilon)) \) are continuous in \( \varepsilon \), hence \( \hat{\pi}_i^* (\bar{r}_i (t_i, s_i, \varepsilon)) \to \hat{\pi}_i^* (\bar{r}_i (t_i, s_i, 0)) \) as \( \varepsilon \to 0 \) (Brandenburger and Dekel, 1993). It suffices to show that \( \hat{\pi}_i^* (\bar{r}_i (t_i, s_i, 0)) = \hat{\pi}_i^* (t_i) \) for each \( t_i \) and \( i \). This is proved by induction. The payoff types and the first order beliefs are the same. For the inductive step, assume that \( \hat{\pi}_i^k (\bar{r}_i (t_i, s_i, 0)) \), \( (\hat{\pi}_i^l (t_i))^k_{l=0} \), we will show that \( \hat{\pi}_i^k (\bar{r}_i (t_i, s_i, 0)) = \hat{\pi}_i^k (t_i) \). Define

\[ D_{-i}^{k-1} = \left\{ \left( \hat{\pi}_{-i}^l (t_{-i}) \right)_{l=0}^{k-1} : t_{-i} \in T_{-i} \right\}. \]

Under the inductive hypothesis, it can be shown (see WY) that

\[ \text{marg}_{\Theta_0 \times D_{-i}^{k-1}} [\mu^i (\phi) \circ \hat{r}_{-i}^{-1}] = \text{marg}_{\Theta_0 \times D_{-i}^{k-1}} [\mu^i (\phi)]. \] (\( \diamond \))

Therefore:

\[ \hat{\pi}_i^k (\bar{r}_i (t_i, s_i, 0)) = \text{marg}_{\Theta_0 \times D_{-i}^{k-1}} [\mu^i (\phi) \circ \hat{r}_{-i}^{-1}] \]

\[ = \text{marg}_{\Theta_0 \times D_{-i}^{k-1}} [\mu^i (\phi)] \]

\[ = \mu^i (\phi) \]

\[ = \hat{\pi}_i^k (t_i) \]

where the first equality is the definition of \( k \)-th level belief; the second from (\( \diamond \)); the third from the inductive hypothesis; the fourth from the fact that \( \mu^i \in \Phi (t_i) \); the last one again by definition. \( \blacksquare \)
A.2.2 Proof of Lemma 3.3

The proof is by induction: For $k = 0$, let $\tilde{t}_i$ be s.t. $\theta_i (\tilde{t}_i) = \theta_i (\tilde{t}_i)$ and $\tau_i (\tilde{t}_i) = u[\{\theta^i \times T_{-i}(s_i)\}]$. Clearly, $\mathcal{ISR}_i^1 (\tilde{t}_i) = \{s_i\}$ and condition (1) is satisfied. Fix $k > 0$, write each $t_{-i} = (\lambda, \kappa)$ where $\lambda = \{\hat{\pi}^k_i (t_{-i})\}_{k=1}^{k-1}$ and $\kappa = \{\hat{\pi}^k_i (t_{-i})\}_{k=k}^{\infty}$. Let $L_{-i}^{k-1} = \{\lambda : \exists \kappa \text{ s.t. } (\lambda, \kappa) \in T_{-i}^*\}$. As inductive hypothesis, assume that: for each finite $t_{-i} = (\lambda, \kappa)$ and $s_{-i} \in \mathcal{SSR}^{k-1}_{-i} (t_{-i})$ s.t. $T_{-i} (s_{-i}) \subseteq \text{supp} \left( \text{marg}_{T_{-i}} \mu^{s_i} (\phi) \right)$, there exists finite $t_{-i}^{s_{-i}} = (\lambda, \kappa^{s_{-i}-\lambda})$ s.t. $\mathcal{ISR}^{k}_{i} (t_{-i}^{s_{-i}}) = \{s_i\}$. Take any $s_{-i} \in \mathcal{SSR}^{k}_{i} (\tilde{t}_{i})$ s.t. $T_{-i} (s_{-i}) \subseteq \text{supp} \left( \text{marg}_{T_{-i}} \mu^{s_i} (\phi) \right)$: we will construct a type $\tilde{t}_{i}$ s.t. for each $k' \leq k$, $\hat{\pi}^{k'}_{i} (\tilde{t}_{i}) = \hat{\pi}^{k'}_{i} (\tilde{t}_{i})$, $\mathcal{ISR}^{k+1}_{i} (\tilde{t}_{i}) = \{s_i\}$: By definition, if $s_{-i} \in \mathcal{SSR}^{k}_{i} (\tilde{t}_{i})$, $\exists \mu^{s_i} \in \Delta^H (\Theta_0^* \times T_{-i}^* \times S_{-i})$ s.t.

1. $\tau_i (\tilde{t}_{i}) = \text{marg}_{\Theta \times T_{-i}} \mu^{s_i} (\phi)$
2. $\text{supp} (\mu^{s_i} (\phi)) \subseteq \Theta_0^* \times \mathcal{SSR}^{k-1}_{-i}$
3. $\{s_i\} = r_i (\mu^{s_i} | \tilde{t}_{i})$

Using the inductive hypothesis, define the mapping

$$\varphi : \bigcup_{h \in \mathcal{H}} \left[ \text{supp} \left( \text{marg}_{\Theta_0^* \times L_{-i}^{k-1} \times S_{-i}} \mu^{s_i} (h) \right) \right] \rightarrow \Theta_0^* \times T_{-i}^*$$

such that: $\varphi (\theta, \lambda, s_{-i}) = (\theta, (\lambda, \kappa^{s_{-i}-\lambda}))$

Define type $\tilde{t}_{i}$ as

$$\tau_i (\tilde{t}_{i}) = \text{marg}_{\Theta_0^* \times L_{-i}^{k-1} \times S_{-i}} \mu^{s_i} (\cdot | \phi) \circ \varphi^{-1}$$

$$= \mu^{s_i} (\phi) \circ \text{proj}_{\Theta_0^* \times L_{-i}^{k-1} \times S_{-i}}^{-1} \circ \varphi^{-1}$$

171
By construction (for the inductive hypothesis), the first $k$ orders of beliefs are the same for $t_i$ and $\tilde{t}_i$ (which is point 1 in the lemma)

$$\hat{\pi}_i^k (\tilde{t}_i) = \text{marg}_{\Theta^* \times L_{-i}^{k-1}} \tau_i$$

$$= \mu^{s_i} (\phi) \circ \text{proj}^{-1}_{\Theta^* \times L_{-i}^{k-1}} \circ \varphi^{-1} \circ \text{proj}^{-1}_{\Theta^* \times L_{-i}^{k-1}}$$

$$= \mu^{s_i} (\phi) \circ \text{proj}^{-1}_{\Theta^* \times L_{-i}^{k-1}}$$

$$= (\mu^{s_i} (\phi) \circ \text{proj}^{-1}_{\Theta^* \times T_{-i}}) \circ \text{proj}^{-1}_{\Theta^* \times L_{-i}^{k-1}}$$

$$= \text{marg}_{\Theta^* \times L_{-i}^{k-1}} \tau_i$$

$$= \hat{\pi}_i^k (t_i)$$

Where the first equality is by definition, the second is from construction of $\tau_i (\tilde{t}_i)$ above; the third is from the definition of $\varphi$, for which

$$\text{proj}^{-1}_{\Theta_0^* \times L_{-i}^{k-1}} \circ \varphi^{-1} \circ \text{proj}^{-1}_{\Theta^* \times L_{-i}^{k-1}} = \text{proj}^{-1}_{\Theta_0^* \times L_{-i}^{k-1}}$$

The fourth and fifth are simply notational, and the last one by definition. We need to show that $\mathcal{ISR}^{k+1}_i (\tilde{t}_i) = \{ s_i \}$. To this end, notice that each $(\theta_0, t_{-i}) \in \text{supp}(\tau_i (\tilde{t}_i))$ is of the form $(\theta_0, (\lambda, \kappa_{s_{-i}}))$, and it is s.t. $\mathcal{ISR}^k_{-i} ((\lambda, \kappa_{s_{-i}})) = \{ s_{-i} \}$. Hence, the array of conditional conjectures (i.e. CPS) that are consistent with $\tilde{t}_i$ and with the restrictions $\mathcal{ISR}^k_{-i}$ are $\tilde{\mu}_i$ s.t.

$$\tau_i (\tilde{t}_i) = \text{marg}_{\Theta_0^* \times T_{-i}} \tilde{\mu}_i (\phi)$$

and

$$\tilde{\mu}_i (\phi) [\Theta_0^* \times \{(t_{-i}, s_{-i}) : \mathcal{ISR}^k_{-i} (t_{-i}) = \{ s_{-i} \}\}] = 1$$

hence, the conditional conjectures are uniquely determined for all $h \in \mathcal{H} (s_{-i})$ for some $s_{-i} : \{ s_{-i} \} = \mathcal{ISR}^k_{-i} (t_{-i})$ and $t_{-i} \in \text{supp} (\text{marg}_{T_{-i}} \tau_i (\tilde{t}_i))$. But since, from hypothesis, $T_{-i} (s_i) \subseteq \text{supp} (\text{marg}_{T_{-i}} \mu^{s_i})$, and that by definition of $T_{-i} (s_i)$ we trivially have that
\[ \bigcup_{t_{-i} \in T_{-i}(s_i)} \mathcal{ISR}^{k+i}_{-i} (t_{-i}) = S_{-i}, \] for all \( h \in \mathcal{H} \). These conjectures are given by \( \tilde{\mu}^i (\phi) = \tau_i (\tilde{t}_i) \circ \kappa^{-1}, \) with \( \kappa \) defined as

\[ \kappa (\theta_0, (\lambda, \kappa^{s_{-i}^\lambda})) = (\theta_0, (\lambda, \kappa^{s_{-i}^\lambda}), s_{-i}) \]

Furthermore, for each \( h \):

\[ \text{marg}_{\Theta^\ast \times S_{-i}} \tilde{\mu}^i (h) = \text{marg}_{\Theta^\ast \times S_{-i}} \mu^{s_i} (h) \]

To see this, given the observation that \( \mu^{s_i} (\phi) \) is completely mixed (i.e. all histories are reachable with positive probability), it suffices to show that

\[ \text{marg}_{\Theta^\ast \times S_{-i}} \tilde{\mu}^i (\phi) = \text{marg}_{\Theta^\ast \times S_{-i}} \mu^{s_i} (\phi). \]

But this is immediate, given that from the definition of \( \kappa \) and \( \varphi \), we have:

\[ \text{proj}^{-1}_{\Theta^\ast \times L_{-i} \times S_{-i}} \circ \kappa \circ \varphi = I. \]

(\( I \) is the identity map). Hence, \( \tilde{\mu}^i \) is uniquely determined for all \( h \), and it is equal to \( \mu^{s_i} \), to which \( s_i \) is the unique best response. Hence \( \mathcal{ISR}^{k+1}_{i} (\tilde{t}_i) = \{ s_i \} \).

The proof of statement (3) in the lemma is identical to WY’s: Define

\[ \tilde{T}_{i} = \{ \tilde{t}_i \} \cup \left( \bigcup_{(\theta, t_{-i}^s) \in \text{supp}(\tau_i(\tilde{t}_i))} T_{i}^{t_{-i}}, \right), \]

\[ \tilde{T}_{j} = \bigcup_{(\theta, t_{-i}^s) \in \text{supp}(\tau_i(\tilde{t}_i))} T_{j}^{t_{-i}} \quad \text{(for} \ j \neq i) \]

### A.3 \( \mathcal{ISR} \) in the Interim Normal Form

The analysis in this section applies to finite models in \( PV\text{-environments} \), we thus focus on Bayesian games \( \Gamma = \langle N, \mathcal{H}, Z, \Theta (T_i, \tau_i, \hat{u}_i)_{i \in N} \rangle \) s.t. \( |\Theta \times T| < \infty \) and
\[ \hat{u}_i : T_i \times \mathcal{Z} \to \mathbb{R} \text{ for each } i \in N. \]

Define the ex-ante payoffs as: For each \((t_i, s_i) \in T_i \times S_i\) and \(\sigma_i \in \Delta (T_{-i} \times S_{-i})\),

\[ U_i (s_i, \sigma_i, t_i) = \int_{T_{-i} \times S_{-i}} u_i (z (s_i, s_{-i}), \theta_i (t_i)) d \sigma_{-i} (t_{-i}, s_{-i}) \]

**Definition A.1.** A strategy \(s_i \in S_i\) is weakly dominated for \(t_i\), if for all \(\sigma_i \in \Delta (T_{-i} \times S_{-i})\) s.t. \(\text{marg}_{T_{-i}} \sigma_{-i} = \tau_i (t_i)\) and s.t. \(\sigma_i [s_{-i}] > 0\) for each \(s_{-i} \in S_{-i}\)

\[ s_i \notin \arg \max_{s'_i \in S_i} U_i (s'_i, \sigma_i, t_i) \]

Say that \(\Gamma^T\) is in generic position if for every \(t_i \in T_i\), \(z \neq z'\) implies that \(u_i (z, \theta_i (t_i)) \neq u_i (z', \theta_i (t_i))\). Notice that in PV-environments, \(t_i\)’s beliefs \(\tau_i (t_i) \in \Delta (T_{-i})\) are payoff irrelevant. The following is thus a well-known fact (e.g., lemma 1.2 in Ben-Porath, 1997).

**Lemma A.1.** If \(\Gamma^T\) is in generic position, \(s_i\) is not weakly dominated for \(t_i\) if and only if \(s_i\) is sequentially rational for \(t_i\).\(^1\)

The next definition introduces Dekel and Fudenberg’s (1990) \(S^\infty W\)-procedure for the interim normal form of \(\Gamma^T\).

**Definition A.2.** For each \(t_i \in T_i\), let \(S^0 W_i^T (t_i) \equiv r_i (t_i)\). For each \(k = 0, 1, 2, \ldots\), and \(t_i \in T_i\), let \(S^k W_i^T (t_i) = \{ s_i \in S_i : (t_i, s_i) \in S^k W_i^T \}\). \(S^k W^T = \times_{i=1,\ldots,n} S^k W_i^T\) and \(S^k W_{-i}^T = \times_{j \neq i, 0} S^k W_j^T\). Recursively, for \(k = 1, 2, \ldots\), and \(t_i \in T_i\)

\[ S^k W_i^T (t_i) = \left\{ \hat{s}_i \in S^{k-1} W_i^T (t_i) : \begin{array}{l} \exists \sigma_{-i} \in \Delta \left( S^{k-1} W_{-i}^T \right) \text{ s.t.} \\
(1). \tau_i (t_i) = \text{marg}_{T_{-i}} \sigma_{-i} \\
(2). \hat{s}_i \in \arg \max_{s'_i \in S_i} U_i (s'_i, \sigma_{-i}, t_i) \end{array} \right\} \]

Finally: \(S^\infty W_i^T (t_i) = \bigcap_{k \geq 0} S^k W_i^T (t_i)\)

\(^1\)As in section 3.3 \(\hat{u}_i (z, t_i) = u_i (z, \theta_i (t_i))\).

\(^2\)Sequentially rational strategies were defined in definition 3.8.
**Proposition A.1.** If $\Gamma^T$ is in generic position, for each $i \in N$, $t_i \in T_i$ and $k \geq 1$, $\mathcal{ISR}^T (t_i) = S^{k-1}W^T_{-i}(t_i)$. Hence $\mathcal{ISR}^T = S^\infty W^T$.

**Proof:** From Lemma A.1, $\mathcal{ISR}^T (t_i) = S^0W^T_{i}(t_i)$. In the following it will be shown that for each $k \geq 1$ and $t_i \in T_i$, $\mathcal{ISR}^T (t_i) = S^{k-1}W^T_{i}(t_i)$ implies $\mathcal{ISR}^T (t_i) = S^{k}W^T_{i}(t_i)$.

**Step 1 (⊆):** Let $\hat{s}_i \in \mathcal{ISR}^T (t_i)$ and $\mu^i \in \Phi_i (t_i)$ be s.t. $\text{supp} (\text{marg}_{S_{-i}} \mu^i (\cdot | \phi)) \subseteq \mathcal{ISR}^T (t_i)$ and $\hat{s}_i \in r_i (\mu^i | t_i)$. Set $\sigma_{-i} = \mu^i (\cdot | \phi)$. Under the inductive hypothesis, $\sigma_{-i} \in \Delta (S^{k-1}W^T_{-i})$ and trivially by construction: $\tau_i (t_i) = \text{marg}_{T_i} \sigma_{-i}$ and $\hat{s}_i \in \arg \max_{s_i \in S^{k-1}W^T_{i} (t_i)} U_i (s_i, \sigma_{-i}, t_i)$. Hence, $\hat{s}_i \in S^{k}W^T_{i} (t_i)$.

**Step 2 (⊇):** Let $\hat{s}_i \in S^{k}W^T_{i} (t_i)$ and $\hat{\sigma}_{-i} \in \Delta (S^{k-1}W^T_{-i})$ s.t. $\tau_i (t_i) = \text{marg}_{T_i} \hat{\sigma}_{-i}$ and $\hat{s}_i \in \arg \max_{s_i \in S^{k-1}W^T_{i} (t_i)} U_i (s_i, \hat{\sigma}_{-i}, t_i)$. From the inductive hypothesis, $\hat{s}_i \in S^{k}W^T_{i} (t_i) \subseteq S^{k-1}W^T_{i} (t_i) = \mathcal{ISR}^T (t_i)$, hence $\exists \mu^i \in \Phi_i (t_i)$ s.t. $\text{supp} (\text{marg}_{S_{-i}} \mu^i (\cdot | \phi)) \subseteq \mathcal{ISR}^T (t_i)$ and $\hat{s}_i \in r_i (\mu^i | t_i)$. Let $\hat{\mu}^i$ such that $\hat{\mu}^i (\cdot | \phi) = \hat{\sigma}_{-i}$, and for each $h$ at which $\hat{\mu}^i (\cdot | h)$ is not specified by Bayes’ Rule, set $\hat{\mu}^i (\cdot | h) = \mu^i (\cdot | h)$. Since $\text{supp} (\hat{\sigma}_{-i}) \subseteq S^{k-1}W^T_{-i} = \mathcal{ISR}^T (t_i)$ under the inductive hypothesis, $\hat{\mu}^i \in \Phi_i (t_i)$ and is concentrated on $\mathcal{ISR}^T (t_i)$. That also $\hat{s}_i \in r_i (\mu^i | t_i)$ holds is immediate, as $\hat{\mu}^i (\cdot | \phi)$ agree with $\hat{\sigma}_{-i}$, and the conditional conjectures at unexpected histories agree with $\mu^i$.
Appendix B

Appendix to Chapter 4

B.1 Topological structures and Conditional Probability Systems

B.1.1 Topological structures

Sets $\Theta_{i,t} \subseteq \mathbb{R}^{n_i,t}$, $\Xi_t \subseteq \mathbb{R}^{h_t}$ and $M_{i,t} \subseteq \mathbb{R}^{\nu_{i,t}}$ are non-empty and compact, for each $i$ and $t$ (Sections 4.3 and 4.4). Let $n_i = \sum_{\tau \in N} n_{i,t}$ and $\nu_i = \sum_{\tau \in N} \nu_{i,t}$. For each $h_{i,t}^\tau = \tau < t$, let $\alpha_{\tau}(h_{i,t}^\tau)$ denote the triple $(\theta_{i,\tau}, m_{\tau}, \xi_{\tau})$ consisting of $i$’s private signal at period $\tau$, the message profile and allocation chosen at stage $\tau$ along history $h_{i,t}^\tau$. For each $k \in \mathbb{N}$, let $d_{(k)}$ denote the Euclidean metric on $\mathbb{R}^k$. We endow the sets $\mathcal{H}_i$ with the following metrics, $d^i (i \in N)$, defined as: For each $h_{i,t}^\tau, h_{i,t}^{\tau'} \in \mathcal{H}_i$ (w.l.o.g.: let $\tau \geq t$)

$$d^i (h_{i,t}^\tau, h_{i,t}^{\tau'}) = \sum_{k=1}^{t-1} d_{(n_{i,k} + \nu_{i,k} + l_k)} (\alpha_k (h_{i,k}^\tau), \alpha_k (h_{i,k}^{\tau'})) + d_{n_{i,t}} (\theta_{i,t}, \theta_{i,t}') + \sum_{k=t+1}^{\tau} 1.$$

It can be checked that $(\mathcal{H}_i, d^i)$ are complete, separable metric spaces.

Sets of strategies are endowed with the supmetrics $d_{S_i}$ defined as:

$$d_{S_i} (s_i, s_i') = \sum_{t=1}^{T} \left( \sup_{h_{i,t}^\tau \in \mathcal{H}_i^{t-1} \times Y_i^t} d_{\nu_{i,t}} (s_i (h_{i,t}^\tau), s_i' (h_{i,t}^\tau)) \right)$$

Under these topological structures, the following lemma implies that CPSs introduced in Section B.1.2 are well-defined.
Lemma B.1. For all public histories \( h \in \mathcal{H} \), \( S_i(h) \) is closed.

Proof. See lemma 2.1 in Battigalli (2003).

B.1.2 Conditional Probability Systems

Let \( \Omega \) be a metric space and \( \mathcal{A} \) its Borel sigma-algebra. Fix a non-empty collection of subsets \( \mathcal{C} \subseteq \mathcal{A} \setminus \emptyset \), to be interpreted as “relevant hypothesis”. A conditional probability system (CPS hereafter) on \((\Omega, \mathcal{A}, \mathcal{C})\) is a mapping \( \mu : \mathcal{A} \times \mathcal{C} \to [0,1] \) such that:

- For all \( B \in \mathcal{C} \), \( \mu(B)[B] = 1 \)
- For all \( B \in \mathcal{C} \), \( \mu(B) \) is a probability measure on \((\Omega, \mathcal{A})\).
- For all \( A \in \mathcal{A}, B, C \in \mathcal{C} \), if \( A \subseteq B \subseteq C \) then \( \mu(B)[A] \cdot \mu(C)[B] = \mu(C)[A] \).

The set of CPS on \((\Omega, \mathcal{A}, \mathcal{C})\), denoted by \( \Delta^\mathcal{C}(\Omega) \), can be seen as a subset of \( [\Delta(\Omega)]^\mathcal{C} \) (i.e. mappings from \( \mathcal{C} \) to probability measures over \((\Omega, \mathcal{A})\)). CPS’s will be written as \( \mu = (\mu(B))_{B \in \mathcal{C}} \in \Delta^\mathcal{C}(\Omega) \). The subsets of \( \Omega \) in \( \mathcal{C} \) are the conditioning events, each inducing beliefs over \( \mathcal{A} \); \( \Delta(\Omega) \) is endowed with the topology of weak convergence of measures and \( [\Delta(\Omega)]^\mathcal{C} \) is endowed with the product topology. Below, for each player \( i \), we will set \( \Omega = \Theta^* \times S \) in games with payoff uncertainty (or \( \Omega = \Theta^* \times \Sigma \) if the game is appended with a model of beliefs). The set of conditioning events is naturally provided by the set of private histories \( \mathcal{H}_i \): for each private history \( h^i_t = (h^t_{t-1}, y^t_i) \in \mathcal{H}_i \), the corresponding event \([h^i_t]\) is defined as:

\[
[h^i_t] = \{y^t_i\} \times (\times_{\tau=t+1}^T \Theta_{i,\tau}) \times \Theta_{-i}^{\pi} \times S(h^{t-1}).
\]

Under the maintained assumptions and topological structures, sets \([h^i_t]\) are compact for each \( h^i_t \), thus \( \Delta^{\mathcal{H}_i}(\Omega) \) is a well-defined space of conditional probability systems. With a slight abuse of notation, we will write \( \mu^i(h^i_t) \) instead of \( \mu^i([h^i_t]) \).
B.2 Proofs of results from Section 4.6

B.2.1 Proof of Proposition 4.1

Proof:

Step 1: \((\Leftrightarrow)\). Fix \(B, (\hat{\sigma}, \hat{\rho})\) and \(\hat{b}_i\). For each \(h^t_i\), let \(P_i^{(\hat{\sigma}, \hat{\rho})}(h^t_i) \in \Delta(\Theta^* \times B_{-i} \times S_{-i})\) denote the probability measure on \(\Theta^* \times B_{-i} \times S_{-i}\) induced by \(\hat{p}_i(h^t_i)\) and \(\hat{\sigma}_{-i}\). For each \(j\), let

\[
\bar{S}_j = \{s_j \in S_j : \exists b_j \in B_j \text{ s.t. } s_j \in \text{supp}(\hat{\sigma}_j(b_j))\}.
\]

We will prove that \(\bar{S}_j \subseteq B_{\mathcal{R}^M_j}\) for every \(j\). For each \(h^t_i = (y^t_i, h^{t-1}_i) \in \mathcal{H}_i\), let \(\varphi^{h^t_i}_j : \bar{S}_j \to S_j(h^t_i)\) be a measurable function such that

\[
\varphi^{h^t_i}_j(s_j)(h^t_j) = \begin{cases} 
    s_j(h^t_j) & \text{if } \tau \geq t \\
    m^t_j & \text{otherwise}
\end{cases}
\]

where \(m^t_j\) is the message (action) played by \(j\) at period \(\tau < t\) in the public history \(h^{t-1}_i\). Thus, \(\varphi^{h^t_i}_j\) transforms any strategy in \(\bar{S}_j\) into one that has the same continuation from \(h^t_i\), and that agrees with \(h^t_i\) for the previous periods. Define the mapping \(L_{h^t_i} : \Theta^* \times B_{-i} \times S_{-i} \to \Theta^* \times B \times S\) such that

\[
L_{h^t_i}((\theta, b_{-i}, s_{-i})) = \left(\theta, \hat{b}_i, b_{-i}, \varphi^{h^t_i}_i\left(\hat{\sigma}_i\left(b_i\right)\right), \varphi^{h^t_i}_{-i}(s_{-i})\right).
\]

(In particular, \(L_{\phi}(\theta, b_{-i}, s_{-i}) = \left(\theta, \hat{b}_i, b_{-i}, \hat{\sigma}_i\left(b_i\right), s_{-i}\right)\).

Define the CPS \(\lambda_i \in \Delta^{\mathcal{H}_i}(\Theta^* \times B \times S)\) such that, for any measurable \(E \subseteq \Theta^* \times B \times S\),

\[
\lambda_i(\phi)[E] = P_i^{(\hat{\sigma}, \hat{\rho})}\left(\hat{b}_i\right)\left[L^{-1}_{\phi}(E)\right]
\]

and for all \(h^t_i \in \mathcal{H}_i\) s.t. \(\lambda_i(h_i^{t-1})[h^t_i] = 0\), let

\[
\lambda_i(h^t_i)[E] = P_i^{(\hat{\sigma}, \hat{\rho})}\left(\hat{b}_i\right)\left[L^{-1}_{h^t_i}(E)\right].
\]
(conditional beliefs \( \lambda_i (h^t_i) \) at histories \( h^t_i \) s.t. \( \lambda_i (h^{t-1}_i) [h^t_i] > 0 \) are determined via Bayesian updating, from the definition of CPS, appendix B.1.2)

Define the CPS \( \mu^i \in \Delta^H \Theta^* \times S \) s.t. \( \forall h^t_i \in \mathcal{H}_i, \mu^i (h^t_i) = \text{marg}_{\Theta^* \times S} \lambda_i (h^t_i) \). By construction, \( \hat{s}_i \in r_i (\mu^i) \). We only need to show that conditions (2) and (3) in the definition of \( \mathcal{BR} \) are satisfied by \( \mu^i \). This part proceeds by induction: The initial step, for \( k = 1 \), is trivial. Hence, \( S_j \subseteq \mathcal{BR}^1_j \) for every \( j \). To complete the proof, let (as inductive hypothesis) \( \bar{S}_j \subseteq \mathcal{BR}^k_j \) for every \( j \). Then \( \mu^i \) constructed above satisfies \( \mu^i (\phi) \subseteq \Theta^* \times \{ \hat{s}_i \} \times \mathcal{BR}^k \) and

\[
\text{supp} (\text{marg}_{S|h^{t-1}} \mu^i (\phi)) = \text{supp} (\text{marg}_{S|h^{t-1}} \mu^i (h^t_i)) \subseteq S|h^{t-1}.
\]

thus \( \hat{s}_i \in \mathcal{BR}^{k+1}_i \). This concludes the first part of the proof.

**Step 2:** (\( \Rightarrow \)). Let \( \mathcal{B} \) be such that for each \( i, B_i = \mathcal{BR}_i \) and let strategy \( \hat{\sigma}_i : B_i \rightarrow S_i \) be the identity map. Define the map \( M_{i,\phi} : \Theta^* \times S \rightarrow \Theta^* \times B_{-i} \) s.t.

\[
M_{i,\phi} (\theta, s_i, s_{-i}) = (\theta, \tilde{\sigma}^{-1}_{-i} (s_{-i}))
\]

Notice that, for each \( i \) and \( s_i \in \mathcal{BR}_i, \exists \mu^{s_i} \in \Delta^H (\Theta^* \times S) \) s.t.

1. \( \hat{s}_i \in r_i (\mu^{s_i}) \)
2. for all \( h^t_i \in \mathcal{H}_i: s_j \in \text{supp} (\text{marg}_{S_j} \mu^{s_i} (h^t_i)) \Rightarrow \exists s'_j \in \mathcal{BR}_j : s_j|h^{t-1} = s'_j|h^{t-1}.
\]

Hence, for each \( h^t_i \neq \phi \), we can define the map \( \rho_{\hat{s}_i, h^t_i} : \text{supp} (\text{marg}_{S_{-i}} \mu^{s_i} (h^t_i)) \rightarrow \mathcal{BR}_{-i} \) that satisfies \( \rho_{\hat{s}_i, h^t_i} (s_{-i}) |h^{t-1} = s_{-i}|h^{t-1} \). Let \( m_{\hat{s}_i, h^t_i} \equiv \tilde{\sigma}^{-1}_{-i} \circ \rho_{\hat{s}_i, h^t_i} \). Define maps

\[ M_{\hat{s}_i, h^t_i} : \Theta^* \times \text{supp} (\mu^{s_i} (h^t_i)) \rightarrow \Theta^* \times B_{-i} \]

\[ M_{\hat{s}_i, h^t_i} (\theta, s_i, s_{-i}) = (\theta, m_{h^t_i} (s_{-i})). \]
Let beliefs $\beta_i : B_i \rightarrow \Delta(\Theta^* \times B_{-i})$ be s.t. for every measurable $E \subseteq \Theta^* \times B_{-i}$

$$\beta_i(b_i)[E] = \mu^{\hat{\sigma}(b_i)}(\phi) \left[ M_{i,\phi}(E) \right]$$

Let beliefs $\hat{p}_i$ be derived from $\hat{\sigma}$ and the initial beliefs via Bayesian updating whenever possible. At all other histories $h'_i \in \mathcal{H}_i$, for every measurable $E \subseteq \Theta^* \times B$, set

$$\hat{p}_i(h'_i)[E] = \mu^{\hat{\sigma}(b_i)}(h'_i) \left[ M_{i,\hat{\sigma}(b_i),h'_i}(E) \right].$$

By construction, $(\hat{\sigma}, \hat{p})$ is an IPE of $(\mathcal{E}, \mathcal{M}, \mathcal{B})$. ■

### B.2.2 The backwards procedure

Fix a public history of length $T - 1, h^{T-1}$. For each $k = 0, 1, \ldots$, let $\mathcal{R}^k_{i,h^{T-1}} \subseteq S^h_{i}$ be such that $(y^T_i, s^h_i) \in \mathcal{R}^k_{i,h^{T-1}}$ if and only if $s^h_i \in \mathcal{R}^k_{i,h^{T-1}}(y^T_i)$, $\mathcal{R}^k_{i,h^{T-1}} = \times_{i \in N} \mathcal{R}^k_{i,h^{T-1}}$ and $\mathcal{R}^k_{i,h^{T-1}} = \times_{j \neq i} \mathcal{R}^k_{j,h^{T-1}}$. For each $h^T_i = (h^{T-1}, y^T_i) \in Y_i^T$, let $\mathcal{R}^0_{i,h^{T-1}}(y^T_i) = S^h_{i}$ and for $k = 1, 2, \ldots$, for each $y^T_i \in Y_i^T$ let

$$\mathcal{R}^k_{i,h^{T-1}}(y^T_i) = \left\{ s_i \in \mathcal{R}^{k-1,h^{T-1}}_{i}(y^T_i) : \exists \pi^{h^{T-1}} \in \Delta(\Theta^* \times S_{-i}^{h^{T-1}}) \right\}$$

1. $\pi^{h^{T-1}}(\{y^T_i\} \times \Theta^*_{-i} \times \mathcal{R}^{k-1,h^{T-1}}_{-i}) = 1$

2. for all $s' \in S^{h^{T-1}}_{i}$:

$$\int_{(\theta, s_{-i}) \in \Theta^* \times S_{-i}^{h^{T-1}}} U_i(s_i, s_{-i}, \theta; h^{T-1}) \cdot d\pi^{h^{T-1}} \geq \int_{(\theta, s_{-i}) \in \Theta^* \times S_{-i}^{h^{T-1}}} U_i(s'_i, s_{-i}, \theta; h^{T-1}) \cdot d\pi^{h^{T-1}} \}$$

and $\mathcal{R}^T_{i,h^{T-1}}(\bar{y}^T_i) = \{ \bigcap_{k=1}^{\infty} \mathcal{R}^k_{i,h^{T-1}}(\bar{y}^T_i) \}$.

Notice that $\mathcal{R}^T_{i,h^{T-1}}$ consists of pairs of types $y^T_i$ and continuation strategies $s_i \in S^{(h^{T-1}, y^T_i)}_{i}$. Hence, each $\mathcal{R}^T_{i,h^{T-1}}$ can be seen as a subset of $S^{h^{T-1}}_{i}$.
For each $t = 1,\ldots, T - 1$, for each $h^t_i = (h^{t-1}_i, y^t_i)$ let:
\[
\mathcal{R}^{0,h^{t-1}}_t (y^t_i) = \left\{ s_i \in S^h_i : \forall h^t \text{ s.t. } h^{t-1} \prec h^t, \forall y^{t+1}_i \text{ s.t. } y^t_i \prec y^{t+1}_i, s_i| (h^t, y^{t+1}_i) \in \mathcal{R}^h_t (y^{t+1}_i) \right\}
\]
and for each $k$, \((y^t_i, s^h_i) \in \mathcal{R}^{k,h^{t-1}}_i \) if and only if \( s^h_i \in \mathcal{R}^{k,h^{t-1}}_i (y^t_i) \). For each $k = 1, 2, \ldots$ and for each $k = 1, 2, \ldots$
\[
\mathcal{R}^{k,h^{t-1}}_i (\tilde{y}^t_i) = \left\{ s_i \in \mathcal{R}^{k-1,h^{t-1}}_i (\tilde{y}^t_i) : \exists \pi^{h_i} \in \Delta (\Theta \times S^{h^{t-1}}) \right\}
\]
1. \( \pi (\{ \tilde{y}^t_i \} \times (\times_{t+1}^T \Theta)) \times \Theta^* \times \mathcal{R}^{k-1,h^{t-1}}_i = 1 \)
2. for all $s' \in S^h_i$,
\[
\int_{(\theta, s-i) \in \Theta \times S_{i-1}^{h^{t-1}}} U_i (s_i, s-i, \theta; h^{t-1}) \cdot d\pi^{h_i}
\]
\[
\geq \int_{(\theta, s-i) \in \Theta \times S_{i-1}^{h^{t-1}}} U_i (s'_i, s-i, \theta; h^{t-1}) \cdot d\pi^{h_i}
\]
and \( \mathcal{R}^{h^{t-1}}_i (\tilde{y}^t_i) = \bigcap_{k=1}^{\infty} \mathcal{R}^{k,h^{t-1}}_i (\tilde{y}^t_i) \).

Finally: \( \mathcal{R}_i^\phi = \left\{ s_i \in S_i : s_i| y^1_i \in \mathcal{R}_i^\phi (y^1_i) \text{ for each } y^1_i \in Y^1_i \right\} \).

**Proposition 4.2.** \( B\mathcal{R}_i = \mathcal{R}_i^\phi \) for each $i$.

**Proof:**

**Step 1** \( (\mathcal{R}_i^\phi \subseteq B\mathcal{R}_i:) \): let \( \hat{s}_i \in \mathcal{R}_i^\phi \). Then, for each $h^t_i = (h^{t-1}_i, y^t_i)$, $s_i|h^t_i \in \mathcal{R}_i^{h^{t-1}} (y^t_i)$ (equivalently: $s_i^{h^{t-1}} \in \mathcal{R}_i^{h^{t-1}}$). Notice that for each $h^{t-1}$ and $s_i^{h^{t-1}} \in \mathcal{R}_i^{h^{t-1}}$, there exists $s_i \in \mathcal{R}_i^\phi$ such that $s_i|h^{t-1} = s_i^{h^{t-1}}$. Thus, for each $j$ and $h^{t-1}$, we can define measurable functions $\rho_j^{h^{t-1}} : \mathcal{R}_j^{h^{t-1}} \rightarrow \mathcal{R}_j^\phi$ such that: $\forall s_j^{h^{t-1}} \in \mathcal{R}_j^{h^{t-1}}$
\[
\rho_j^{h^{t-1}} \left( s_j^{h^{t-1}} \right) | h^{t-1} = s_j^{h^{t-1}}.
\]
(Functions $\rho_j^{h^{t-1}}$ assign to strategies in $\mathcal{R}_j^{h^{t-1}}$, strategies in $\mathcal{R}_j^\phi$ with the same continuation from $h^{t-1}$.) As usual, denote by $\rho_{i=1}^{h^{t-1}}$ the product $\times_{j \neq i} \rho_j^{h^{t-1}}$. 181
For each $h_{l-1}$, let $\varphi_{j}^{h_{l-1}} : S_{j} \rightarrow S_{j} (h_{l-1})$ be a measurable function such that

$$
\varphi_{j}^{h_{l-1}} (s_{j}) (h_{j}^{\tau}) = \begin{cases} s_{j} (h_{j}^{\tau}) & \text{if } \tau > t \\
m_{j} & \text{otherwise}
\end{cases}
$$

where $m_{j}^{\tau}$ is the message (action) played by $j$ at period $\tau < t$ in the public history $h_{l-1}$. (As usual, denote by $\varphi_{-i}^{h_{l-1}}$ the product $\times_{j \neq i} \varphi_{j}^{h_{l-1}}$.)

For each $h_{l-1}$, define the measurable mapping $\varrho_{-i}^{h_{l-1}} : \mathcal{R}_{-i}^{h_{l-1}} \rightarrow S_{-i} (h_{l-1})$ such that

$$
\forall s_{-i}^{h_{l-1}} \in \mathcal{R}_{-i}^{h_{l-1}}, \\
\varrho_{-i}^{h_{l-1}} (s_{-i}^{h_{l-1}}) = \varphi_{-i}^{h_{l-1}} \circ \rho_{-i}^{h_{l-1}} (s_{-i}^{h_{l-1}}).
$$

It will be shown that: for each $k = 0, 1, ..., \hat{s}_{i} \in \mathcal{R}_{i}^{\phi,k}$ implies $\hat{s}_{i} \in B \mathcal{R}_{i}^{k}$.

The initial step is trivially satisfied ($B \mathcal{R}_{i}^{0} = S_{i} = \mathcal{R}_{i}^{\phi,0}$).

For the inductive step, suppose that the statement is true for $n = 0, ..., k-1$: Since $\hat{s}_{i} \in \mathcal{R}_{i}^{\phi,k}$, for each $h_{i}^{l} = (h_{l-1}^{i}, y_{i}^{l})$ there exists $\pi_{i}^{h_{i}^{l}} \in \Delta \left( \Theta^{*} \times S_{-i}^{h_{l-1}} \right)$ that satisfy

$$
\hat{s}_{i} | h_{i}^{l} \in \arg \max_{s_{i}^{l} \in S_{i}^{h_{i}^{l}}} \int_{\Theta^{*} \times S_{-i}^{h_{l-1}}} U_{i} (s_{i}^{l}, s_{-i}, \theta; h_{l-1}^{i}) \cdot d\pi_{i}^{h_{i}^{l}},
$$

and such that $\pi^{\phi} \left( \Theta^{*} \times \mathcal{R}_{-i}^{\phi,k-1} \right) = 1$ and for all $h_{i}^{l} \neq \phi$,

$$
\pi_{i}^{h_{i}^{l}} \left( \{ y_{i}^{l} \} \times (\times_{\tau=t+1}^{T} \Theta_{i,\tau}) \times \Theta_{-i}^{*} \times \mathcal{R}_{-i}^{h_{l-1}} \right) = 1.
$$

Now, consider the CPS $\mu^{i} \in \Delta^{h_{i}} (\Theta^{*} \times S)$ such that, for all measurable $E \subseteq \Theta^{*} \times S_{-i}$,

$$
\mu^{i} (\phi) [\{ \hat{s}_{i} \} \times E] = \pi^{\phi} (E).
$$

By definition of CPS, $\mu^{i} (\phi)$ defines $\mu (h_{i}^{l})$ for all $h_{i}^{l}$ s.t. $\mu^{i} (\phi) [h_{i}^{l}] > 0$. Let $h_{i}^{l}$ be such that $\mu^{i} (\phi) [h_{i}^{l-1}] > 0$ and $\mu^{i} (\phi) [h_{i}^{l}] = 0$. Define the measurable mapping $M_{h_{i}^{l}} : \Theta^{*} \times \mathcal{R}_{-i}^{h_{l-1}} \rightarrow \Theta^{*} \times S (h_{l-1})$ such that for all $\left( \theta, s_{-i}^{h_{l-1}} \right) \in \Theta^{*} \times S (h_{l-1})$,

$$
M_{h_{i}^{l}} (\theta, s_{-i}^{h_{l-1}}) = \left( \theta, \varphi_{i}^{h_{l-1}} (\hat{s}_{i}) ; \varrho_{-i}^{h_{l-1}} (s_{-i}^{h_{l-1}}) \right).
$$

182
and set \( \mu^i(h^t_i) \) equal to the pushforward of \( \pi^{h^t_i} \) under \( M_{h^t_i} \), i.e. such that for every measurable \( E \subseteq \Theta^* \times S \)

\[
\mu^i(h^t_i)[E] = \pi^{h^t_i} \left[ M_{h^t_i}^{-1}(E) \right].
\]

Again, by definition of CPS, \( \mu^i(h^t_i) \) defines \( \mu(h^t_i) \) for all \( h^t_i \succ h^t_i \) that receive positive probability under \( \mu^i(h^t_i) \). For other histories, proceeds as above, setting \( \mu^i(h^t_i) \) equal to the pushforward of \( \pi^{h^t_i} \) under \( M_{h^t_i} \), and so on.

By construction, \( \hat{s}_i \in r_i(\mu^i) \) (condition 1 in the definition of \( \mathcal{BR}^k_i \)). Since by construction \( \mu^i \left( \Theta^* \times \{ \hat{s}_i \} \times \mathcal{R}^{k-1}_{-i} ; \phi \right) = 1 \), under the inductive hypothesis

\[
\mu^i \left( \Theta^* \times \{ \hat{s}_i \} \times \mathcal{BR}^{k-1}_{-i} ; \phi \right) = 1
\]

(condition 2 in the definition of \( \mathcal{BR}^k_i \)). From the definition of \( \varphi^{h^{t-1}_i}(\hat{s}_i) \), CPS \( \mu^i \) satisfies condition (3.1) at each \( h^t_i \). From the definition of \( \varphi^{h^{t-1}_i} \), under the inductive hypothesis, \( \mu^i \) satisfies condition (3.2).

**Step 2 (\( \mathcal{BR}_i \subseteq \mathcal{R}_i^\phi \)):** let \( \hat{s}_i \in \mathcal{R}_i^\phi \) and \( \mu^i \in \Delta \mathcal{H}_i \left( \Theta^* \times S \right) \) be such that \( \hat{s}_i \in r_i(\mu^i) \). For each \( h^t_i = (h^{t-1}_i, y^t_i) \), define the mapping \( \psi_{h^t_i} : S_{-i} \rightarrow S^{h^{t-1}_{-i}}_{-i} \) s.t. \( \psi_{h^t_i}(s_{-i})|h^{t-1}_i = s_{-i}|h^{t-1}_i \). (Function \( \psi_{h^t_i} \) transforms each strategy profile of the opponents into its continuation from \( h^{t-1}_i \).) Define also \( \Psi_{h^t_i} : \Theta^* \times S \rightarrow \Theta^* \times S^{h^{t-1}_i}_{-i} \) such that

\[
\Psi_{h^t_i}(\theta, s_i, s_{-i}) = (\theta, \psi_{h^t_i}(s_{-i}))
\]

For each \( h^t_i \in \mathcal{H}_i \), let \( \pi^{h^t_i} \in \Delta \left( \Theta^* \times S^{h^{t-1}_{-i}}_{-i} \right) \) be such that for every measurable \( E \subseteq \Theta^* \times S^{h^{t-1}_{-i}}_{-i} \)

\[
\pi^{h^t_i}[E] = \mu^i(h^t_i) \left[ \Psi_{h^t_i}^{-1}(E) \right].
\]

so that the implied joint distribution over payoff states and continuation strategies \( s_{-i}|h^{t-1}_i \) is the same under \( \mu^i(\cdot; h^t_i) \) and \( \pi^{h^t_i} \). We will show that \( \hat{s}_i|h^t_i \in \mathcal{R}^{h^{t-1}_i}(y^t_i) \) for
each \( h^t_i = (h^{t-1}, y^t_i) \). Notice that, by construction,

\[
\hat{s}_i|h^t_i \in \arg \max_{s_i \in S^t_i} \int U_i(s_i, s_{-i}; h^t_i) \cdot d\pi^{h^t_i}.
\]

The argument proceeds by induction on the length of histories.

**Initial Step** \((T - 1)\). Fix history \( h^T_i = (h^{T-1}, y^T_i) \): for each \( k \), if \( \hat{s}_i \in \mathcal{BR}^k_i \), then \( \hat{s}_i|h^T_i \in \mathcal{R}^{h^{T-1},k}_i \). For \( k = 0 \), it is trivial. For the inductive step, let \( \pi^{h^T} \) be defined as above: under the inductive hypothesis, \( \pi^{h^T}_i \left( \Theta^*_i \times \mathcal{R}^{h^{T-1},k-1}_i \right) = 1 \) (condition 1), while \( \hat{s}_i \in r_i(\mu^t) \) implies that condition (2) is satisfied.

**Inductive Step:** suppose that for each \( \tau = t + 1, ..., T \), \( \hat{s}_i \in \mathcal{BR}_i \), implies \( \hat{s}_i|h^T_i \in \mathcal{R}^{h^T-1}_i(y^\tau) \) for each \( h^\tau_i = (h^{T-1}, y^\tau_i) \). We will show that for each \( k \), \( h^t_i = (h^{t-1}, y^t_i) \), \( \hat{s}_i|h^t_i \in \mathcal{R}^{k,h^t-1}_i(y^t) \). We proceed by induction on \( k \): under the inductive hypothesis on \( \tau \), \( \hat{s}_i|h^t_i \in \mathcal{R}^{0,h^t-1}_i(y^t) \). For the inductive step on \( k \), suppose that \( \hat{s}_i \in \mathcal{BR}_i \), implies \( \hat{s}_i|h^t_i \in \mathcal{R}^{n,h^t-1}_i(y^t) \) for \( n = 0, ..., k - 1 \), and suppose (as contra-positive) that \( \hat{s}_i|h^t_i \notin \mathcal{R}^{k,h^t-1}_i(y^t) \). Then, for \( \pi^{h^t_i} \) defined as above, it must be that \( \text{supp}(\pi^{h^t_i}) \notin \Theta^*_i \times \mathcal{R}^{k-1,h^t-1}_i \), which, under the inductive hypothesis on \( n \), implies that \( \exists s_{-i} \in \text{supp}(\text{marg}_{s_{-i}}(\mu^t_i(h^t_i))) \) s.t. \( \exists s'_{-i} \in \mathcal{BR}_{-i} : s'_{-i}|h^{t-1} = s_{-i}|h^{t-1} \), which contradicts that \( \mu^t \) justifies \( \hat{s}_i \) in \( \mathcal{BR}_i \). ■

**B.3 Proofs of results from Sections 4.7 and 4.8**

**B.3.1 Proof of Proposition 4.3**

**Step 1** (If): For the if part, fix an arbitrary type space \( \mathcal{B} \), and consider a direct mechanism \( \mathcal{M} \). Let \((p^i)_{i \in \mathcal{N}}\) be any beliefs system such that, \( \forall i \in \mathcal{N}, \forall (\theta, b_{-i}) \in \Theta^* \times \mathcal{B}_{-i}, p^i(h^0_i) = \beta_i(h^0_i) \) and for each \( h^t_i \in \mathcal{H}_i \) such that \( P^{\pi^* \mathcal{M}}(h^t_i - 1) | h^t_i > 0 \), \( p^i(h^t_i) \) is obtained via Bayesian updating. If instead \( h^t_i \) is such that \( P^{\pi^* \mathcal{M}}(h^t_i - 1) | h^t_i = 0 \), and
That is, at unexpected histories, each $i$ believes that the opponents have not deviated from the truth-telling strategy: If "unexpected reports" were observed, player $i$ rather revises his beliefs about the opponents’ types, not their behavior.

Notice that if $U_i(s^*, \theta) \geq U_i(s'_i, s^*_{-i}, \theta)$ for all $\theta$, then for any $p^i(\phi) \in \Delta(\Theta^* \times B_{-i})$, 

$$
\int_{\Theta^* \times B_{-i}} u_i \left( g^{s^*}(\theta), \theta \right) \cdot dp^i(\phi) \\
\geq \int_{\Theta^* \times B_{-i}} u_i \left( g^{(s'_i, s^*_{-i})|h^i_t}(\theta), \theta \right) \cdot dp^i(\phi) .
$$

Hence, the incentive compatibility constraints are satisfied at the beginning of the game, and so at all histories reached with positive probability according to the initial conjectures and strategy profile. Being $\sigma^* \in \Sigma^*$, only truthful histories receive positive probability. At zero probability histories, we maintain that the belief system satisfies (B.1). With these beliefs, the only payoff-relevant component of the opponents’ strategies at history $h^i_t$ is the “truthful report”: from the point of view of player $i$, what $\sigma^*_{-i}$ specifies at non-truthful histories is irrelevant. Let $\sigma^*_i(h^i_t)$ be a best response to such beliefs and $\sigma^*_{-i}$ in the continuation game: Notice that under these beliefs, any $\sigma_{-i} \in \Sigma^*$ determines the same $\sigma^*_i(h^i_t)$. Hence, for any $i$ we can chose $\sigma^*_i \in \Sigma^*$ so that the strategy profile thus constructed is an IPE of the Bayesian game.

**Step 2 (only if):** Since perfect implementability implies interim implementability, the “only if” immediately follows the results by Bergemann and Morris (2005), who showed that if a SCF is interim implementable on all type spaces, then it is ex-post implementable. ■
B.3.2 Proof of Proposition 4.5

By contradiction, suppose $\mathcal{BR} = B \neq \{s^c\}$. By continuity of $u_i$ and compactness of $\Theta^*$, $B(h^t)$ is compact for each $h^t$. (Because if $B = \mathcal{BR}$, strategies in $B$ must be best responses to conjectures concentrated on $B$, see definition of $\mathcal{BR}$).

It will be shown that for each $t$ and for each public history $h^{t-1}$, $s[B(h^{t-1})] = s^c[h^{t-1}]$, contradicting the absurd hypothesis. The proof proceeds by induction on the length of the history, proceeding backwards from public histories $h^T$ to the empty history $h^0$.

**Initial Step:** $[s[B(h^{T-1})] = s^c[h^{T-1}]$ for each $h^{T-1}$.

Suppose, by contradiction, that $\exists h^{T-1} = (y^{T-1}, x^{T-1}) : s[B(h^{T-1})] \neq s^c(h^{T-1})$.

Then, by the contraction property,

$$\exists y^T_i \text{ and } \theta^t_i \in B_i(h^{T-1}, y^T_i) : \theta^t_i \neq s^c_i(h^{T-1}, y^T_i) \text{ such that:}$$

$$\text{sign} \left[ s^c_i(h^{T-1}, y^T_i) - \theta^t_i \right] = \text{sign} \left[ \alpha^T_i(y^T_i, y^-_i) - \alpha^T_i(\bar{y}^{T-1}, \theta^t_i, \theta^t_{-i,t}) \right]$$

for all $y^-_i = (y^{T-1}, \theta_{-i,t})$ and $\theta^t_{-i,t} \in B_{-i}(h^{T-1}, y^-_i)$.

Fix such $y^T_i$ and $\theta^t_i \neq s^c_i(h^{T-1}, y^T_i)$, and suppose (w.l.o.g.) that $s^c_i(h^{T-1}, y^T_i) > 0$. Define:

$$\delta(h^{T-1}, y^T_i) := \min_{y^T_i, \theta^t_i \in Y^T_i \text{ and } \theta^t_{-i,t} \in B_{-i}(h^{T-1}, y^-_i)} \left[ \alpha^T_i(y^T_i, y^-_i) - \alpha^T_i(\bar{y}^{T-1}, \theta^t_i, \theta^t_{-i,t}) \right]$$

(by compactness of $Y^T$ and $B(h^{T})$, $\delta(h^{T}, y^T_i)$ is well-defined). Also, from $\theta^t_i \neq s^c_i(h^{T-1}, y^T_i)$ and the Contraction Property, $\delta(h^{T-1}, y^T_i) > 0$.

For any $\varepsilon > 0$, let

$$\psi(h^{T-1}, y^T_i, \theta^t_i, \varepsilon) = \max_{\theta_{-i,t} \in \Theta_{-i,t}} \left\{ \alpha^T_i(\bar{y}^{T-1}, \theta^t_i + \varepsilon, \theta_{-i,t}) - \alpha^T_i(\bar{y}^{T-1}, \theta^t_i, \theta_{-i,t}) \right\}$$

(B.3)
Thus, by continuity, there exists as \(0\) strictly increasing in \(\theta_{i,T}, \psi (h_{i,T}^{T-1}, y_{i}^{T-1}, \theta', \theta_{i,T}, \varepsilon)\) is increasing in \(\varepsilon\) and \(\psi (h_{i,T}^{T-1}, y_{i}^{T-1}, \theta', \theta_{i,T}, \varepsilon) \rightarrow 0\) as \(\varepsilon \rightarrow 0\).

Let \((f_{t}(y_{i}^{T}))_{t=1}^{T-1} = x^{T-1}\). From strict EPIC, we have that for each \(\varepsilon\),

\[
v_{i} (x^{T-1}, f_{T} (y_{T}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}), \alpha^{T} (y_{T}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}), \alpha^{-T} (y_{T}^{T-1})) > v_{i} (x^{T-1}, f_{T} (y_{T}^{T-1}, \theta'_{i,T}, \theta_{-i,T}), \alpha^{T} (y_{T}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}), \alpha^{-T} (y_{T}^{T-1}))
\]

and

\[
v_{i} (x^{T-1}, f_{T} (y_{T}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}), \alpha^{T} (y_{T}^{T-1}, \theta'_{i,T}, \theta_{-i,T}), \alpha^{-T} (y_{T}^{T-1})) < v_{i} (x^{T-1}, f_{T} (y_{T}^{T-1}, \theta'_{i,T}, \theta_{-i,T}), \alpha^{T} (y_{T}^{T-1}, \theta'_{i,T}, \theta_{-i,T}), \alpha^{-T} (y_{T}^{T-1}))
\]

Thus, by continuity, there exists \(a^{T} (\varepsilon)\) such that

\[
\alpha^{T} (y_{T}^{T-1}, \theta'_{i,T}, \theta_{-i,T}) < a^{T} (\varepsilon) < \alpha^{T} (y_{T}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}) \quad \text{ (B.4)}
\]

such that

\[
v_{i} (x^{T-1}, f_{T} (y_{T}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}), a^{T} (\varepsilon), \alpha^{-T} (y_{T}^{T-1})) = v_{i} (x^{T-1}, f_{T} (y_{T}^{T-1}, \theta'_{i,T}, \theta_{-i,T}), a^{T} (\varepsilon), \alpha^{-T} (y_{T}^{T-1}))
\]

From the “within-period SCC” (def. 4.16),

\[
v_{i} (x^{T-1}, f_{T} (y_{T}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}), a^{*}, \alpha^{-T} (y_{T}^{T-1})) > v_{i} (x^{T-1}, f_{T} (y_{T}^{T-1}, \theta'_{i,T}, \theta_{-i,T}), a^{*}, \alpha^{-T} (y_{T}^{T-1}))
\]

whenever \(a^{*} > a^{T} (\varepsilon)\).

Thus, to reach the contradiction, it suffices to show that for any \(y_{i}^{T-1} \in Y_{-i}, \alpha^{T} (y_{i}^{T-1}, y_{-i}^{T-1}) \geq a^{T} (\varepsilon)\): If this is the case, reporting \(\theta'_{i,T}\) is (conditionally) strictly dominated by reporting \(\theta'_{i,T} + \varepsilon\) at \(h_{i}^{T} = (h_{i}^{T-1}, y_{i}^{T-1})\), hence it cannot be that \(B_{i} = B_{R_{i}}\) and \(\theta'_{i,T} \in \)


$B_i \left( h^{T-1}, y_i^T \right)$. To this end, it suffices to choose $\varepsilon$ sufficiently small that

$$
\psi \left( h^{T-1}, y_i^T, \theta_{i,T}', \varepsilon \right) < \delta
$$

(B.5)

and operate the substitutions as follows

$$
\alpha_i^T \left( y_i^T, y_{i-1}^T \right) \geq \alpha_i^T \left( \tilde{y}^{T-1}, \theta_{i,T}', \theta'_{i-1,T} \right) + \delta \left( h^{T-1}, y_i^T \right)
$$

$$
\geq \alpha_i^T \left( \tilde{y}^{T-1}, \theta_{i,T}' + \varepsilon, \theta'_{i-1,T} \right) + \delta \left( h^{T-1}, y_i^T \right) - \psi \left( h^{T-1}, y_i^T, \theta_{i,T}', \varepsilon \right)
$$

$$
> \alpha_i^T \left( \tilde{y}^{T-1}, \theta_{i,T}' + \varepsilon, \theta'_{i-1,T} \right)
$$

$$
> a^T \left( \varepsilon \right)
$$

Thus: $\alpha_i^T \left( y_i^T, y_{i-1}^T \right) > a^T \left( \varepsilon \right)$ for any $y_{i-1}^T$. This concludes the initial step.

**Inductive Step:** [for $t = 1, \ldots, T - 1$: if $s \left[ B \left( h^\tau \right) \right] = s^c \left[ h^\tau \right]$ for all $h^\tau$ and all $\tau > t$ then $s \left[ B \left( h^T \right) \right] = s^c \left[ h^T \right]$ for all $h^T$]

Suppose, by contradiction, that $\exists h^{t-1} = (\tilde{y}^{t-1}, x^{t-1}) : s \left[ B \left( h^{t-1} \right) \right] \neq s^c \left( h^{t-1} \right)$.

Then, by the contraction property,

$$
\exists y_i^t \text{ and } \theta_{i,t}' \in B_i \left( h^{t-1}, y_i^t \right) : \theta_{i,t}' \neq s^c_i \left( h^{t-1}, y_i^t \right) \text{ such that:}
$$

$$
\text{sign} \left[ s^c_i \left( h^{t-1}, y_i^t \right) - \theta_{i,t}' \right] = \text{sign} \left[ \alpha_i^t \left( y_i^t, y_{i-1}^t \right) - \alpha_i^t \left( \tilde{y}^{t-1}, \theta_{i,t}', \theta'_{i-1,t} \right) \right]
$$

$$
\text{for all } y_{i-1}^t = (y_{i-1}^{t-1}, \theta_{i-1,t}) \text{ and } \theta_{i-1,t}' \in B_{i-1} \left( h^{t-1}, y_{i-1}^t \right).
$$

Fix such $y_i^t$ and $\theta_{i,t}' \neq s^c_i \left( h^{t-1}, y_i^t \right)$, and suppose (w.l.o.g.) that $s^c_i \left( h^{t-1}, y_i^t \right) > \theta_{i,t}'$.

Similar to the initial step, it will be shown that there exists $\theta_{i,t}^e = \theta_{i,t}' + \varepsilon$ for some $\varepsilon > 0$ such that for any conjecture consistent with $B_{i-1}$, playing $\theta_{i,t}^e$ is strictly better than playing $\theta_{i,t}'$ at history $(h^{t-1}, y_i^t)$, contradicting the hypothesis that $\mathcal{BR} = B$.

For any $\varepsilon > 0$, set $\theta_{i,t}^e = \theta_{i,t}' + \varepsilon$; for each realization of signals $\tilde{\theta}_i = \left( \tilde{\theta}_{i,k} \right)_{k=1}^T$ and opponents’ reports $\tilde{m}_{i-1} = (\tilde{m}_{i-1,k})_{k=1}^T$, for each $\tau > t$, denote by $s_{i,\tau}^e \left( \theta_{i,t}^e, \tilde{m}_{i-1}, \tilde{\theta}_i \right)$ the action taken at period $\tau$ if $\theta_{i,t}^e$ is played at $t$, $s_{i}^c$ is followed in the following stages,
and the realized payoff type and opponents’ messages are \( \tilde{\theta}_i \) and \( \tilde{m}_{-i} \), respectively. 
(By continuity of the aggregators functions, \( s_{i,T}^c \left( \theta_{i,t}^{c}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \) is continuous in \( \varepsilon \), and converges to \( s_{i,T}^c \left( \theta_{i,t}^{c}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \) as \( \varepsilon \to 0 \).

For each realization \( \tilde{\theta}_i = \left( \tilde{\theta}_{i,k} \right)_{k=1}^T \) and reports \( \tilde{m}_{-i} = (\tilde{m}_{-i,k})_{k=t}^T \) and for each \( \tau > t \), \( s_{i,T}^c \left( \theta_{i,t}^{c}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \) may be one of five cases:

1. \( s_{i,T}^c \left( \theta_{i,t}^{c}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \in (\theta_{i,T}^{-}, \theta_{i,T}^{+}) \), then
   \[
   \alpha_t^\tau \left( y_t^\tau, y_{-i}^\tau \right) = \alpha_t^\tau \left( \tilde{y}_t^{\tau-1}, \theta_{i,t}, \left( s_{i,k}^c \left( \theta_{i,t}^{c}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^T, y_{-i}^\tau \right)
   \]
   for all \( y_{-i}^\tau \), and we can choose \( \varepsilon \) sufficiently small that \( s_{i,T}^c \left( \theta_{i,t}^{c}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \in (\theta_{i,T}^{-}, \theta_{i,T}^{+}) \), i.e.
   \[
   \alpha_t^\tau \left( y_t^\tau, y_{-i}^\tau \right) = \alpha_t^\tau \left( \tilde{y}_t^{\tau-1}, \theta_{i,t}, \left( s_{i,k}^c \left( \theta_{i,t}^{c}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^T, y_{-i}^\tau \right)
   \]
   for all \( y_{-i}^\tau \).

2. \( s_{i,T}^c \left( \theta_{i,t}^{c}, \tilde{m}_{-i}, \tilde{\theta}_i \right) = \theta_{i,T}^{+} \) and
   \[
   \alpha_t^\tau \left( y_t^\tau, y_{-i}^\tau \right) > \alpha_t^\tau \left( \tilde{y}_t^{\tau-1}, \theta_{i,t}, \left( s_{i,k}^c \left( \theta_{i,t}^{c}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^T, y_{-i}^\tau \right)
   \]
   at the argmax over \( y_{-i}^\tau \), then we can choose \( \varepsilon \) sufficiently small that \( s_{i,T}^c \left( \theta_{i,t}^{c}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \)
   \( = \theta_{i,T}^{+} \) as well.

3. \( s_{i,T}^c \left( \theta_{i,t}^{c}, \tilde{m}_{-i}, \tilde{\theta}_i \right) = \theta_{i,T}^{+} \) and
   \[
   \alpha_t^\tau \left( y_t^\tau, y_{-i}^\tau \right) = \alpha_t^\tau \left( \tilde{y}_t^{\tau-1}, \theta_{i,t}, \left( s_{i,k}^c \left( \theta_{i,t}^{c}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^T, y_{-i}^\tau \right)
   \]
   for all \( y_{-i}^\tau \). Then, either \( s_{i,T}^c \left( \theta_{i,t}^{c}, \tilde{m}_{-i}, \tilde{\theta}_i \right) = \theta_{i,T}^{+} \) as well, or \( s_{i,T}^c \left( \theta_{i,t}^{c}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \in (\theta_{i,T}^{-}, \theta_{i,T}^{+}) \), i.e.
   \[
   \alpha_t^\tau \left( y_t^\tau, y_{-i}^\tau \right) = \alpha_t^\tau \left( \tilde{y}_t^{\tau-1}, \theta_{i,t}, \left( s_{i,k}^c \left( \theta_{i,t}^{c}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^T, y_{-i}^\tau \right)
   \]

189
for all $y_{i}^{-\tau}$. In either case,

$$
\alpha_{i}^{\tau} \left( \tilde{y}_{i}^{t-1}, \theta_{i,t}', \left( s_{i,k}^{c} \left( \theta_{i,t}', \tilde{m}_{-i}, \tilde{\theta}_{i} \right) \right)^{\tau}_{k=t+1}, y_{-i}^{\tau} \right) \\
= \alpha_{i}^{\tau} \left( \tilde{y}_{i}^{t-1}, \theta_{i,t}', \left( s_{i,k}^{c} \left( \theta_{i,t}', \tilde{m}_{-i}, \tilde{\theta}_{i} \right) \right)^{\tau}_{k=t+1}, y_{-i}^{\tau} \right)
$$

for all $y_{i}^{-\tau}$

4. $s_{i,\tau}^{c} \left( \theta_{i,t}', \tilde{m}_{-i}, \tilde{\theta}_{i} \right) = \theta_{i,T}^{\tau}$ and

$$
\alpha_{i}^{\tau} \left( y_{i}^{\tau}, y_{-i}^{\tau} \right) < \alpha_{i}^{\tau} \left( \tilde{y}_{i}^{t-1}, \theta_{i,t}', \left( s_{i,k}^{c} \left( \theta_{i,t}', \tilde{m}_{-i}, \tilde{\theta}_{i} \right) \right)^{\tau}_{k=t+1}, y_{-i}^{\tau} \right)
$$

at the argmax over $y_{-i}^{\tau}$. Then we can choose $\varepsilon$ sufficiently small that $s_{i,\tau}^{c} \left( \theta_{i,t}', \tilde{m}_{-i}, \tilde{\theta}_{i} \right) = \theta_{i,T}^{\tau}$ as well.

5. $s_{i,\tau}^{c} \left( \theta_{i,t}', \tilde{m}_{-i}, \tilde{\theta}_{i} \right) = \theta_{i,T}^{\tau}$ and

$$
\alpha_{i}^{\tau} \left( y_{i}^{\tau}, y_{-i}^{\tau} \right) = \alpha_{i}^{\tau} \left( \tilde{y}_{i}^{t-1}, \theta_{i,t}', \left( s_{i,k}^{c} \left( \theta_{i,t}', \tilde{m}_{-i}, \tilde{\theta}_{i} \right) \right)^{\tau}_{k=t+1}, y_{-i}^{\tau} \right)
$$

for all $y_{i}^{-\tau}$. Then, either $s_{i,\tau}^{c} \left( \theta_{i,t}', \tilde{m}_{-i}, \tilde{\theta}_{i} \right) = \theta_{i,T}^{\tau}$ as well, or $s_{i,\tau}^{c} \left( \theta_{i,t}', \tilde{m}_{-i}, \tilde{\theta}_{i} \right) \in (\theta_{i,T}^{\tau}, \theta_{i,T}^{\tau})$, i.e.

$$
\alpha_{i}^{\tau} \left( y_{i}^{\tau}, y_{-i}^{\tau} \right) = \alpha_{i}^{\tau} \left( \tilde{y}_{i}^{t-1}, \theta_{i,t}', \left( s_{i,k}^{c} \left( \theta_{i,t}', \tilde{m}_{-i}, \tilde{\theta}_{i} \right) \right)^{\tau}_{k=t+1}, y_{-i}^{\tau} \right)
$$

for all $y_{i}^{-\tau}$. In either case,

$$
\alpha_{i}^{\tau} \left( \tilde{y}_{i}^{t-1}, \theta_{i,t}', \left( s_{i,k}^{c} \left( \theta_{i,t}', \tilde{m}_{-i}, \tilde{\theta}_{i} \right) \right)^{\tau}_{k=t+1}, y_{-i}^{\tau} \right) \\
= \alpha_{i}^{\tau} \left( \tilde{y}_{i}^{t-1}, \theta_{i,t}', \left( s_{i,k}^{c} \left( \theta_{i,t}', \tilde{m}_{-i}, \tilde{\theta}_{i} \right) \right)^{\tau}_{k=t+1}, y_{-i}^{\tau} \right)
$$

for all $y_{i}^{-\tau}$
That is, for each $\tau > t$, and for each $\left(\tilde{\theta}_i, \tilde{m}_{-i}\right)$, in all five cases there exists $\varepsilon = \min_{\tilde{\theta}_i, \tilde{m}_{-i}, \tau} \varepsilon\left(\tilde{\theta}_i, \tilde{m}_{-i}, \tau\right)$ (by compactness, this is well-defined and such that $\varepsilon > 0$). Hence, if the continuation strategies are self-correcting, if $f$ is aggregator-based, for any $\varepsilon \in (0, \varepsilon)$, reporting $\theta^e_{i,t}$ or $\theta^e_{i,t}$ at period $t$ does not affect the allocation chosen at periods $\tau > t$ (the opponents’ self-correcting report cannot be affected by $i$-th components of the public history). Hence, for $\varepsilon \in (0, \varepsilon)$, for each $\theta_{-i} \in \Theta_{-i}$, the allocations induced following $s^e_i$ at periods $\tau > t$ and playing $\theta^e_{i,t}$ or $\theta^e_{i,t}$ at history $h^1_i$, respectively $\xi'$ and $\xi^e$, are such that $\xi'_{\tau} = \xi^e_{\tau}$ for all $\tau \neq t$.

Consider types of player $i$, $\theta^e_{i,t}, \theta^e_{i,t} \in \Theta^e_i$ such that for each $\tau < t$, $\theta^e_{i,\tau} = \theta^e_{i,\tau} = \tilde{\theta}_{i,\tau}$ (the one actually reported on the path), for all $\tau > t$ and $\theta^e_{i,\tau} = s^e_{i,\tau}$ as above, while at $t$ respectively equal to $\theta^e_{i,t}$ and $\theta^e_{i,t}$. Thus, the induced allocations are $\xi^e$ and $\xi'$ discussed above, and for each $\tau \neq t$, $\alpha^e_i(\theta^e) = \alpha^e_i(\theta') \equiv \tilde{\alpha}^e_i$.

From strict EPIC, we have that for any $\theta_{-i}$

$$v_i\left(\xi^e, \alpha^e_i(\theta^e), \{\tilde{a}^e_i\}_{\tau \neq t}\right) > v_i\left(\xi', \alpha^e_i(\theta'), \{\tilde{a}^e_i\}_{\tau \neq t}\right)$$

and

$$v_i\left(\xi^e, \alpha^e_i(\theta'), \{\tilde{a}^e_i\}_{\tau \neq t}\right) < v_i\left(\xi', \alpha^e_i(\theta'), \{\tilde{a}^e_i\}_{\tau \neq t}\right)$$
Thus, by continuity, there exists $a^t(\varepsilon)$

$$\alpha_i^t(\tilde{y}^{t-1}, \theta_{i,t}, \theta_{-i,t}) < a^t(\varepsilon) < \alpha_i^t(\tilde{y}^{t-1}, \theta_{i,t}^e, \theta_{-i,t})$$  \hfill (B.6)

such that

$$v_i\left(\xi, a^t(\varepsilon), \{\hat{\alpha}_{i}^{\tau}\}_{\tau \neq t}\right) = v_i\left(\xi', a^t(\varepsilon), \{\hat{\alpha}_{i}^{\tau}\}_{\tau \neq t}\right)$$

From the Single Crossing Condition,

$$v_i\left(\xi, a^*, \{\hat{\alpha}_{i}^{\tau}\}_{\tau \neq t}\right) > v_i\left(\xi', a^*, \{\hat{\alpha}_{i}^{\tau}\}_{\tau \neq t}\right)$$

whenever $a^* > a^t(\varepsilon)$.

Thus, since the continuations in periods $\tau > t$ are the same under both $\theta_{i,t}$ and $\theta_{i,t}^e$, to reach the desired contradiction it suffices to show that for any $y_{-i}^t \in Y_{-i}^t$, $\alpha_i^t(y_i^t, y_{-i}^t) > a^t(\varepsilon)$. (This, for any realization of $\tilde{\theta}_{-i}$).

As in the initial step, define:

$$\delta := \min_{y_{-i}^t \in Y_{-i}^t \text{ and } \theta_{-i,t} \in B_{-i}(h_{i,t}^e, y_{-i}^t)} \left[ \alpha_i^t(y_i^t, y_{-i}^t) - \alpha_i^t(\tilde{y}_{i,t}^e,\theta_{i,t},\theta_{-i,t}) \right]$$  \hfill (B.7)

For any $\varepsilon > 0$, let

$$\psi(\varepsilon) = \max_{\theta_{-i,t} \in \Theta_{-i,t}} \left\{ \alpha_i^t(\tilde{y}_{i,t}^e,\theta_{i,t}^e,\theta_{-i,t}) - \alpha_i^t(\tilde{y}_{i,t}^e,\theta_{i,t},\theta_{-i,t}) \right\}$$  \hfill (B.8)

Since $\alpha_i^t$ is strictly increasing in $\theta_{i,t}$, $\psi(\varepsilon)$ is increasing in $\varepsilon$ and $\psi(\varepsilon) \to 0$ as $\varepsilon \to 0$.

To obtain the desired contradiction, it suffices to choose $\varepsilon$ sufficiently small that

$$\psi(\varepsilon) < \delta$$  \hfill (B.9)
and operate the substitutions as follows

\[
\alpha_i^t (y_i, y_{-i}) \geq \alpha_i^t (\bar{y}_{t-1}^t, \theta_{i,t}', \theta_{-i,t}') + \delta \\
\geq \alpha_i^t (\bar{y}_{t-1}^t, \theta_{i,t}', \theta_{-i,t}') + \delta - \psi (\varepsilon) \\
> \alpha_i^t (\bar{y}_{t-1}^t, \theta_{i,t}', \theta_{-i,t}') \\
> a^t (\varepsilon).
\]

**B.3.3 Proof of Proposition 4.6**

The proof is very similar to those of proposition 4.5.

**Initial Step:** \([s [B (h^{T-1})] = s^c [h^{T-1}] \text{ for each } h^{T-1}].\)

The initial step is the same, to conclude (in analogy with equation B.4), that there exists \(a^T (\varepsilon)\) such that

\[
\alpha_i^T (\bar{y}^{T-1}, \theta_{i,T}', \theta_{-i,T}) < a^T (\varepsilon) < \alpha_i^T (\bar{y}^{T-1}, \theta_{i,T} + \varepsilon, \theta_{-i,T}) \tag{B.10}
\]

such that

\[
v_i (x^{T-1}, f_T (\bar{y}^{T-1}, \theta_{i,T} + \varepsilon, \theta_{-i,T}), a^T (\varepsilon), \alpha^{-T} (\bar{y}^{T-1})) \\
= v_i (x^{T-1}, f_T (\bar{y}^{T-1}, \theta_{i,T}', \theta_{-i,T}), a^T (\varepsilon), \alpha^{-T} (\bar{y}^{T-1})) \tag{B.11}
\]

From the “Strengthened SCC” (def. 4.17),

\[
v_i (x^{T-1}, f_T (\bar{y}^{T-1}, \theta_{i,T} + \varepsilon, \theta_{-i,T}), a^*, \alpha^{-T} (\bar{y}^{T-1})) \\
> v_i (x^{T-1}, f_T (\bar{y}^{T-1}, \theta_{i,T}', \theta_{-i,T}), a^*, \alpha^{-T} (\bar{y}^{T-1}))
\]

whenever \(a^* > a^T (\varepsilon)\)

From this point, the argument proceeds unchanged, concluding the initial step.

**Inductive Step:** [for \(t = 1, ..., T - 1\): if \(s [B (h^t)] = s^c [h^t]\) for all \(h^t\) and all \(\tau > t\) then \(s [B (h^t)] = s^c [h^t]\) for all \(h^t\)]
The argument proceeds as in proposition 4.5, to show that for each \( \tau > t \), and for each \( (\tilde{\theta}, \tilde{m}_{-i}) \), if continuation strategies are self-correcting, there exists \( \varepsilon \left( \tilde{\theta}, \tilde{m}_{-i}, \tau \right) > 0 \) such that:

for all \( \varepsilon \in \left( 0, \varepsilon \left( \tilde{\theta}, \tilde{m}_{-i}, \tau \right) \right) \),

\[
\alpha_i^\tau \left( \tilde{y}_i^{t-1}, \theta_i', \left( s_{i,k} \left( \theta_i', \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^\tau, y_{-i}^\tau \right) = \alpha_i^\tau \left( \tilde{y}_i^{t-1}, \theta_i', \left( s_{i,k} \left( \theta_i', \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^\tau, y_{-i}^\tau \right)
\]

for all \( y_{-i}^\tau \).

Consider types of player \( i \), \( \theta_i', \theta_i^\varepsilon \in \Theta_i^+ \) such that for each \( \tau < t \), \( \theta_i', \tau = \theta_i^\varepsilon = \tilde{\theta}_i, \rho \) (the one actually reported on the path), for all \( \tau > t \) and \( \theta_i, \tau = s_{i,\tau}^\varepsilon \) as above, while at \( t \) respectively equal to \( \theta_i^\varepsilon \) and \( \theta_i' \). By construction, such types are such that for any \( \tau \neq t \), \( \alpha_i^\tau (\theta^\varepsilon) = \alpha_i^\tau (\theta') \).

From strict EPIC, we have that for any \( \theta_{-i} \)

\[
v_i \left( f(\theta^\varepsilon), \alpha_i(\theta^\varepsilon), \{\hat{a}_{i}^\tau\}_{\tau \neq t} \right) > v_i \left( f(\theta'), \alpha_i(\theta^\varepsilon), \{\hat{a}_{i}^\tau\}_{\tau \neq t} \right)
\]

and

\[
v_i \left( f(\theta^\varepsilon), \alpha_i(\theta'), \{\hat{a}_{i}^\tau\}_{\tau \neq t} \right) < v_i \left( f(\theta'), \alpha_i(\theta'), \{\hat{a}_{i}^\tau\}_{\tau \neq t} \right)
\]

Thus, by continuity, there exists \( \alpha_i^t (\varepsilon) \)

\[
\alpha_i^t \left( \tilde{y}_i^{t-1}, \theta_i', \theta_{-i}, t \right) < \alpha_i^t (\varepsilon) < \alpha_i^t \left( \tilde{y}_i^{t-1}, \theta_i^\varepsilon, \theta_{-i}, t \right) \tag{B.12}
\]

such that

\[
v_i \left( \xi^\varepsilon, \alpha_i^t (\varepsilon), \{\hat{a}_{i}^\tau\}_{\tau \neq t} \right) = v_i \left( \xi_i', \alpha_i^t (\varepsilon), \{\hat{a}_{i}^\tau\}_{\tau \neq t} \right)
\]

From the single crossing condition,

\[
v_i \left( f(\theta^\varepsilon), a^*, \{\hat{a}_{i}^\tau\}_{\tau \neq t} \right) > v_i \left( f(\theta''), a^*, \{\hat{a}_{i}^\tau\}_{\tau \neq t} \right)
\]

whenever \( a^* > \alpha_i^t (\varepsilon) \)
To reach the desired contradiction it suffices to show that for any $y_{-i}^t \in Y_{-i}^t$, $\alpha_i^t(y_i^t, y_{-i}^t) > a^t(\varepsilon)$. The remaining part of the proof is identical to proposition 4.5. \qed
Bibliography


