NORTHWESTERN UNIVERSITY

Essays on Robustness and Uncertainty in Game Theory

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Economics

By

Gabriel J. Ziegler

EVANSTON, ILLINOIS

June 2020

Abstract

In this dissertation I examine issues related to uncertainty and robustness in game theory. In Chapter 1 a strategic setting is analyzed where players face Knightian uncertainty about the strategic choices of their opponents. That is, in contrast to the usual Bayesian framework and in line with experimental evidence, players might not be able to form probabilistic belief about the uncertainty they face. Instead players entertain a set of beliefs about this uncertainty. In joint work with Peio Zuazo-Garin, I provide a general model that captures this situation and show that, under certain assumptions, well-known solution concepts, such as rationalizability, are still appropriate.

Chapter 2 reinterprets the model with Knightian uncertainty by interpreting cautiousness as robustness to ambiguity. As before, each players strategic uncertainty is represented by a possibly non-singleton set of beliefs, but now a rational player also wants to make a choice that is robust to this ambiguity. Thus, a rational player chooses a strategy that is a best-reply to every belief in this set. Again in joint work with Peio Zuazo-Garin, I show that the interplay between these two features precludes the conflict between strategic reasoning and cautiousness and therefore solves the inclusion-exclusion problem raised by Samuelson (1992). Notably, my approach provides a simple foundation for the iterated elimination of weakly dominated strategies.

In Chapter 3 I study an information provider who commits to provide information to multiple receivers seeking robustness towards the reasoning of these receivers. The

robustness consideration arises naturally in a setting where information is provided bilaterally. Such a scenario precludes the possibility of commitment to a grand information structure. Consequently, in a strategic situation, each receiver needs to reason about what information other receivers get. Since the information provider does not know this reasoning process, a motivation for a robustness requirement arises: the provider seeks an information structure that performs well no matter how the receivers actually reason. In this chapter, I provide a general method to study how to optimally provide information under these constraints. The main result is a representation theorem, which makes the problem tractable by reformulating it as a linear program in belief space. Furthermore, I provide novel bounds on the correlation among receivers beliefs, which provide even more tractability in some special cases. I illustrate the main result by solving for the optimal provision of information in a stylized model of contract research organizations, which are an integral part of the pharmaceutical industry.

Acknowledgment

There is no doubt that I would not currently sit in Evanston finishing writing this dissertation without the help of my committee members Eddie Dekel, Alessandro Pavan, and Marciano Siniscalchi. I am especially indebted to Marciano Siniscalchi, who was an excellent advisor and mentor throughout my studies at Northwestern University. He was always open to discuss research ideas even if they turned out not to be of great interest. In case they did turn out to be of greater interest he was always available to reflect on ongoing progress and provide very valuable suggestions about next steps. My conversations with him did not only shape my research, but also my more general view about game theory and economics.

Peio Zuazo-Garin is another person that deserves endless credit for me being able to write this dissertation. His private and intellectual friendship was very valuable not only academically but also more generally. We have spent so many hours discussing and developing ideas which other people might call crazy, but at least some of them turned out to be very intriguing nevertheless. For most of the time we had to communicate online due to the location difference, but when we met in person I deeply enjoyed our conversations about game theory while having a beer together. Florian Rundhammer deserves also credit. Without his support, especially in the beginning of my studies at the University of Graz, I might have been totally lost in the academic environment from day one.

There are no words to say how important the support of my family was. My parents, Sissi and Manfred, and my brothers, David, Emanuel, Raphael, and Joachim, always aided me throughout my whole life. Needless to say, without their support and encouragement many of my achievements would have been much more challenging if not even impossible. During my PhD studies, I lost two family members who shaped my worldview: my grandfather Jakob and my great-grandaunt Annemaria. Due to the location difference, I was not able to spent more time with them late in their lives. For this I am truly repentant.

Dedicated to my mother Sissi and my father Manfred.

Contents

List of Tables

List of Figures

CHAPTER 1

Strategic Knightian Uncertainty[1]

"It is a world of change in which we live, and a world of uncertainty. We live only by knowing *something* [...]; while the problems of life [...] arise from the fact that we know so little.

Frank H. Knight (1921, p.199)

[1]This chapter was developed together with Peio Zuazo-Garin. Parts of the exposition, but none of the results (besides Proposition 1) were used in Ziegler and Zuazo-Garin (2020). For consistency reasons the grammatical first person in singular form and associated personal pronouns will be used throughout this dissertation.

1.1. Introduction

Economists commonly use iterated strategy elimination procedures as solution concepts in games. Such procedures thus constitute one of the cornerstones for modeling agents' behavior in economic theory. The predictive power of iterated elimination procedures is in general lower than that of equilibrium-related notions; however, since the latter requires players to correctly forecast their opponents' behavior (see Aumann and Brandenburger, 1995), the former seems more appropriate in situations of multiple equilibria wherein either the players or the economic analyst lack accurate data about past play or such data appears uninformative about future behavior.[2] For instance, this is the case in many application of auction theory, e.g., wireless spectrum, carbon emission rights and online advertising.[3] Consequently, thorough understanding of the forces behind iterated elimination is relevant from both a purely theoretical perspective and a more applied point of view, and is key to effective mechanism design and correct identification in empirical analyses.[4]

The conceptual appeal of iterated elimination procedures is that they carry the intuitive game-theoretic appeal of *strategic reasoning*: if a player is certain that some of her

[2]In Dekel and Fudenberg's (1990) words (p. 243): "Nash equilibrium and its refinements describe situations with little or no 'strategic uncertainty,' in the sense that each player *knows* and is *correct* about the beliefs of the other players regarding how the game will be played. While this will sometimes be the case, it is also interesting to understand what restrictions on predicted play can be obtained when the players' strategic beliefs may be inconsistent, that is, using only the assumption that it is common knowledge that the players are rational."

[3]See Milgrom (1998), Cramton and Kerr (2002), and Varian (2007), respectively.

[4]See Bergemann and Morris (2009, 2011); Bergemann, Morris, and Tercieux (2011) and Aradillas-Lopez and Tamer (2008), respectively.

opponent's strategies are not going to be played, then she might deem some of her own strategies to be unreasonable. This is clearly exemplified by the informal argument for competitive prices in Bertrand duopoly models. Consider a market consisting of profitable, identical firms A and B: If A slightly lowers its mark-up it should absorb all the demand and increases its profit; now, this is easy to forecast by B, which might in turn decide to lower its mark-up more than slightly and thus absorb itself all the demand and increase her profit with respect to the losses obtained under A's, hypothetical, initial slight cut. Obviously, this logic leads to the standard zero mark-up conclusion. Sketches of this elementary intuition in modern economic theory can be traced back to Keynes (1936): "It is not a case of choosing those [faces] that, to the best of one's judgment, are really the prettiest, nor even those that average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practice the fourth, fifth and higher degrees."

However, the formalization of these ideas in game theory was not reached until Pearce (1984) and Bernheim (1984) introduced the concept of *rationalizability*. Shortly after this concept was developed, Brandenburger and Dekel (1987) and Tan and da Costa Werlang (1988) gave a foundation for this concept in terms of strategic reasoning as outlined above within a setting where every player is Bayesian. That is, each player of the game is assumed to be a subjective expected utility maximizer as envisioned by Savage (1954). In a strategic setting, this means that every player has a unique subjective belief about the coplayer's actions.

To this day, most of game theory makes this expected utility assumption either explicitly or implicitly.[5] Using a distinction introduced by Knight (1921), most game theoretic analysis is therefore conducted under the assumption of *risk* without allowing for the possibility of *uncertainty*. Knight explains the difference: "Uncertainty must be taken in a sense radically distinct from the familiar notion of Risk, from which it has never been properly separated.... The essential fact is that 'risk' means in some cases a quantity susceptible of measurement, while at other times it is something distinctly not of this character; and there are far-reaching and crucial differences in the bearings of the phenomena depending on which of the two is really present and operating.... It will appear that a measurable uncertainty, or 'risk' proper, as we shall use the term, is so far different from an unmeasurable one that it is not in effect an uncertainty at all."

In this chapter, I study the implications of this sort of *Knightian uncertainty* on strategic reasoning. For this, I make use of the interpretation and formalization of Knightian uncertainty due to Bewley (2002). In this formalization, each player's strategic uncertainty is represented by a possibly non-singleton set of beliefs. Thus, this allows for Knightian uncertainty on the one hand, but also incorporates standard Bayesian analysis as a special case if the set of beliefs is a singleton set. The main departure of Bewley's decision theory is dropping the completeness axiom and although the main approach in this chapter

[5] In Section 1.2, I discuss related literature including departures from the Bayesian paradigm in game theory.

is conceptual and focused on the implications of Knightian uncertainty on reasoning-based processes in strategic situations, the results provide a methodological contribution for the use of incomplete preferences in game theory, which is a subject of interest in itself aside from its interpretation as Knightian uncertainty. For example, Aumann (1962) questions the completeness axiom: "Of all the axioms of utility theory, the completeness axiom is perhaps the most questionable. [. . .] [W]e find it hard to accept even from the normative viewpoint. Does 'rationality' demand that an individual make definite preference comparisons [. . .]"

For a given set of beliefs, there are two natural interpretation on how a player will choose a strategy that is reasonable. First, a requirement in spirit of Bewley (2002) is to require a strategy will be chosen if and only if it is a best-reply to *every* belief in the player's set of beliefs. I will dub strategies that satisfy this requirement as *rational*. Since such a choice procedure corresponds to incomplete preferences, this requirement is more demanding than usual rationality postulates because it also requires that a strategy being a best-reply to all beliefs actually exists. Thus, my notion of rationality incorporates two behavioral restrictions: (*i*) optimal choice and (*ii*) being able to make a choice. Second, a dual notion to my definition of rationality is in spirit of Lehrer and Tepper (2011): a strategy will be chosen if there *exists one* belief on the set of beliefs that make this strategy a best-reply. These strategies will be called *justifiable*. In contrast to rational strategies, justifiable strategies always exist. The reason is that justifiable strategies correspond to preferences which do satisfy completeness. However, as the axiomatic study of Lehrer and Tepper (2011) highlights these preferences fail to satisfy transitivity instead.

When it comes to reasoning-based analysis in games where players have sets of beliefs, there are also two natural formalization of how players reason. Both are related to the notions of reasonable choices explained above. I say that a player *believes* a certain event if *at least one* belief in her set of beliefs assigns probability one to this event. This is a weak notion of belief because it allows for one player to belief two mutually exclusive events at the same time. On the other hand, similar to rationality, a stronger requirement is to have *all* beliefs in the set to assign probability one to the event under consideration. If this is the case, I will say that a player *fully believes* the event. Thus, a player cannot fully believe two mutually exclusive events, but she can believe both of these events.

Based on the above, I build a framework to explore the behavioral implications of common (full) believe of rationality and justifiability, respectively. Note that either choice requirement drops an assumption that many economists find crucial to be able to conduct any analysis; namely, completeness or transitivity. Thus, maybe surprisingly, the behavioral implications of these choice requirements together with common (full) belief restrictions give rise to predictions that are more or less standard predictions of game theory: whereas common belief in justifiably corresponds to undominated strategies (S^1), all the other combinations correspond the iterated deletion of strictly dominated strategies (S^∞). Table 1.1 summarizes these results and provides references to the formal statements of each result. Therefore, Knightian uncertainty itself does not provide sharper predictions of common belief of rationality in a standard Bayesian setting, which is just S^∞. Furthermore, Knightian uncertainty also does enable to make more behavioral predictions beyond common belief of rationality (S^∞), unless one wants to impose the very weak requirements of common belief of justifiably in the Knightian uncertainty case. Then the only strategies that can be ruled out for each player are strictly dominated strategies.

Table 1.1: Behavioral implications of strategic Knightian uncertainty.

		Epistemic Assumption	
		Belief ($\exists$)	Full Belief ($\forall$)
Choice Assumption	Justifiability ($\exists$)	S^1 (Theorem 2)	S^∞ (Theorem 4)
	Rationality ($\forall$)	S^∞ (Theorem 1)	S^∞ (Theorem 3)

The results summarized in Table 1.1 might suggest that strategic Knightian uncertainty does not really affect the behavioral implications in games relative to the Bayesian benchmark. However, the Knightian uncertainty allows for more flexibility because the Bayesian benchmark is covered as a special case. This flexibility is very salient in the case of belief under Knightian uncertainty. As mentioned above, in my setting a player is allowed to believe two mutually exclusive events.

This flexibility of strategic Knightian uncertainty allows to overcome the inclusion-exclusion problem of Samuelson (1992). Samuelson points out that in a standard Bayesian setting there is a conflict cautious behavior and strategic reasoning. On the one hand, cautiousness requires to assign positive probability on all eventualities (i.e. *inclusion* of eventualities). On the other hand, strategic reasoning allows to rule out certain behavior of the other players (i.e. *exclusion* of eventualities). Therefore, for a player with a unique subjective belief as in the Bayesian case satisfying both, cautiousness and strategic reasoning, is impossible. However, in my Knightian uncertainty framework this problem can be easily solved because multiple beliefs are allowed.

I explore this possibility by introducing cautiousness in my framework. I say that a player is *weakly cautious* if *at least one* belief in her set of beliefs has support equal to the full state space. Intuitively, this means a cautious player assigns positive probability to every eventuality and therefore does not rule out anything. Furthermore, I consider a more demanding criterion dubbed *strong cautiousness*: a strongly cautious player has a set of beliefs with non-empty interior. This implies, intuitively, that a strongly cautious player has (*i*) a set of beliefs that is big and (*ii*) many beliefs with full support. Thus, a strongly cautious player is also weakly cautious. Building on these definitions I show that the Dekel-Fudenberg procedure ($S^{\infty}W$) characterizes the behavioral implications of rationality and weak cautiousness and common belief thereof and that strict rationalizability ($S^{\infty}S^{+}$) characterizes the behavioral implications rationality and strong cautiousness and common belief thereof. Hence, this chapter not only provides the behavioral consequences under strategic Knightian uncertainty but by doing so provides a concise framework for the characterization of the standard iterated deletion procedures which would run into the inclusion-exclusion problem of Samuelson (1992). Furthermore, my analysis sheds light on the requirements to overcome this inclusion-exclusion problem. In addition to allowing for multiple beliefs another requirement is that all of these belief have to matter when it comes to choosing strategies in a game. My notion of rationality does exactly this using an universal quantifier over beliefs. If I only consider justifiable strategies, then (either version of) cautiousness does not provide sharper predictions than players not choosing strictly dominated strategies. The results on cautiousness are summarized in Table 1.2.

Table 1.2: Strategic Knightian uncertainty and cautiousness.

		Cautiousness Assumption	
		Weak	Strong
Choice Assumption	Justifiability (∃)	S^1 (Corollary 2)	S^1 (Corollary 2)
	Rationality (∀)	$S^\infty W$ (Theorem 5)	$S^\infty SR$ (Theorem 6)

The rest of this chapter is structured as follows. Section 1.2 reviews the related literature. Section 1.3 recalls both the game- and decision-theoretic preliminaries and introduces the epistemic framework. All the epistemic characterization results are stated in Section 1.4. Section 1.5 concludes.

1.2. Related Literature

As mentioned in the introduction, Brandenburger and Dekel (1987) and Tan and da Costa Werlang (1988) were among the first to apply Bayesian decision making to give foundations for solution concepts in game theory. In particular, they show that the behavioral implications of rationality common belief of ratioanlity are exactly rationalizable strategies (S^∞). A decade earlier, Armbruster and Böge (1979) and Böge and Eisele (1979) establieshed similar results already. These results were extended to incorporate exogenous uncertainty, for example about the payoffs of the players, by Battigalli and Siniscalchi (2002, 2003) and Dekel, Fudenberg, and Morris (2007). The former also study extensions to dynamic games with the natural adaption of a Bayesian framework to dynamic games. Still within a Bayesian framework but with the use of p-beliefs of Monderer and Samet (1987), Börgers (1994) provides a foundation for the Dekel-Fudenberg procedure[6] ($S^{\infty}W$) via approximate common belief in rationality and cautiousness. All of these papers have in common to provide a foundation of iterated strategy elimination procedures. Aumann (1987) studies a similar question with the additional assumption of a common prior and therefore provides a foundation for (objective) correlated equilibrium. A reasoning based foundation for the widely applied concept of Nash equilibrium was provided Aumann and Brandenburger (1995). This analysis sheds light on the demanding assumptions underlying Nash equilibrium.

Leaving the Bayesian paradigm was spured by the inclusion-exclusion problem of Samuelson (1992). As a response, interactive reasoning based on lexicographic probability systems due to Blume, Brandenburger, and Dekel (1991) were introduced to game theory. Brandenburger (1992) shows that the Dekel-Fudenberg procedure coincides with *permissibility*, which is an iterative procedure based on lexicographic probability systems. Reasoning based foundations within the same decision theoretic setting were provided recently by Catonini and De Vito (2020).[7] To the best of my knowledge, a foundation for what I call *strict rationalizability* does not appear in the literature. Another solution con-

[6]This solution concept was introduced by Dekel and Fudenberg (1990) in a different context.
[7]Asheim and Dufwenberg (2003) provide a foundation in a slightly different setting.

cept that runs into the inclusion-exclusion problem is *iterated admissibility*, i.e. the iterative deletion of weakly dominated strategies. Lexicographic probability systems seem useful also to provide a foundation for iterated admissibility. Within this stream of literature, seminal work by Brandenburger, Friedenberg, and Keisler (2008) establishes foundations for iterated admissibility in terms of rationality and common assumption of rationality (RCAR); however, the same authors reveal a vexatious reality: RCAR is empty in every non-trivial game. By now, there are solutions available either using lexicographic probability systems or using my framework of strategic Knightian uncertainty. The latter is presented in Chapter 2, which also provides a more thorough discussion about the solutions with lexicographic probability systems.

Knightian uncertainty can be interpreted as a form of ambiguity.[8] Combining ambiguity with interactive reasoning is relatively new. Ahn (2007) provides a theory of hierarchies of ambiguity. Like us, he assumes that each player is allowed to have a set of beliefs, a set of beliefs about the set of beliefs of the other players, and so on. He goes on by constructing a universal type strcutre akin to the standard Bayesian version of Mertens and Zamir (1985) and Brandenburger and Dekel (1993). In my setting of strategic Knightian uncertainty, I rely on Ahn's construction. Independently from my work, Dominiak and Schipper (2020) study ambiguity and interactive reasoning in games. In contrast to my work, they use Choquet expected utility. With their version of *belief*, they provide a characterization of the behavioral implications of common belief of Choquet

[8] In fact, I put forward this interpretation in Chapter 2, where I study the foundation of iterated admissibility.

rationality, which is different from standard rationalizability. A seminal work by Epstein (1997) goes beyond ambiguity by allowing general preferences, which include a wide class of decision theories but maintains the completeness assumption.[9] In his model common belief in rationality is behaviorally equivalent to appropriately adapted iterative deletion procedures. Chen, Luo, and Qu (2016) extend Epstein's analysis to include, among other things, incomplete preferences á la Bewley (2002). Their notion of *belief* can be interpreted as deeming the complement of given event as impossible. Thus, their notion is similar to my notion of *full belief*. In turn, my Theorem 3 and Theorem 4 can be seen as special cases of their model adding the additional restriction that Knightian uncertainty is commonly believed as well, i.e. ruling out other preferences that are allowed in the general class considered by Chen et al. (2016). However, in their formalization it is implicit that even if preferences are incomplete, players will have a preferred strategy. For incomplete preferences this might not be true. Thus, I explicitly demand that a rational player has such a preferred strategy. If a player is not able to find a preferred strategy due to incomplete preferences, he will be called irrational in my framework. In my model these types of irrational players are allowed for.

Moving beyond reasoning-based foundations of solution concepts, allowing for non-Bayesian behavior in game theory is studied more thoroughly and reviewing it is beyond the scope of this chapter.[10] However, there are a few exceptions that are direclty related to my work. First, Kokkala et al. (2019) allow players to have incomplete preferences and

[9] See also Epstein and Wang (1996) for a construction of a universal preference space.

[10] A recent paper in this strand of the literature is Kokkala, Berg, Virtanen, and Poropudas (2019), which includes many references to this literature.

provide a suitable version of rationalizability for their setting. In contrast to my setting, they allow players to have incomplete preferences over outcomes of the game. In my case, incompleteness arises from multiple beliefs only but players have a complete ordering over outcomes. Furthermore, Kokkala et al. do not provide reasoning-based foundations for their proposed version of rationalizability. Second, Lopomo, Rigotti, and Shannon (2011, 2017) study the implications of Knightian uncertainty as formalized here on moral hazard and mechanism design, respectively. A key condition that arises in Lopomo et al. (2017) is that the set of beliefs has full dimensionality, which is related to, but weaker, than my notion of strong cautiousness. Chiesa, Micali, and Zhu (2015) study Vickrey auctions where bidders face Knightian uncertainty about their own evaluations. Knightian uncertainty was also employed in models of general equilibirum by, for example, Rigotti and Shannon (2005), Kajii and Ui (2009), Easley and O'Hara (2010), Dana and Riedel (2013), Chambers (2014), Beissner and Riedel (2019), and Chambers and Melkonyan (2020).

1.3. Preliminaries

This section presents the main standard concepts and formalism related to game and decision theory. Subsection 1.3.1 recalls the formalization of strategic-form games and states the relevant solution concepts. Since my analysis models players as individual decision makers whose beliefs may display Knightian uncertainty, Subsection 1.3.2 recalls the necessary decision-theoretical toolbox and Subsection 1.3.3 illustrates how games can be envisioned as decision problems as is standard in the literature since Tan and da

Costa Werlang (1988).[11] Subsection 1.3.4 formalizes the idea of (higher-order) reasoning in games. Finally, Subsection 1.3.5 defines the relevant concepts of *justifiable* and *rational choices* as well as the notions of *belief* and *full belief*.

1.3.1. Games and iterated strategy elimination

A *game* consists of a tuple $G := \langle I, (S_i, u_i)_{i\in I}\rangle$ where I is a finite set of *players*, and for each player i there is a finite set of (pure) *strategies* S_i and a *utility function* $u_i : S \to \mathbb{R}$, where $S := \prod_{i\in I} S_i$ denotes the set of strategy *profiles*. For each player i a randomization of own strategies $\sigma_i \in \Delta(S_i)$ is referred to as a *mixed* strategy,[12] and a probability measure $\mu_i \in \Delta(S_{-i})$, where $S_{-i} := \prod_{j\neq i} S_j$, as a *conjecture*. When necessary, with some abuse of notation, I use s_i to refer to the degenerate mixed strategy that assigns probability one to s_i. Each conjecture μ_i and possibly mixed strategy σ_i naturally induce *expected* utility $U_i(\mu_i; \sigma_i)$ and based on this, each player i's *best-reply* correspondence is defined by assigning to each conjecture μ_i the subset of pure strategies $BR_i(\mu_i)$ that maximize its corresponding expected utility.[13] For the rest of the chapter I consider game G to be fixed and therefore drop most explicit mentions to it.

Following the duality results of Pearce (1984), best-replies correspondences allow for easily formalizing the standard procedures of iterated elimination whose foundations I later study in Section 1.4:

[11] Di Tillio (2008) provides a more modern and detailed formulation.

[12] Throughout this chapter, for any topological space X, as usual, $\Delta(X)$ denotes the set of probability measures on the Borel σ-algebra of X.

[13] That is, given conjecture μ_i the expected utility is $U_i(\mu_i; \sigma_i) := \sum_{(s_{-i};s_i)\in S} \mu_i[s_{-i}] \cdot \sigma_i[s_i] \cdot u_i(s_{-i}; s_i)$ for each possibly mixed strategy σ_i, and the set of best-replies is $BR_i(\mu_i) := \arg\max_{s_i\in S_i} U_i(\mu_i; s_i)$.

Rationalizability

Strategy s_i is *rationalizable* if it survives the iterated elimination of strictly dominated strategies; i.e., if it is not strictly dominated given strategy profiles $S_{-i} \times S_i$, it is not strictly dominated given strategy profiles $S^1_{-i} \times S^1_i$ consisting only of strategies surviving the first elimination round, etc. Thus, formally, strategy s_i is rationalizable if $s_i \in S_i^\infty := \bigcap_{n\geq 0} S_i^n$, where $S_i^0 := S_i$ and for any $n \in \mathbb{N}$,

$$S_i^n := \left\{ s_i \in S_i^{n-1} \;\middle|\; \begin{array}{l} \text{There exists some } \mu_i \in \Delta(S_{-i}) \text{ such that:} \\ (i) \quad \operatorname{supp} \mu_i \subseteq \prod_{j\neq i} S_j^{n-1}, \\ (ii) \quad s_i \in BR_i(\mu_i) \end{array} \right\}.$$

Note that with this notation S_i^1 denotes the set of undominated strategies for player i.

The Dekel-Fudenberg Procedure

The procedure due to Dekel and Fudenberg (1990) consists of performing an initial elimination of weakly dominated strategies followed by iterated elimination of strictly dominated strategies; in such case I say that strategy s_i survives the Dekel-Fudenberg procedure if

$s_i \in S^\infty W_i := \bigcap_{n\geq 0} S^n W_i$, where

$$S^0 W_i := \left\{ s_i \in S_i \;\middle|\; \begin{array}{l} \text{There exists some } \mu_i \in \Delta(S_{-i}) \text{ such that:} \\ \\ (i) \quad \operatorname{supp} \mu_i = \prod_{j\neq i} S_j, \\ \\ (ii) \quad s_i \in BR_i(\mu_i) \end{array} \right\},$$

and for any $n \geq 1$,

$$S^n W_i := \left\{ s_i \in S^{n-1} W_i \;\middle|\; \begin{array}{l} \text{There exists some } \mu_i \in \Delta(S_{-i}) \text{ such that:} \\ \\ (i) \quad \operatorname{supp} \mu_i \subseteq \prod_{j\neq i} S^{n-1} W_j, \\ \\ (ii) \quad s_i \in BR_i(\mu_i) \end{array} \right\}.$$

Strict Rationalizability

Finally, I say that strategy s_i is *strictly rationalizable* if it survives an initial elimination of of non-strict best-replies followed by iterated elimination of strictly dominated strategies. Formally, $s_i \in S^\infty S_i^+ := \bigcap_{n\geq 0} S^n S_i^+$, where

$$S^0 S_i^+ := \left\{ s_i \in S_i \;\middle|\; \begin{array}{l} \text{There exists some } \mu_i \in \Delta(S_{-i}) \text{ such that:} \\ \\ (i) \quad \operatorname{supp} \mu_i \subseteq \prod_{j\neq i} S_j, \\ \\ (ii) \quad BR_i(\mu_i) = \left\{ s_i' \in S_i \;\middle|\; \begin{array}{l} \text{For any } s_{-i} \in \prod_{j\neq i} S_j, \\ \\ u_i(s_{-i}; s_i') = u_i(s_{-i}; s_i) \end{array} \right\} \end{array} \right\},$$

and for any $n \geq 1$,

$$S^n S_i^+ := \left\{ s_i \in S^{n-1} S_i^+ \;\middle|\; \begin{array}{ll} \multicolumn{2}{l}{\text{There exists some } \mu_i \in \Delta(S_{-i}) \text{ such that:}} \\ (i) & \operatorname{supp} \mu_i \subseteq \prod_{j \neq i} S^{n-1} S_j^+, \\ (ii) & s_i \in BR_i(\mu_i) \end{array} \right\}.$$

The inclusion relation between these solution concepts is easy to see: for any player i, $S^\infty S_i^+ \subseteq S^\infty W_i \subseteq S_i^\infty$ and furthermore, the inclusions might be strict.

1.3.2. Decision problems and preferences

I follow the reformulation of Anscombe and Aumann's (1963) framework by Fishburn (1970). The decision maker faces *decision environment* (Z, Θ) where: (i) Z is a set of *outcomes*, which can be informally understood as the elements that will ultimately yield direct utility to the decision maker; and (ii) Θ is a set of *states* (of the world) about which the decision maker might face uncertainty, and which may affect how her choices relate to outcomes. I refer to randomizations of outcomes, $\ell \in \Delta(Z)$, as *lotteries*. A *preference* is a binary relation $\succsim$ over the set of *acts*, $\mathcal{F}$, which is the collection of all maps $f : \Theta \to \Delta(Z)$ that assign a lottery to each state. $\mathscr{M}(\Theta)$ denotes the set of closed and convex nonempty subsets of Θ.[14] Throughout the chapter I focus on two classes of preferences: (i) *Bewley* preferences

[14] To be more mathematically precise, Z is assumed to be finite, Θ is compact and metrizable and the elements of $\mathcal{F}$, simple and measurable in the Borel σ-algebra of Θ. Space $\mathscr{M}_i(S_{-i} \times T_{-i})$ is endowed with the topology induced by the Hausdorff metric and is therefore compact and metrizable.

as introduced by Bewley (2002) and (*ii*) *justifiable* preferences due to Lehrer and Tepper (2011).[15] The main point of departure from the preferences of a standard Bayesian decision maker (i.e., one whose preferences satisfy the axioms by Anscombe and Aumann (1963) is that completeness or transitivity of the preferences is dropped, respectivley. Theorem 1 by Gilboa et al. (2010) provides the following convenient representation for Bewley preferences: $\succsim$ is a Bewley preference if and only if there exist a non-constant utility-function $u : Z \to \mathbb{R}$ and a set of *ambiguous beliefs* $M \in \mathscr{M}(\Theta)$ such that for every pair of acts f, g,[16]

$$f \succsim g \iff \int_\Theta \mathbb{E}_{f(\theta)}[u(z)]\,\mathrm{d}\mu \geq \int_\Theta \mathbb{E}_{g(\theta)}[u(z)]\,\mathrm{d}\mu \text{ for every } \mu \in M.$$

Similarly, Theorem 1 of Lehrer and Tepper (2011) gives a representation for justifiable preferences: $\succsim$ is a justifiable preference if and only if there exist a non-constant utility-function $u : Z \to \mathbb{R}$ and a set of *ambiguous beliefs* $M \in \mathscr{M}(\Theta)$ such that for every pair of acts f, g,[17]

$$f \succsim g \iff \int_\Theta \mathbb{E}_{f(\theta)}[u(z)]\,\mathrm{d}\mu \geq \int_\Theta \mathbb{E}_{g(\theta)}[u(z)]\,\mathrm{d}\mu \text{ for at least one } \mu \in M.$$

[15] For Bewley preferences, I rely on a more modern version by Gilboa, Maccheroni, Marinacci, and Schmeidler (2010). Bewley's (2002) version requires the decision-maker to have a designated default act always chosen unless ranked strictly lower than some alternative. This is commonly known in the literature as *inertia* (see, Bewley, 2002, or Lopomo et al., 2011). Furthermore, the version of Gilboa et al. (2010) allows for infinite state spaces which are necessary in my framework.

Lehrer and Tepper (2011) consider an axiomatization with a finite state space only. They claim (see Footnote 7) that an extension to infinite states is possible in a similar manner as in Gilboa et al. (2010). I make use of such an extenstion without providing the details.

[16] More precisely, M is non-empty, closed, and convex. Moreover, M is unique and u is unique up to positive affine transformations.

[17] The details of the previous footnote apply to this representation as well.

In both cases, a decision maker's epistemic attitude with respect to the source of uncertainty may not be represented by a single belief, as in the standard case, but rather by a possibly non-singleton set of beliefs that reflects the decision maker's possible ambiguity towards that source of uncertainty.

1.3.3. Games as decision problems

Players are envisioned as individual decision makers facing a decision problem where their opponents' strategies are part of the description of the states of the world and strategies are the feasible acts. For obvious reasons, for each player i, game G is a very specific decision problem (Z_i, Θ_i, F_i) consisting of:

- *Outcomes.* In contexts of complete (payoff-relevant) information, player i's utility depends only on the strategy profiles chosen in the game; hence, I identify outcomes with the latter: $Z_i := S$.

- *States.* Player i's primary source of uncertainty (and the only payoff-relevant one) is strategic: it refers to her opponents' behavior (S_{-i}). However, player i's beliefs about her opponents' strategies could be affected by an additional non payoff-relevant unobserved parameters about which she might face uncertainty, say T_{-i}.[18] I identify the set of states of the world with these joint sources of uncertainty: $\Theta_i := S_{-i} \times T_{-i}$.

- *Acts and feasible acts.* Player i's set of acts is $\mathcal{F}_i := \Delta(S)^{S_{-i} \times T_{-i}}$. Notice that within the context of a game this set of acts is not feasible. First, player i cannot make her choice

[18] To ensure appropriate construction, T_{-i} is assumed to be compact and metrizable.

contingent on a parameter t_{-i} that she does not observe. Second, in situations of simultaneous choice, player i cannot make her choice contingent on her opponents' choices. Still, player i might (and typically will) have preferences on modeled but unavailable options. The set of player i's feasible acts is then identified with her mixed strategies:

$$F_i := \left\{ f \in \mathcal{F}_i \;\middle|\; \begin{array}{l} \text{There exists a } \sigma_i \in \Delta(S_i) \text{ such that:} \\ \\ f(s_{-i}, t_{-i})[(s'_{-i}; s'_i)] = \begin{cases} \sigma_i[s'_i] & \text{if } s'_{-i} = s_{-i}, \\ \\ 0 & \text{otherwise,} \end{cases} \\ \\ \text{for any } (s_{-i}, t_{-i}) \in S_{-i} \times T_{-i} \text{ and any } (s'_{-i}; s'_i) \in S \end{array} \right\}.$$

In addition, remember that game G already incorporates utility functions; thus, each player i's set of Bewley preferences under consideration needs to be restricted to these preferences whose risk attitude is represented by utility function u_i. Now, Theorem 1 by Gilboa et al. (2010) implies that for any set of parameters T_{-i}, each Bewley preference for decision environment $(S, S_{-i} \times T_{-i})$ whose risk attitude is represented by u_i is biunivocally associated with ambiguous beliefs $M_i \subseteq \Delta(S_{-i} \times T_{-i})$.[19] Thus, there is no loss of generality in switching the focus from Bewley preferences to ambiguous beliefs, the collection of which I denote by $\mathscr{M}_i(S_{-i} \times T_{-i})$. The same applies for justifiable preferences via Theorem 1 of Lehrer and Tepper (2011).

[19]Remember that M_i is non-empty, closed, and convex. Of course, M_i is a subset of $\Delta(S_{-i})$ in cases in which I omit set of parameters T_{-i}.

1.3.4. Reasoning in Games

In this section I present the epistemic framework that I employ in my analysis. Formally, for each player I specify a choice and a representation of her beliefs on her opponents' strategies, her beliefs on her opponents' beliefs over their opponents' strategies, etc. These elements suffice to assess whether under such specifications, the player is being rational or has certain higher-order beliefs on her opponents' rationality. To formalize these concepts some previous methodological work is required. As seen above, when Knightian uncertainty is allowed for, the representation of uncertainty may require non-singleton sets of beliefs. It follows that standard type structures as introduced by Harsanyi (1967) and standard belief-hierarchies à la Mertens and Zamir (1985) are not suitable for analyzing strategic reasoning: they fail to capture the possibility of Knightian uncertainty. Instead, I rely on a modified version of type structure that accounts for ambiguous beliefs.[20]

The study of strategic reasoning requires an instrument that formalizes players' beliefs about their opponents' choices, players' beliefs about their opponents' beliefs about their opponents' choices and so on. When players have complete preferences this hierarchical uncertainty can easily be represented through type structures. Thus, it is convenient to extend the definition of the latter so that can deal with the possibility of ambiguity. Formally, an *ambiguous type structure* consists of a list $\mathcal{T} := \langle T_i, M_i \rangle_{i \in I}$ where for each player i there is:[21]

[20]These type structures are regarded to Ahn's (2007) *ambiguous hierarchies* what Harsanyi's (1967) type structures are to Mertens and Zamir's (1985) belief hierarchies.

[21]I assume each T_i to be compact and metrizable and each M_i, continuous. See Footnote 18.

(i) A set of (*ambiguous*) *types* T_i.

(ii) An *ambiguous belief map* $M_i : T_i \to \mathscr{M}_i(S_{-i} \times T_{-i})$, where $T_{-i} := \prod_{j \neq i} T_j$, that associates each type with ambiguous beliefs on opponents' strategy-type pairs.

It is easy to see why ambiguous type structures capture the idea of hierarchical reasoning mentioned at the beginning of the paragraph. For any player i's type t_i it is possible to compute the following by recursive marginalization:[22]

(1) *First-order* ambiguous beliefs that represent type t_i's uncertainty about her opponents' strategies, $M_{i,1}(t_i) \in \mathscr{M}_{i,1} := \mathscr{M}_i(S_{-i})$, which is easily obtained by taking the marginals on S_{-i} of the beliefs in $M_i(t_i)$.

(2) *Second-order* ambiguous beliefs that represent type t_i's uncertainty about her opponents' strategy-first-order ambiguous beliefs pairs, $M_{i,2}(t_i) \in \mathscr{M}_{i,2} := \mathscr{M}_i(\prod_{j \neq i}(S_j \times \mathscr{M}_{j,1}))$.

...

(n) n^{th}*-order* ambiguous beliefs that represent type t_i's uncertainty about her opponents' strategy-$(n-1)^{\text{th}}$-order ambiguous beliefs pairs, $M_{i,n}(t_i) \in \mathscr{M}_{i,n} := \mathscr{M}_i(\prod_{j \neq i}(S_j \times \mathscr{M}_{j,n-1}))$.

[22]The conceptual simplicity that follows contrasts the notational complexity that it requires; technically, for each $n \in \mathbb{N}$:

$$M_{i,n+1}(t_i) = \left\{ \mu_i \in \Delta\left(\prod_{j \neq i}(S_j \times \mathscr{M}_{j,n})\right) \middle| \begin{array}{l} \text{There exists some } \mu_i' \in M_i(t_i) \text{ such that:} \\ \mu_i[E] = \mu_i'\left[\left(\prod_{j \neq i}(\mathrm{id}_{S_j} \times M_{j,n})\right)^{-1}(E)\right] \\ \text{for every measurable } E \subseteq \prod_{j \neq i} S_j \times \mathscr{M}_{j,n} \end{array} \right\}.$$

...

Ambiguous type structure $\mathcal{T}$ is said to be *complete* if every map M_i is surjective, that is, if for every possible ambiguous beliefs the ambiguous type structure may admit, there exists some type that is mapped to such ambiguous beliefs.[23]

1.3.5. Behavior and Beliefs

The analysis of each player i's reasoning is focused on strategy-type pairs (s_i, t_i), which specify both player i's choice, and as described above, her ambiguous beliefs on her opponents' choices, her ambiguous beliefs on her opponents' first-order ambiguous beliefs, etc. Thus, each strategy-type pair (s_i, t_i) enables questions such as the following to be addressed: Is player i rational given her beliefs? Do her preferences embody some kind of ambiguity? What are her higher-order beliefs about her opponents' rationality and ambiguity? Next, I first formalize the notions of rationality and justifiability that I employ. Second, I introduce the formalizations of belief and full belief.

Justifiability

I say that strategy s_i is *justifiable* for type t_i if s_i is a best-reply to at least one first-order ambiguous belief induced by t_i; thus, the set of strategy-type pairs in which player i is

[23]As shown by Ahn (2007), the answers to the following modified questions in (Dekel and Siniscalchi, 2015, p. 629): "Is there a[n] [ambiguous] type structure that generates *all* [ambiguous] hierarchies of beliefs? Is there a[n] [ambiguous] type structure into which every other [ambiguous] type structure can be embedded?" are yes, and yes. Within a Bayesian framework, Friedenberg (2010) studies such a richness requirement more generally.

justified is formalized as follows:

$$J_i := \left\{ (s_i, t_i) \in S_i \times T_i \,\middle|\, s_i \in \bigcup_{\mu_i \in M_i(t_i)} BR_i(\operatorname*{marg}_{S_{-i}} \mu_i) \right\}.$$

These sets are non-empty because the best-reply correspondence is non-empty. Furthermore, J_i is closed, hence also an event:[24]

Lemma 1. *Let G be a game and $\mathcal{T}$ an ambiguous type structure. For any player i, J_i is a closed subset of $S_i \times T_i$.*

Proof. Fix player i and convergent sequence of strategy-type pairs $\{s_i^n, t_i^n\}_{n \in \mathbb{N}} \subseteq R_i$ with limit (s_i, t_i). Notice that due to finiteness of S_i I can assume without loss of generality that $s_i^n = s_i$ for any $n \in \mathbb{N}$. Now, for each $n \in \mathbb{N}$ set $M_i^n = M_i(t_i^n)$ and let $M_i = M_i(t_i)$. By definition it holds that $s_i \in BR_i(\operatorname{marg}_{S_{-i}} \mu_i^n)$ for at least one $\mu_i^n \in M_i^n$ for any $n \in \mathbb{N}$. Without loss (otherwise take an appropriate subsequence) μ_i^n converges to μ_i. $\mu_i \in M_i$, because M_i^n's as functions of the types are continuous and M_i^n's converge to M_i, which is closed. Since $s_i \in BR_i(\operatorname{marg}_{S_{-i}} \mu_i^n)$ for any $n \in \mathbb{N}$, upper-hemicontinuity of the best-reply correspondence implies that $s_i \in BR_i(\operatorname{marg}_{S_{-i}} \mu_i)$. Hence, I conclude that $(s_i, t_i) \in J_i$ and therefore, that the latter is closed. ■

[24] In the following, for beliefs I will use the Lévy-Prohorov metric (denoted with d) which induces the weak*-topology. For any $\mu \in \Delta(X)$ and any closed $M \subseteq \Delta(X)$, define $d(\mu, M) = \min_{\mu' \in M} d(\mu, \mu')$.

Rationality

I say that strategy s_i is *rational* for type t_i if s_i is a best-reply to every first-order ambiguous belief induced by t_i; thus, the set of strategy-type pairs in which player i is rational is formalized as follows:

$$R_i := \left\{ (s_i, t_i) \in S_i \times T_i \,\middle|\, s_i \in \bigcap_{\mu_i \in M_i(t_i)} BR_i(\operatorname*{marg}_{S_{-i}} \mu_i) \right\}.$$

First, I claim that R_i is a well-defined event:

Lemma 2. *Let G be a game and $\mathcal{T}$ an ambiguous type structure. For any player i, R_i is a closed subset of $S_i \times T_i$.*

Proof. Fix player i and convergent sequence of strategy-type pairs $\{s_i^n, t_i^n\}_{n \in \mathbb{N}} \subseteq R_i$ with limit (s_i, t_i). Notice that due to finiteness of S_i I can assume without loss of generality that $s_i^n = s_i$ for any $n \in \mathbb{N}$. Now, for each $n \in \mathbb{N}$ set $M_i^n = M_i(t_i^n)$ and let $M_i = M_i(t_i)$. By definition it holds that $s_i \in \bigcap_{\mu_i' \in M_i^n} BR_i(\operatorname{marg}_{S_{-i}} \mu_i')$ for any $n \in \mathbb{N}$. Pick now arbitrary $\mu_i \in M_i$. Furthermore, $\{d(\mu_i, M_i^n)\}_{n \in \mathbb{N}}$ converges to 0 and thus, that there exists some sequence $\{\mu_i^n\}_{n \in \mathbb{N}}$ with limit μ_i such that $\mu_i^n \in M_i^n$ for any $n \in \mathbb{N}$. Since $s_i \in BR_i(\operatorname{marg}_{S_{-i}} \mu_i^n)$ for any $n \in \mathbb{N}$, upper-hemicontinuity of the best-reply correspondence implies that $s_i \in BR_i(\operatorname{marg}_{S_{-i}} \mu_i)$. Hence, I conclude that $(s_i, t_i) \in R_i$ and therefore, that the latter is closed. ∎

Second, note that the definition implicitly requires each type t_i, in order to be eligible for rational behavior, to satisfy that the intersection of the best-replies to the ambiguous first-order beliefs induced by it is non empty. This is a consistency requirement in the vein of Bayesian updating for conditional probability systems in the literature of dynamic games: When a conditional probability system fails to satisfy Bayesian updating it may not admit sequential best-replies.[25]

I refer to the types that admit rational choices as *decisive*. The foundation of decisiveness in terms of preferences is provided by Proposition 1 below. Decisive types are those induced by preferences that are possibly incomplete but display completeness *at the top*: the decision maker is indifferent between two acts that are not less preferred than another act.[26]

Proposition 1 (Behavioral foundation of decisiveness). *Let G be a game and $\mathcal{T}$ an ambiguous type structure. Then, any player i's type t_i is decisive if and only if there exists a subset of feasible acts $F_i^* \subseteq F_i$, such that $\succsim_i$, the Bewley preference represented by $(u_i, M_i(t_i))$, satisfies*

$$f \sim_i g \succ_i h,$$

for every $f, g \in F_i^*$ and every $h \in F_i \setminus F_i^*$.

Proof. Fix player i, type t_i and event $E_{-i} \subseteq S_{-i} \times T_{-i}$ and let $\succsim_i$ denote the Bewley preference represented by $(u_i, M_i(t_i))$. The 'if' part is immediate, so I focus on the 'only if' part. To

[25] I thank Pierpaolo Battigalli for this observation. This issue, which refers to the distinction between a choice being optimal or undominated, is discussed in further detail below.

[26] Despite the following characterization relying on an axiom evoking existence, G being a finite game implies that the verification of the condition requires only finitely many bets.

see it simply take $S_i^* := \bigcap_{\mu_i \in M_i(t_i)} BR_i(\text{marg}_{S_{-i}} \mu_i)$ and set:

$$F_i^* := \left\{ f_i \in \mathcal{F}_i \;\middle|\; \begin{array}{ll} \multicolumn{2}{l}{\text{There exists a } \sigma_i \in \Delta(S_i) \text{ such that:}} \\ (i) & f_i(s_{-i}, t_{-i})[(s'_{-i}; s'_i)] = \begin{cases} \sigma_i[s'_i] & \text{if } s'_{-i} = s_{-i}, \\ 0 & \text{otherwise,} \end{cases} \\ & \text{for any } (s_{-i}, t_{-i}) \in S_{-i} \times T_{-i} \text{ and any } (s'_{-i}; s'_i) \in S, \\ (ii) & \sigma_i[S_i^*] = 1 \end{array} \right\}.$$

Clearly, $F^* \subseteq F$ and $f \sim_i g \succ_i h$ for every $f, g \in F_i^*$ and every $h \in F_i \setminus F_i^*$. ■

Notice that in the presence of incomplete preferences ‘undomination’ (an act not being strictly worse than *some* other act) and ‘optimality’ (an act being at least as good as *every* other act) are two different concepts, which is not the case under completeness: An optimal act is always undominated but an undominated act might not be optimal; furthermore, every Bewley preference admits undominated acts, but there may not exist optimal ones. Decisiveness ensures the existence of the latter, which in turn, restores the equivalence of undomination and optimality. In consequence, imposing decisiveness on incomplete preferences is similar in spirit to the requirement of Bayesian updating for conditional probability systems in the literature of extensive-form games.[27] As for

[27]The definition of conditional probability systems (originally due to Rényi, 1955) requires the decision maker to update her beliefs according to the chain rule whenever possible; this requirement is usually referred to as *Bayesian updating*.

decisiveness, Bayesian updating guarantees the existence of optimal strategies by forcing them to be equivalent to undominated ones.

Belief and Full Belief

Hereafter I refer to measurable subsets $E \subseteq S \times T$ as events. A standard Bayesian decision maker is said to *belief* event E when the unique subjective belief induced by her preference has support included in E. With sets of beliefs two extensions are natural: (*i*) existence of one belief with support in E and (*ii*) all beliefs being supported in E. The first extension, I will call *belief* and the latter will be referred to as *full belief*. Formally, for any player i, any type t_i and any event $E_{-i} \subseteq S_{-i} \times T_{-i}$ I say that type t_i *believes* E_{-i} if at least one belief in $M_i(t_i)$ assigns probability one to E_{-i}, i.e. there exists $\mu_i \in M_i(t_i)$ such that $\mu_i[E] = 1$. I denote the set of player i's strategy-type pairs in which the type beliefs E_{-i} by $B_i(E_{-i})$. Similarly, I say that type t_i *fully believes* E_{-i} if all beliefs in $M_i(t_i)$ assigns probability one to E_{-i}, i.e. $\mu_i[E] = 1$ for all $\mu_i \in M_i(t_i)$. Henceforth, $F_i(E_{-i})$ denotes the set of player i's strategy-type pairs in which the type fully beliefs E_{-i}. Note that both definitions of beliefs reduce to the Bayesian case, i.e. if the set of beliefs is a singleton.

I end this section with proving that $B_i(E_{-i})$ and $F_i(E_{-i})$ are well-defined events whenever E_{-i} is closed because both sets are closed:

Lemma 3. *Let G be a game and $\mathcal{T}$ an ambiguous type structure. For any player i and any closed event $E \subset S_{-i} \times T_{-i}$, $B_i(E)$ is a closed subset of $S_i \times T_i$.*

Proof. Fix player i and any closed event $E \subseteq S_{-i} \times T_{-i}$. Consider convergent sequence of strategy-type pairs $\{s_i^n, t_i^n\}_{n \in \mathbb{N}} \subseteq B_i$ with limit (s_i, t_i). W.l.o.g assume $s_i^n = s_i$ for any $n \in \mathbb{N}$. Now, for each $n \in \mathbb{N}$ set $M_i^n = M_i(t_i^n)$ and let $M_i = M_i(t_i)$. By definition it holds that $\mu_i^n[E] = 1$ for at least one $\mu_i^n \in M_i^n$ for any $n \in \mathbb{N}$. Without loss (otherwise take an appropriate subsequence) μ_i^n converges to μ_i. $\mu_i \in M_i$, because M_i^n's as functions of the types are continuous and M_i^n's converge to M_i, which is closed. By the Portmanteau theorem, $\mu_i[E] \geq \limsup_{n \to \infty} \mu_i^n[E] = 1$. Thus, $\mu_i[E] = 1$ for at least one $\mu_i \in M_i$. Hence, I conclude that $(s_i, t_i) \in B_i(E)$ and therefore, that the latter is closed. ■

Lemma 4. *Let G be a game and $\mathcal{T}$ an ambiguous type structure. For any player i and any closed event $E \subset S_{-i} \times T_{-i}$, $F_i(E)$ is a closed subset of $S_i \times T_i$.*

Proof. Fix player i and any closed event $E \subseteq S_{-i} \times T_{-i}$. Consider convergent sequence of strategy-type pairs $\{s_i^n, t_i^n\}_{n \in \mathbb{N}} \subseteq B_i$ with limit (s_i, t_i). W.l.o.g assume $s_i^n = s_i$ for any $n \in \mathbb{N}$. Now, for each $n \in \mathbb{N}$ set $M_i^n = M_i(t_i^n)$ and let $M_i = M_i(t_i)$. By definition it holds that $\mu_i^n[E] = 1$ for all $\mu_i^n \in M_i^n$ for any $n \in \mathbb{N}$. Pick now arbitrary $\mu_i \in M_i$. Furthermore, $\{d(\mu_i, M_i^n)\}_{n \in \mathbb{N}}$ converges to 0 and thus there exists some sequence $\{\mu_i^n\}_{n \in \mathbb{N}}$ with limit μ_i such that $\mu_i^n \in M_i^n$ for any $n \in \mathbb{N}$. By the Portmanteau theorem, $\mu_i[E] \geq \limsup_{n \to \infty} \mu_i^n[E] = 1$. Thus, $\mu_i[E] = 1$ for all $\mu_i \in M_i$. Hence, I conclude that $(s_i, t_i) \in F_i(E)$ and therefore, that the latter is closed. ■

1.4. Epistemic Implications on Behavior

This section presents the main results of this chapter. Namely, the behavioral implications of strategic Knightian uncertainty as summarized in Table 1.1. Furthermore, I will illustrate in Subsection 1.4.2 how my framework allows to overcome the inclusion-exclusion problem and the corresponding results (summarized in Table 1.2) are stated there as well.

1.4.1. Main characterizations

The set of strategy-type pairs in which player i exhibits common belief of rationality is given by $CBR_i := \bigcap_{n\geq 0} CBR_{i,n}$, where each $CBR_{i,n}$ is defined recursively by setting:

$$CBR_{i,0} := S_i \times T_i,$$

$$CBR_{i,n} := CBR_{i,n-1} \cap B_i(\prod_{j\neq i} R_j \cap CBR_{j,n-1}),$$

for every $n \in \mathbb{N}$. That is, CBR_i brings together all the strategy-type pairs (s_i, t_i) where player i's type t_i beliefs that every player $j \neq i$ is rational and beliefs that every player $j \neq i$ beliefs that every player $k \neq j$ is rational, and so on.

In a similar vein, common belief of justifiability is defined by replacing R_i with J_i everywhere and the strategy-type pairs corresponding to it will be denoted by CBJ_i. If *belief* B_i is replaced with *full belief* F_i everywhere, one obtains common full belief of rationality (CFR_i) and common full belief of justifiability (CFJ_i), respectively. All these sets are actually events, because of the closedness of all the primitive events defined before (Lemma 1–4):

Corollary 1. *Let G be a game and $\mathcal{T}$ an ambiguous type structure. For any $n \in \mathbb{N}$, the following are events: $R_i \cap CBR_{i,n}$, $J_i \cap CBJ_{i,n}$, $R_i \cap CFR_{i,n}$, and $J_i \cap CFJ_{i,n}$. Furthermore, $R_i \cap CBR_{i,n}$. Furthermore, $R_i \cap CBR_i$, $J_i \cap CBJ_i$, $R_i \cap CFR_i$, and $J_i \cap CFJ_i$ are events.*

Proof. The first claim follows directly from the previous analysis (Lemma 1–4). The second claim holds because all four sets are intersections of closed sets, hence closed. ■

Based on the above, each of these four epistemic conditions has implications for behavior in a game:

Theorem 1 (Implications of common belief of rationality). *Let G be a game. The following holds:*

(*i*) *For any ambiguous type structure, any player i, and any strategy-type pair (s_i, t_i), if type t_i is consistent with common belief of rationality and s_i is rational for t_i, then s_i is rationalizable; i.e.,*

$$\operatorname*{proj}_{S_i}(R_i \cap CBR_i) \subseteq S_i^{\infty}.$$

(*ii*) *For any player i and any strategy s_i, if s_i is rationalizable then there exist a complete ambiguous type structure $\mathcal{T}$ and a type t_i consistent common belief of rationality for which s_i is rational; i.e.,*

$$S_i^{\infty} \subseteq \operatorname*{proj}_{S_i}(R_i \cap CBR_i).$$

Proof. Let's check first the finitely many iteration case, that is, that for each player i it holds that for any $n \geq 0$,

$$\operatorname*{proj}_{S_i}(R_i \cap CBR_{i,n}) \subseteq S_i^{n+1},$$

and that the inclusion is an equality when the ambiguous type structure is complete. For convenience, for each player i define $X_{i,0} := S_i \times T_i$ and for any $n \in \mathbb{N}$, $X_{i,n} := (R_i \cap CBR_{i,n-1})$. Now, I proceed by induction on n:

Initial Step ($n = 0$). For the right-hand inclusion, fix strategy-type pair $(\bar{s}_i, \bar{t}_i) \in R_i \cap CBR_{i,0}$ and denote $\bar{M}_i := M_i(\bar{t}_i)$. Then, for any belief $\mu_i^1 \in \bar{M}_i$, $\bar{s}_i$ is a best-reply for $\operatorname{marg}_{S_{-i}} \mu_i^1$. Thus, $\bar{\mu}_i^1 := \operatorname{marg}_{S_{-i}} \mu_i^1$ is a conjecture for which $\bar{s}_i$ is a best-reply. Hence, $s_i \in S_i^1$.

For the left-hand inclusion, fix strategy $\bar{s}_i \in S_i^1$ and conjecture $\bar{\mu}_i$ for which $\bar{s}_i$ is a best-reply. Pick then arbitrary $\eta_i \in \Delta(T_{-i})$ and define belief μ_i^1 as $\mu_i^1 := \bar{\mu}_i \times \eta_i$. Obviously, $\bar{s}_i$ is a best-reply to the marginal on S_{-i} induced by μ_i^1, which is precisely $\bar{\mu}_i$. Next, set $\bar{M}_i := \{\mu_i^1\}$ and pick (by completeness) type $\bar{t}_i \in T_i$ such that $M_i(\bar{t}_i) = \bar{M}_i$. The following two hold:

- $(\bar{s}_i, \bar{t}_i) \in CBR_{i,0}$. This holds trivially since $CBR_{i,0} = S_i \times T_i$.
- $(\bar{s}_i, \bar{t}_i) \in R_i$. This follows immediately from the fact that $\bar{s}_i$ is a best-reply to the conjecture induced by the unique belief in $M_i(\bar{t}_i) = \{\mu_i^1\}$, which is, precisely, $\bar{\mu}_i$.

Thus, I conclude that $(\bar{s}_i, \bar{t}_i)$ is strategy-type pair in $R_i \cap CBR_{i,0}$ that induces $\bar{s}_i$.

Inductive Step. Suppose that $n \geq 0$ is such that the claim holds. I verify that it also holds for $n + 1$. For the right-hand inclusion, fix strategy-type pair $(\bar{s}_i, \bar{t}_i) \in R_i \cap CBR_{i,n+1}$ and denote $\bar{M}_i := M_i(\bar{t}_i)$. Then, there exists some belief $\mu_i^{n+2} \in \bar{M}_i$ that puts probability one on $X_{-i,k} := \prod_{j \neq i} X_{j,k}$ for every $k = 0, \ldots, n+1$ and such that $\bar{s}_i$ is a best-reply for $\operatorname{marg}_{S_{-i}} \mu_i^{n+2}$. Thus, $\bar{\mu}_i^{n+2} := \operatorname{marg}_{S_{-i}} \mu_i^{n+2}$ is a conjecture that puts probability one on S_{-i}^k for every $k = 0, \ldots, n+1$ and for which $\bar{s}_i$ is a best-reply. Hence, $s_i \in S_i^{n+2}$.

For the left-hand inclusion, fix strategy $\bar{s}_i \in S_i^{n+2}$ and conjecture $\bar{\mu}_i$ that puts probability one on S_{-i}^{n+1} and for which $\bar{s}_i$ is a best-reply. Now, the induction hypothesis implies that for any $s_{-i} \in S_{-i}^{n+1}$ there exists some $t_{-i}(s_{-i}) \in T_{-i}$ such that $(s_{-i}, t_{-i}(s_{-i})) \in R_{-i} \cap CBR_{-i,n} = X_{-i,n+1}$. Then, for any measurable $E_{-i} \subseteq S_{-i} \times T_{-i}$ set:

$$\mu_i^{n+2}[E_{-i}] = \bar{\mu}_i \left[\left\{ s_{-i} \in S_{-i}^{n+1} \middle| (s_{-i}, t_{-i}(s_{-i})) \in E_{-i} \right\} \right].$$

Notice that finiteness of S_{-i} guarantees that μ_i^{n+2} is well-defined, and that the fact that $\bar{\mu}_i$ puts probability one on S_{-i}^{n+1} ensures that μ_i^{n+2} puts probability one on $X_{-i,n+1}$. Furthermore, it follows from monotonicity of the belief operator that μ_i^{n+2} puts probability one on $X_{-i,k}$ for every $k = 0, \ldots, n$. Obviously, $\bar{s}_i$ is a best-reply to the marginal on S_{-i} induced by μ_i^{n+2}, which is precisely $\bar{\mu}_i$. Next, set $\bar{M}_i = \{\mu_i^{n+2}\}$ and pick (using completeness) type $\bar{t}_i \in T_i$ such that $M_i(\bar{t}_i) = \bar{M}_i$. The following two hold:

- $(\bar{s}_i, \bar{t}_i) \in CBR_{i,n+1}$. To see it simply remember from above that μ_i^{n+2} puts probability one on $X_{-i,k}$ for every $k = 0, \ldots, n+1$.

- $(\bar{s}_i, \bar{t}_i) \in R_i$. This follows immediately from the fact that $\bar{s}_i$ is a best-reply to the conjecture induced by the unique belief in $M_i(\bar{t}_i) = \{\mu_i^{n+2}\}$, which is, precisely, $\bar{\mu}_i$.

Thus, I conclude that $(\bar{s}_i, \bar{t}_i)$ is strategy-type pair in $R_i \cap CBR_{i,n+1}$ that induces $\bar{s}_i$.

I prove next that, indeed:

$$\operatorname*{proj}_{S_i}(R_i \cap CBR_i) \subseteq S_i^\infty,$$

and that the inclusion is an equality when the ambiguous type structure is complete. For the right-hand inclusion fix strategy-type pair $(\bar{s}_i, \bar{t}_i) \in R_i \cap CBR_i$ and simply notice that since $(\bar{s}_i, \bar{t}_i) \in R_i \cap CBR_{i,n}$ for any $n \geq 0$. From above it follows that that $\bar{s}_i \in S_i^n$ for any $n \geq 1$. Thus, $\bar{s}_i \in S_i^\infty$.

For the left-hand inclusion, fix strategy $\bar{s}_i \in S_i^\infty$. Since, in particular, $\bar{s}_i \in S_i^{n+1}$ for any $n \geq 0$, the arguments above imply that for any $n \geq 0$ there exists some type $t_i^n \in T_i$ such that $(\bar{s}_i, t_i^n) \in R_i \cap CBR_{i,n}$. Now, let $\bar{M}_i$ denote the closure of the convex-hull of $\bigcup_{n\geq 0} M_i(t_i^n)$ and pick (using completeness) type $\bar{t}_i \in T_i$ such that $M_i(t_i) = \bar{M}_i$. Obviously, $\bar{s}_i$ is a best-reply is to every conjecture induced by the beliefs in $M_i(\bar{t}_i)$ and $\bar{t}_i$ is consistent with common belief in rationality. Thus, $(\bar{s}_i, \bar{t}_i) \in R_i \cap CBR_i$ and hence, $\bar{s}_i \in \operatorname{proj}_{S_i}(R_i \cap CBR_i)$. ■

Theorem 1 provides a complete characterization that generalizes the well-known result by Tan and da Costa Werlang (1988) from standard Bayesian rational players to ones that might display some Knightian uncertainty (or, formally equivalently, from players with complete preferences to ones that might have incomplete preferences due to multiple beliefs). Part (*i*) shows that whenever a player chooses maximizing w.r.t. possibly ambiguous higher-order beliefs that represent common belief in rationality, then the resulting strategy is necessarily rationalizable. Part (*ii*) shows the partial converse: it is not true that every time a rationalizable strategy is chosen this is due to the player being rational and best-responding to higher-order ambiguous beliefs that represent common belief in rationality, but still, it holds that every rationalizable strategy is a rational choice for some type that is consistent with common belief in rationality. Note that in this characterization no claim or implication is done about the presence of Knightian uncertainty: any rationalizable strategy could be played under either presence or absence of ambiguity. Indeed, parts of the construction for the proof of part (*ii*) resemble the construction in the Bayesian case. That is, I construct types with singleton sets of beliefs.

Theorem 2 (Implications of common belief of justifiability). *Let G be a game. The following holds:*

(*i*) *For any ambiguous type structure, any player i, and any strategy-type pair* (s_i, t_i), *if type* t_i *is consistent with common belief of justifiability and* s_i *is justifiable for* t_i, *then* s_i *is undominated; i.e.,*

$$\operatorname*{proj}_{S_i}(J_i \cap CBJ_i) \subseteq S_i^1.$$

(*ii*) *For any player i and any strategy* s_i, *if* s_i *is undominated then there exist a complete ambiguous type structure* $\mathcal{T}$ *and a type* t_i *consistent common belief of justifia1bility for which* s_i *is justifiable; i.e.,*

$$S_i^1 \subseteq \operatorname*{proj}_{S_i}(J_i \cap CBJ_i).$$

Proof. Let's check first the finitely many iteration case, that is, that for each player i it holds that for any $n \geq 0$,

$$\operatorname*{proj}_{S_i}(J_i \cap CBJ_{i,n}) \subseteq S_i^1,$$

and that the inclusion is an equality when the ambiguous type structure is complete. For convenience, for each player i define $X_{i,0} := S_i \times T_i$ and for any $n \in \mathbb{N}$, $X_{i,n} := (R_i \cap CBR_{i,n-1})$. Now, I proceed by induction on n:

Initial Step ($n = 0$).

For the right-hand inclusion, fix strategy-type pair $(\bar{s}_i, \bar{t}_i) \in J_i \cap CBJ_{i,0}$ and denote $\bar{M}_i := M_i(\bar{t}_i)$. Then, there exists a belief $\mu_i^1 \in \bar{M}_i$, $\bar{s}_i$ is a best-reply for $\operatorname{marg}_{S_{-i}} \mu_i^1$. Thus, $\bar{\mu}_i^1 := \operatorname{marg}_{S_{-i}} \mu_i^1$ is a conjecture for which $\bar{s}_i$ is a best-reply. Hence, $s_i \in S_i^1$.

For the left-hand inclusion, fix strategy $\bar{s}_i \in S_i^1$ and conjecture $\bar{\mu}_i$ for which $\bar{s}_i$ is a best-reply. Pick then arbitrary $\eta_i \in \Delta(T_{-i})$ and define belief μ_i^1 as $\mu_i^1 := \bar{\mu}_i \times \eta_i$. Obviously, $\bar{s}_i$ is a best-reply to the marginal on S_{-i} induced by μ_i^1, which is precisely $\bar{\mu}_i$. Next, set $\bar{M}_i := \{\mu_i^1\}$ and pick (by completeness) type $\bar{t}_i \in T_i$ such that $M_i(\bar{t}_i) = \bar{M}_i$. The following two hold:

- $(\bar{s}_i, \bar{t}_i) \in CBJ_{i,0}$. This holds trivially since $CBJ_{i,0} = S_i \times T_i$.
- $(\bar{s}_i, \bar{t}_i) \in J_i$. This follows immediately from the fact that $\bar{s}_i$ is a best-reply to the conjecture induced by the unique belief in $M_i(\bar{t}_i) = \{\mu_i^1\}$, which is, precisely, $\bar{\mu}_i$.

Thus, I conclude that $(\bar{s}_i, \bar{t}_i)$ is strategy-type pair in $J_i \cap CBJ_{i,0}$ that induces $\bar{s}_i$.

Inductive Step. Suppose that $n \geq 0$ is such that the claim holds. I verify that it also holds for $n+1$. For the right-hand inclusion, fix strategy-type pair $(\bar{s}_i, \bar{t}_i) \in J_i \cap CBJ_{i,n+1}$ and denote $\bar{M}_i := M_i(\bar{t}_i)$. Note that $J_i \cap CBJ_{i,n+1} \subseteq J_i \cap CBJ_{i,0}$. So by the initial step, $s_i \in S_i^1$.

For the left-hand inclusion, fix strategy $\bar{s}_i \in S_i^1$ and conjecture $\bar{\mu}_i$ for which $\bar{s}_i$ is a best-reply. Extend this conjecture to a belief μ_i^1 as in the initial step. Furthermore, consider any arbitrary belief μ_i^{n+1} that puts probability one on $J_{-i} \cap CBJ_{-i,n} = X_{-i,n+1}$ (and by monotonicity also on $X_{-i,k}$ for every $k = 0, \ldots, n+1$). Next, set $\bar{M}_i$ to be the convex hull of $\{\mu_i^1, \mu_i^{n+1}\}$ and pick (using completeness) type $\bar{t}_i \in T_i$ such that $M_i(\bar{t}_i) = \bar{M}_i$. The following two hold:

- $(\bar{s}_i, \bar{t}_i) \in CBJ_{i,n+1}$. To see it simply remember that from above it follows that μ_i^{n+1} puts probability one on $X_{-i,k}$ for every $k = 0, \ldots, n+1$.

- $(\bar{s}_i, \bar{t}_i) \in J_i$. This follows immediately from the fact that $\bar{s}_i$ is a best-reply to the conjecture induced by μ_i^1.

Thus, I conclude that $(\bar{s}_i, \bar{t}_i)$ is strategy-type pair in $J_i \cap CBJ_{i,n+1}$ that induces $\bar{s}_i$.

I prove next that, indeed:

$$\operatorname*{proj}_{S_i} (J_i \cap CBJ_i) \subseteq S_i^1,$$

and that the inclusion is an equality when the ambiguous type structure is complete. The right-hand inclusion follows because $J_i \cap CBJ_i \subseteq J_i \cap CBJ_{i,n}$ for any $n \geq 0$. For the left-hand inclusion, fix strategy $\bar{s}_i \in S_i^1$. The analysis above implies that for any $n \geq 0$ there exists some type $t_i^n \in T_i$ such that $(\bar{s}_i, t_i^n) \in J_i \cap CBJ_{i,n}$. Now, let $\bar{M}_i$ denote the closure of the convex-hull of $\bigcup_{n \geq 0} M_i(t_i^n)$ and pick (using completeness) type $\bar{t}_i \in T_i$ such that $M_i(t_i) = \bar{M}_i$.

Obviously, $\bar{s}_i$ is a best-reply is to at least one conjecture induced by the beliefs in $M_i(\bar{t}_i)$ and $\bar{t}_i$ is consistent with common belief in justifiability. Thus, $(\bar{s}_i, \bar{t}_i) \in R_i \cap CBR_i$ and hence, $\bar{s}_i \in \mathrm{proj}_{S_i}(J_i \cap CBJ_i)$. ■

Theorem 3 provides a complete characterization only of undominated strategies, which highlights that the combination of justfiability with belief do not impose severe restrictions about behavior in a game where players might display some Knightian uncertainty. As both notions are very weak, the higher-order restrictions do not have any bite. As before, part (i) shows unsurprisingly that whenever a player chooses justifiable w.r.t. possibly ambiguous higher-order beliefs that represent common belief in justifiability, then the resulting strategy is necessarily undominated. Part (ii) shows the partial converse: it is not true that every time an undominated strategy is chosen this is due to the player being justifiable and corresponding to higher-order ambiguous beliefs that represent common belief in justifiability, but still, it holds that every undominated strategy is a justifiable choice for some type that is consistent with common belief in justifiability. As before, in this characterization no claim or implication is done about the presence of Knightian uncertainty.

Theorem 3 (Implications of common full belief of rationality)**.** *Let G be a game. The following holds:*

(i) *For any ambiguous type structure, any player i, and any strategy-type pair* (s_i, t_i)*, if type* t_i *is consistent with common full belief of rationality and* s_i *is rational for* t_i*, then* s_i *is* *rationalizable; i.e.,*

$$\operatorname*{proj}_{S_i}(R_i \cap CFR_i) \subseteq S_i^\infty.$$

(*ii*) *For any player i and any strategy s_i, if s_i is rationalizable then there exist a complete ambiguous type structure $\mathcal{T}$ and a type t_i consistent common full belief of rationality for which s_i is rational; i.e.,*

$$S_i^\infty \subseteq \operatorname*{proj}_{S_i}(R_i \cap CFR_i).$$

Proof. Let's check first the finitely many iteration case, that is, that for each player i it holds that for any $n \geq 0$,

$$\operatorname*{proj}_{S_i}(R_i \cap CFR_{i,n}) \subseteq S_i^{n+1},$$

which holds because $CFR_{i,n} \subseteq CBR_{i,n}$ for any $n \geq 0$ and by (the proof of) Theorem 1. To prove that the inclusion is an equality when the ambiguous type structure is complete, it suffices to verify that the construction in the proof of Theorem 1 corresponds to *full* belief of rationality.

I prove next that, indeed $\operatorname{proj}_{S_i}(R_i \cap CFR_i) \subseteq S_i^\infty$, and that the inclusion is an equality when the ambiguous type structure is complete. The right-hand inclusion follows as above from Theorem 1 and monotonicity. For the left-hand inclusion, fix strategy $\bar{s}_i \in S_i^\infty$. Since, in particular, $\bar{s}_i \in S_i^{n+1}$ for any $n \geq 0$, the analysis above implies that for any $n \geq 0$ there exists some type $t_i^n \in T_i$ such that $(\bar{s}_i, t_i^n) \in R_i \cap CFR_{i,n}$. Let $Z_n := \{(s_i, t_i) : s_i = \bar{s}_i\} \cap R_i \cap CFR_{i,n}$ which is non-empty and closed by Corollary 1. By construction, Z_n is a decreasing sequence and has the finite intersection property. Since the type structure is assumed to be compact, it follows that $Z := \cap_n Z_n$ is non-empty, i.e. there exists $\bar{t}_i$ such that $(\bar{s}_i, \bar{t}_i) \in Z \subseteq R_i \cap CFR_i$. Hence, $\bar{s}_i \in \mathrm{proj}_{S_i}(R_i \cap CFR_i)$. ■

The interpretation of Theorem 3 is similar to Theorem 1 stated above. Indeed, the construction for part (*ii*) of the theorem is actually the same as in the respective part of Theorem 1, which—as mentioned above–resembles the Bayesian construction and there *belief* and *full belief* collapse to the same notion.

Theorem 4 (Implications of common full belief of justifiability)**.** *Let G be a game. The following holds:*

(*i*) *For any ambiguous type structure, any player i, and any strategy-type pair* (s_i, t_i)*, if type* t_i *is consistent with common full belief of justifiability and* s_i *is justifiable for* t_i*, then* s_i *is rationalizable; i.e.,*

$$\mathrm{proj}_{S_i}(J_i \cap CBJ_i) \subseteq S_i^\infty.$$

(*ii*) *For any player i and any strategy* s_i*, if* s_i *is rationalizable then there exist a complete ambiguous type structure* $\mathcal{T}$ *and a type* t_i *consistent common full belief of justifiability for which* s_i *is justifiable; i.e.,*

$$S_i^\infty \subseteq \mathrm{proj}_{S_i}(J_i \cap CBJ_i).$$

Proof. Let's check first the finitely many iteration case, that is, that for each player *i* it holds that for any $n \geq 0$,

$$\mathrm{proj}_{S_i}(J_i \cap CFJ_{i,n}) \subseteq S_i^{n+1},$$

and that the inclusion is an equality when the ambiguous type structure is complete. For convenience, for each player i define $X_{i,0} := S_i \times T_i$ and for any $n \in \mathbb{N}$, $X_{i,n} := J_i \cap CFJ_{i,n-1}$. Now, I proceed by induction on n:

Initial Step ($n = 0$). Same as in the proof of Theorem 2.

Inductive Step. Suppose that $n \geq 0$ is such that the claim holds. I verify that it also holds for $n+1$. For the right-hand inclusion, fix strategy-type pair $(\bar{s}_i, \bar{t}_i) \in J_i \cap CFJ_{i,n+1}$ and denote $\bar{M}_i := M_i(\bar{t}_i)$. Then, there exists some belief (in fact, that's true for all beliefs) $\mu_i^{n+2} \in \bar{M}_i$ that puts probability one on $X_{-i,k} := \prod_{j \neq i} X_{j,k}$ for every $k = 0, \ldots, n+1$ and such that $\bar{s}_i$ is a best-reply for $\mathrm{marg}_{S_{-i}} \mu_i^{n+2}$. Thus, $\bar{\mu}_i^{n+2} := \mathrm{marg}_{S_{-i}} \mu_i^{n+2}$ is a conjecture that puts probability one on S_{-i}^k for every $k = 0, \ldots, n+1$ and for which $\bar{s}_i$ is a best-reply. Hence, $s_i \in S_i^{n+2}$.

For the left-hand inclusion, fix strategy $\bar{s}_i \in S_i^{n+2}$ and conjecture $\bar{\mu}_i$ that puts probability one on S_{-i}^{n+1} and for which $\bar{s}_i$ is a best-reply. Now, the induction hypothesis implies that for any $s_{-i} \in S_{-i}^{n+1}$ there exists some $t_{-i}(s_{-i}) \in T_{-i}$ such that $(s_{-i}, t_{-i}(s_{-i})) \in J_{-i} \cap CFJ_{-i,n} = X_{-i,n+1}$. Then, for any measurable $E_{-i} \subseteq S_{-i} \times T_{-i}$ set:

$$\mu_i^{n+2}[E_{-i}] = \bar{\mu}_i \left[\left\{ s_{-i} \in S_{-i}^{n+1} \,\middle|\, (s_{-i}, t_{-i}(s_{-i})) \in E_{-i} \right\} \right].$$

Notice that finiteness of S_{-i} guarantees that μ_i^{n+2} is well-defined, and that the fact that $\bar{\mu}_i$ puts probability one on S_{-i}^{n+1} ensures that μ_i^{n+2} puts probability one on $X_{-i,n+1}$. Furthermore, it follows from monotonicity of the belief operator that μ_i^{n+2} puts probability one on $X_{-i,k}$ for every $k = 0, \ldots, n$. Obviously, $\bar{s}_i$ is a best-reply to the marginal on S_{-i} induced by μ_i^{n+2}, which is precisely $\bar{\mu}_i$. Next, set $\bar{M}_i = \{\mu_i^{n+2}\}$ and pick (using completeness) type $\bar{t}_i \in T_i$ such that $M_i(\bar{t}_i) = \bar{M}_i$. The following two hold:

- $(\bar{s}_i, \bar{t}_i) \in CFJ_{i,n+1}$. To see it simply remember that the arguments above imply that μ_i^{n+2} puts probability one on $X_{-i,k}$ for every $k = 0, \dots, n+1$.
- $(\bar{s}_i, \bar{t}_i) \in R_i$. This follows immediately from the fact that $\bar{s}_i$ is a best-reply to the conjecture induced by the unique belief in $M_i(\bar{t}_i) = \{\mu_i^{n+2}\}$, which is, precisely, $\bar{\mu}_i$.

Thus, I conclude that $(\bar{s}_i, \bar{t}_i)$ is strategy-type pair in $J_i \cap CFJ_{i,n+1}$ that induces $\bar{s}_i$.

I prove next that, indeed:

$$\operatorname*{proj}_{S_i} (J_i \cap CFJ_i) \subseteq S_i^\infty,$$

and that the inclusion is an equality when the ambiguous type structure is complete. For the right-hand inclusion fix strategy-type pair $(\bar{s}_i, \bar{t}_i) \in R_i \cap CBR_i$ and simply notice that since $(\bar{s}_i, \bar{t}_i) \in J_i \cap CFJ_{i,n}$ for any $n \geq 0$, the arguments above imply that $\bar{s}_i \in S_i^n$ for any $n \geq 1$. Thus, $\bar{s}_i \in S_i^\infty$.

For the left-hand inclusion, fix strategy $\bar{s}_i \in S_i^\infty$. Since, in particular, $\bar{s}_i \in S_i^{n+1}$ for any $n \geq 0$, the analysis above implies that for any $n \geq 0$ there exists some type $t_i^n \in T_i$ such that $(\bar{s}_i, t_i^n) \in J_i \cap CFJ_{i,n} = X_{i,n+1}$. Let $Z_n := \{(s_i, t_i) : s_i = \bar{s}_i\} \cap X_{i,n+1}$ which is non-empty and closed by Corollary 1. By construction, Z_n is a decreasing sequence and has the finite intersection property. Since the type structure is assumed to be compact, it follows that $Z := \cap_n Z_n$ is non-empty, i.e. there exists $\bar{t}_i$ such that $(\bar{s}_i, \bar{t}_i) \in Z \subseteq J_i \cap CFJ_i$. Hence, $\bar{s}_i \in \operatorname{proj}_{S_i}(J_i \cap CFJ_i)$. ■

Theorem 4 has again a similar interpretation as the previous ones. However, in contrast to Theorem 2 replacing *belief* with *full belief* imposes severe restrictions on higher-order reasoning so that the characterization corresponds to rationalizability instead of just undominated strategies. As before, the characterization does not rely on the presence (nor on the absence) of Knightian uncertainty itself.

1.4.2. Knightian Uncertainty, Cautiousness, and the Inclusion-Exclusion Problem

The framework of Knightian uncertainty is more flexible than the standard Bayesian one, because the latter is nested in the former. Therefore, the characterization results Subsection 1.4.1 do not explicitly require nor rule-out the Bayesian case. However, as mentioned in the introduction, one possiblity that arises from having multiple beliefs is the resolution of the of inclusion-exclusion problem of Samuelson (1992). That is, cautious behavior and strategic sophistication in a game together are not possible within a Bayesian framework.

I argue next that Knightian uncertainty, intuitively thought of the decision maker considering every state of the world when deciding which choice is best, can be interpreted as a product of ambiguity in the sense that types that exhibit cautiousness tend to represent preferences that also display Knightian uncertainty. First, I formalize the two notions of cautiousness that take part in the characterizations result and discuss their relation with presence of Knightian uncertainty.

Definition 1 (Weak and strong cautiousness)**.** *Let G be a game and $\mathcal{T}$ an ambiguous type structure. Then, for any player i and any type t_i I say that type t_i is:*

(*i*) weakly cautious *if it is has a belief with support $S_{-i} \times T_{-i}$. That is, there exists $\mu_i \in M_i(t_i)$ such that* $\operatorname{supp} \mu_i = S_{-i} \times T_{-i}$. *I denote the set of player i's strategy-type pairs in which the type is weakly cautious by WC_i.*

(*ii*) strongly cautious *if $M_i(t_i)$ has non-empty topological interior. I denote the set of player i's strategy-type pairs in which the type is strongly cautious by SC_i.*

If at an intuitive level cautiousness is associated with the idea that a decision maker takes every possible contingency into account, then this is present in both concepts in Definition 1. Weak cautiousness requires that the whole set of states $S_{-i} \times T_{-i}$ is believed and nothing is ruled out, that is, loosely speaking, that every state is taken into account by the decision maker. Strong cautiousness is of course more demanding: if a type is maximizing w.r.t. to some belief in $\Delta(S_{-i} \times T_{-i})$, her choice must be also optimal for every belief resulting from some small *enough* perturbation of the original belief. Thus, intuitively, not only is every state taken into account, the possibility of small increases in the likeliness of each state also is.

The relation of these notions of cautiousness with Knightian uncertainty is easy to see. In principle, it is possible that a type displays weak cautiousness but not Knightian uncertainty. This is the case of every type whose set of ambiguous beliefs consists of a single belief with full-support on $S_{-i} \times T_{-i}$. However, if in addition to weak cautiousness the type also exhibits some form of strategic sophistication in the sense of believing in some proper subset of $S_{-i} \times T_{-i}$, then, necessarily, the type displays Knightian uncertainty: the corresponding ambiguous beliefs *a fortiori* contain at least two different beliefs. Strong cautiousness necessarily implies Knightian uncertainty regardless of whether some proper event is believed or not. This is due to no open set consisting of just a singleton belief. Hence, in either case, not only does the introduction of Knightian uncertainty allow for making strategic reasoning and cautiousness compatible, but indeed, it is necessary when strategic reasoning has any bite. This allows to overcome the inclusion-exclusion problem of Samuelson (1992).

I continue now with two solution concepts that, intuitively, do contain some flavor of cautiousness: the Dekel-Fudenberg procedure and strict rationalizability. The definition of the former requires that the strategies that survive the first iterated round are best-replies to some belief that has full-support on the opponents' strategies, that is, some belief that gives some consideration to every possible behavior. The link with the notion of weak cautiousness seems clear. Similarly, the definition of strict rationalizability requires that non-strict best-replies are eliminated in the first round; this is equivalent to requiring that the best-replies of the player remain unaltered under any possible small enough tremble in the belief in any direction (i.e., by adding some small enough mass to any strategy by the opponent). Again, it seems natural how this idea could relate to the notion of strong cautiousness, which requires ambiguous beliefs to consider perturbations that may increase the likeliness of every state. Indeed, as shown below in Theorem 5 and Theorem 6, providing foundations for the Dekel-Fudenberg procedure and strict rationalizability only requires to incorporate the aforementioned notions of cautiousness in the common belief constraints. To this end, consider the following collections of strategy-type pairs:

- The set of strategy-type pairs in which player i exhibits common belief in rationality and weak cautiousness is given by $CBRWC_i := \bigcap_{n\geq 0} CBRWC_{i,n}$, where each $CBRWC_{i,n}$ is defined recursively by setting:

$$CBRWC_{i,0} := S_i \times T_i,$$

$$CBRWC_{i,n} := CBRWC_{i,n-1} \cap B_i(\bigcap_{j\neq i} R_j \cap WC_j \cap CBRWC_{j,n-1}),$$

  for every $n \in \mathbb{N}$. The interpretation of $CBRWC_i$ is clear: it is the collection of the strategy-type pairs (s_i, t_i) where player i's type t_i believes that every player $j \neq i$ is rational and weakly cautious, believes that every player $j \neq i$ believes that every player $k \neq j$ is rational and weakly cautious, and so on.

- The set of strategy-type pairs in which player i exhibits common belief in rationality and strong cautiousness is given by $CBRSC_i := \bigcap_{n\geq 0} CBRSC_{i,n}$, where each $CBRSC_{i,n}$ is defined recursively by setting:

$$CBRSC_{i,0} := S_i \times T_i,$$

$$CBRSC_{i,n} := CBRSC_{i,n-1} \cap B_i(\bigcap_{j\neq i} R_j \cap SC_j \cap CBRSC_{j,n-1}),$$

for every $n \in \mathbb{N}$. Again, it is clear that $CBRSC_i$ is the collection of the strategy-type pairs (s_i, t_i) where player i's type t_i believes that every player $j \neq i$ is rational and strongly cautious, believes that every player $j \neq i$ believes that every player $k \neq j$ is strongly rational and strongly cautious, and so on.

Now, I am ready to provide foundations for the Dekel-Fudenberg procedure and strict rationalizability:

Theorem 5 (Foundation of the Dekel-Fudenberg procedure)**.** *Let G be a game. It holds that:*

(i) *For any ambiguous type structure, any player i, and any strategy-type pair* (s_i, t_i)*, if type* t_i *is consistent weak cautiousness and common belief of rationality and weak cautiousness and* s_i *is rational for* t_i*, then* s_i *survives the Dekel-Fudenberg procedure; i.e.,*

$$\operatorname*{proj}_{S_i}(R_i \cap WC_i \cap CBRWC_i) \subseteq S^{\infty}W_i.$$

(ii) *For any player i and any strategy* s_i*, if* s_i *survives the Dekel-Fudenberg procedure then there exist a complete ambiguous type structure* $\mathcal{T}$ *and a type* t_i *consistent with weak cautiousness and common belief in rationality and weak cautiousness for which* s_i *is rational; i.e.*

$$S^{\infty}W_i \subseteq \operatorname*{proj}_{S_i}(R_i \cap WC_i \cap CBRWC_i).$$

Proof. Let's check first the finitely many iteration case, that is, that for each player i it holds that for any $n \geq 0$,

$$\operatorname*{proj}_{S_i}(R_i \cap WC_i \cap CBRWC_{i,n}) \subseteq S^n W_i,$$

and that the inclusion is an equality when the ambiguous type structure is complete. For convenience, for each player i define $X_{i,0} := S_i \times T_i$ and for any $n \in \mathbb{N}$, $X_{i,n} := R_i \cap WC_i \cap CBRWC_{i,n-1}$. Now, I proceed by induction on n:

Initial Step ($n = 0$). For the right-hand inclusion, fix strategy-type pair $(\bar{s}_i, \bar{t}_i) \in R_i \cap WC_i$ and denote $\bar{M}_i = M_i(\bar{t}_i)$. Then, since $\bar{t}_i$ is weakly cautious there exists some belief $\mu_i^1 \in \bar{M}_i$ whose support is $S_{-i} \times T_{-i}$, and since $(\bar{s}_i, \bar{t}_i) \in R_i$ it holds that $\bar{s}_i$ is a best-reply for $\operatorname{marg}_{S_{-i}} \mu_i^1$. Thus, $\bar{\mu}_i^1 := \operatorname{marg}_{S_{-i}} \mu_i^1$ is a conjecture in with full-support on S_{-i} for which $\bar{s}_i$ is a best-reply. Hence, $\bar{s}_i \in S^0 W_i$.

For the left-hand inclusion, fix strategy $\bar{s}_i \in S^0 W_i$ and conjecture $\bar{\mu}_i$ with full-support on S_{-i} for which $\bar{s}_i$ is a best-reply. Then, take arbitrary full-support belief $\eta_i \in \Delta(T_{-i})$ and set $\mu_i^1 := \bar{\mu}_i \times \eta_i$ and $\bar{M}_i := \{\mu_i^1\}$. Since $\mathcal{T}$ is complete, there exists some type $\bar{t}_i \in T_i$ such that $M_i(\bar{t}_i) = \bar{M}_i$. As μ_i^1 has full-support in $S_{-i} \times T_{-i}$ it has to be the case that $\bar{t}_i$ is weakly cautious, and hence, that $(\bar{s}_i, \bar{t}_i) \in WC_i$, and the fact that $\bar{s}_i$ is a best-reply to the marginal on S_{-i} induced by the unique belief in $M_i(\bar{t}_i)$ ensures that $(\bar{s}_i, \bar{t}_i) \in R_i$. Thus, I conclude that $(\bar{s}_i, \bar{t}_i)$ is a strategy-type pair in $R_i \cap WC_i$ that induces $\bar{s}_i$.

Inductive Step. Suppose that $n \geq 0$ is such that the claim holds. Next I verify it for $n+1$. For the right-hand inclusion, fix strategy-type pair $(\bar{s}_i, \bar{t}_i) \in R_i \cap WC_i \cap CBRWC_{i,n+1}$ and denote

$\bar{M}_i := M_i(\bar{t}_i)$. Then, since $(\bar{s}_i, \bar{t}_i) \in R_i \cap WC_i \cap CBRWC_{i,n}$ the induction hypothesis implies that $\bar{s}_i \in S^n W_i$, and since $(\bar{s}_i, \bar{t}_i) \in CBRWC_{i,n+1}$, that there exists some belief $\mu_i^{n+1} \in M_i(\bar{t}_i)$ which assign probability 1 to $X_{-i,n+1} := \prod_{j \neq i} X_{j,n+1}$. Now, I also know that $(\bar{s}_i, \bar{t}_i) \in R_i$ and thus, that $\bar{s}_i$ is a best reply to the marginal on S_{-i} induced by μ_i^{n+1}, which by the induction hypothesis puts probability one on $S^n W_{-i}$. Hence, I conclude that $\bar{s}_i \in S^{n+1} W_i$.

For the left-hand inclusion, fix strategy $\bar{s}_i \in S^{n+1} W_i$ and pair of conjectures $\bar{\mu}_i^1$ and $\bar{\mu}_i^{n+1}$ such that: (*i*) $\bar{\mu}_i^1$ has full-support on S_{-i}, (*ii*) $\bar{\mu}_i^{n+1}$ puts probability one on $S^n W_{-i}$ (and hence, so does it on $S^k W_{-i}$ for every $k = 1, \dots, n$) and (*iii*) $\bar{s}_i$ is a best-reply to both conjectures. Use these two conjectures now to define:

- Belief μ_i^1 with full support in $S_{-i} \times T_{-i}$. Just pick arbitrary belief full-support belief $\eta_i^1 \in \Delta(T_{-i})$ and set $\mu_1^1 := \bar{\mu}_i^1 \times \eta_i^1$.

- Belief μ_i^{n+1} that puts probability one on $X_{-i,n+1}$. To do so simply notice that from the induction hypothesis for any $s_{-i} \in S^n W_{-i}$ there exists some types $t_{-i}(s_{-i})$ such that $(s_{-i}, t_{-i}(s_{-i})) \in X_{-i,n+1}$. Then, for any measurable $E_{-i} \subseteq S_{-i} \times T_{-i}$ set:

$$\mu_i^{n+1}[E_{-i}] = \bar{\mu}_i^{n+1} \left[\{ s_{-i} \in S^n W_{-i} | \, (s_{-i}, t_{-i}(s_{-i})) \in E_{-i} \} \right].$$

  Notice that finiteness of S_{-i} guarantees that μ_i^{n+1} is well-defined, and that the fact that $\bar{\mu}_i^{n+1}$ puts probability one on $S^n W_{-i}$ ensures that μ_i^{n+1} puts probability one on $X_{-i,n+1}$. Furthermore, it follows from monotonicity of the belief operator that μ_i^{n+1} puts probability one on $X_{-i,k}$ for every $k = 0, \dots, n$.

Notice in addition that since for both $k = 1, n+1$, the marginal of μ_i^k on S_{-i} is precisely $\bar{\mu}_i^k$. Thus, $\bar{s}_i$ is a best-reply to μ_i^k. Then, let $\bar{M}_i$ denote the convex hull of $\{\mu_i^1, \mu_i^{n+1}\}$ and pick (by completeness) type $\bar{t}_i \in T_i$ such that $M_i(\bar{t}_i) = \bar{M}_i$. The following three hold:

- $(\bar{s}_i, \bar{t}_i) \in CBRCW_{i,k}$ for any $k = 0, \dots, n+1$. To see it, simply notice that as seen above, μ_i^{n+1}, which is an element of $M_i(\bar{t}_i)$, puts probability one on every $X_{i,k}$.
- $(\bar{s}_i, \bar{t}_i) \in WC_i$. This is implied by μ_i^1, which is an element of $M_i(\bar{t}_i)$ that has full-support on $S_{-i} \times T_{-i}$.
- $(\bar{s}_i, \bar{t}_i) \in R_i$. This follows immediately from the fact that $\bar{s}_i$ is a best-reply to the conjectures induced by both μ_i^1 and μ_i^{n+1}, and hence, to the conjecture induced by each belief in $M_i(\bar{t}_i)$.

Thus, I conclude that $(\bar{s}_i, \bar{t}_i)$ is strategy-type pair in $R_i \cap WC_i \cap CBRWC_{i,n+1}$ that induces $\bar{s}_i$.

I prove next that, indeed:

$$\operatorname*{proj}_{S_i} (R_i \cap WC_i \cap CBRWC_i) \subseteq S^\infty W_i,$$

and that the inclusion is an equality when the ambiguous type structure is complete. For the right-hand inclusion fix strategy-type pair $(\bar{s}_i, \bar{t}_i) \in R_i \cap WC_i \cap CBRWC_i$ and simply notice that since $(\bar{s}_i, \bar{t}_i) \in R_i \cap WC_i \cap CBRWC_{i,n}$ for any $n \geq 0$, the arguments above imply that $\bar{s}_i \in S^n W_i$ for any $n \geq 0$. Thus, $\bar{s}_i \in S^\infty W_i$.

For the left-hand inclusion, fix strategy $\bar{s}_i \in S^\infty W_i$. Since, in particular, $\bar{s}_i \in S^n W_i$ for any $n \geq 0$, the arguments above imply that for any $n \geq 0$ there exists some type $t_i^n \in T_i$

such that $(\bar{s}_i, t_i^n) \in R_i \cap WC_i \cap CBRWC_{i,n}$. Now, let $\bar{M}_i$ denote the closure of the convex-hull of $\bigcup_{n\geq 0} M_i(t_i^n)$ and (using completeness) pick type $\bar{t}_i \in T_i$ such that $M_i(t_i) = \bar{M}_i$. Obviously, $\bar{s}_i$ is a best-reply is to every conjecture induced by the beliefs in $M_i(\bar{t}_i)$ and $\bar{t}_i$ is weakly cautious and is consistent with common belief in rationality and weak cautiousness. Thus, $(\bar{s}_i, \bar{t}_i) \in R_i \cap WC_i \cap CBRWC_i$ and hence, $\bar{s}_i \in \text{proj}_{S_i}(R_i \cap WC_i \cap CBRWC_i)$.

■

Theorem 6 (Foundation of strict rationalizability). *Let G be a game. It holds that:*

(*i*) *For any ambiguous type structure, any player i, and any strategy-type pair (s_i, t_i), if type t_i is consistent with strong cautiousness and common belief of rationality and strong cautiousness and s_i is rational for t_i, then s_i is strictly rationalizable; i.e.,*

$$\underset{S_i}{\text{proj}}(R_i \cap SC_i \cap CBRSC_i) \subseteq S^\infty S_i^+.$$

(*ii*) *For any player i and any strategy s_i, if s_i is strictly rationalizable then there exist a complete ambiguous type structure $\mathcal{T}$ and a type t_i consistent with strong cautiousness and common belief of rationality and strong cautiousness for which s_i is rational; i.e.,*

$$S^\infty S_i^+ \subseteq \underset{S_i}{\text{proj}}(R_i \cap SC_i \cap CBRSC_i).$$

For the poof define for any $A_{-i} \subseteq \prod_{j \neq i} S_j$ and any $s_i \in S_i$

$$[s_i]_{A_{-i}} := \left\{ s_i' \in S_i \;\middle|\; \begin{array}{l} \text{For any } s_{-i} \in A_{-i}, \\ \\ u_i(s_{-i}; s_i') = u_i(s_{-i}; s_i) \end{array} \right\}.$$

Then the first step of strict rationalizability can be rewritten as

$$S^0 S_i^+ := \left\{ s_i \in S_i \;\middle|\; \begin{array}{l} \text{There exists some } \mu_i \in \Delta(S_{-i}) \text{ such that:} \\ \\ (i) \quad \operatorname{supp} \mu_i \subseteq \prod_{j \neq i} S_j, \\ \\ (ii) \quad BR_i(\mu_i) = [s_i]_{\prod_{j \neq i} S_j} \end{array} \right\}.$$

The proof of Theorem 6 makes use of the following lemma.

Lemma 5. *Let $\varepsilon > 0$ and G be a game. Consider player i's conjecture μ_i with full support on S_{-i}. If $s_i \in BR_i(\mu_i')$ for every $\mu_i' \in B_\varepsilon(\mu_i)$,[28] then $BR_i(\mu_i) = [s_i]_{S_{-i}}$.*

Proof. Fix $\varepsilon > 0$ and $s_i \in S_i$ such that $s_i \in BR_i(\mu_i')$ for every $\mu_i' \in B_\varepsilon(\mu_i)$. Consider $s_i' \in S_i$ such that $s_i' \neq s_i$ and $s_i' \in BR_i(\mu_i)$. First, s_i' cannot be weakly dominated by s_i since μ_i has full support. So there are two cases:

1. s_i' is payoff equivalent to s_i (on S_{-i}), i.e. $s_i' \in [s_i]_{S_{-i}}$, or

2. there exists $s_{-i}' \in S_{-i}$ such that $u_i(s_{-i}'; s_i') > u_i(s_{-i}'; s_i)$.

[28] I.e., in the open ball with center μ_i and radius ε.

In the first case there is nothing to prove. The second case would imply

$$s_i \notin BR_i\Big((1-\delta)\mu_i + \delta s'_{-i}\Big)$$

for $\delta < \varepsilon$. This violates $s_i \in BR_i(\mu'_i)$ for every $\mu'_i \in B_\varepsilon(\mu_i)$. Thus, the second case is impossible. ■

Proof of Theorem 6. Let's check first the finitely many iteration case, that is, that for each player i it holds that for any $n \geq 0$,

$$\operatorname*{proj}_{S_i}(R_i \cap SC_i \cap CBRSC_{i,n}) \subseteq S^n S_i^+,$$

and that the inclusion is an equality when the ambiguous type structure is complete. For convenience, for each player i define $X_{i,0} := S_i \times T_i$ and for any $n \in \mathbb{N}$, $X_{i,n} := R_i \cap SC_i \cap CBRSC_{i,n-1}$. Now, I proceed by induction on n:

Initial Step ($n = 0$). For the right-hand inclusion, fix strategy-type pair $(\bar{s}_i, \bar{t}_i) \in R_i \cap SC_i \cap CBRSC_{i,0}$ and denote $\bar{M}_i := M_i(\bar{t}_i)$. Pick arbitrary full-supported belief μ_i in the interior of $\bar{M}_i$,[29] and set $\bar{\mu}_i^1 := \operatorname{marg}_{S_{-i}} \mu_i$. Then, there exists some $\varepsilon > 0$ such that $\bar{s}_i \in BR_i(\mu'_i)$ for every $\mu'_i \in B_\varepsilon(\bar{\mu}_i^1)$. By Lemma 5 this implies $BR_i(\bar{\mu}_i^1) = [\bar{s}_i]_{S_{-i}}$. Thus, $\bar{\mu}_i^1$ is a conjecture whose best-reply is exactly $[\bar{s}_i]_{S_{-i}}$. Hence, $\bar{s}_i \in S^0 S_i^+$.

For the left-hand inclusion, fix strategy $\bar{s}_i \in S^0 S_i^+$ and conjecture $\bar{\mu}_i$ whose best reply is exactly $[s_i]_{S_{-i}}$. It follows from finiteness of S_i and from continuity of the expected utility

[29] Since $\bar{M}_i$ has non-empty interior, this is obviously possible.

payoff function that there exists some $\varepsilon > 0$ such that $BR_i(\mu_i) = [s_i]_{S_{-i}}$ for any $\mu_i \in \overline{B}_\varepsilon(\bar{\mu}_i)$.[30] Define then:

$$\bar{M}_i := \overline{B}_\varepsilon(\bar{\mu}_i) \times \Delta(T_{-i}),$$

which is clearly closed and convex and (using completeness) pick type $\bar{t}_i \in T_i$ such that $M_i(\bar{t}_i) = \bar{M}_i$. The following three hold:

- $(\bar{s}_i, \bar{t}_i) \in CBRSC_{i,0}$. This is trivially true due to the fact that $CBRSC_{i,0} = S_{-i} \times T_{-i}$.
- $(\bar{s}_i, \bar{t}_i) \in SC_i$. It is obvious by construction that $M_i(\bar{t}_i)$ has non-empty interior.
- $(\bar{s}_i, \bar{t}_i) \in R_i$. This follows immediately from the fact that the conjecture induced by each belief in $M_i(\bar{t}_i)$ is a convex combination of conjectures in $\overline{B}_\varepsilon(\bar{\mu}_i)$.

Thus, I conclude that $(\bar{s}_i, \bar{t}_i)$ is strategy-type pair in $R_i \cap SC_i \cap CBRSC_{i,0}$ that induces $\bar{s}_i$.

Inductive Step. Suppose that $n \geq 0$ is such that the claim holds. Next I verify it holds for $n + 1$. For the right-hand inclusion, fix strategy-type pair $(\bar{s}_i, \bar{t}_i) \in R_i \cap SC_i \cap CBRSC_{i,n+1}$ and denote $\bar{M}_i := M_i(\bar{t}_i)$. Then, since $(\bar{s}_i, \bar{t}_i) \in R_i \cap SC_i \cap CBRSC_{i,n}$ the induction hypothesis implies that $\bar{s}_i \in S^n S_i^+$, and since $(\bar{s}_i, \bar{t}_i) \in CBRSC_{i,n+1}$, that there exists some belief $\mu_i^{n+1} \in M_i(\bar{t}_i)$ which assign probability 1 to $X_{-i,n+1} := \prod_{j \neq i} X_{j,n+1}$. Furthermore, $(\bar{s}_i, \bar{t}_i) \in R_i$ and thus, that $\bar{s}_i$ is a best reply to the marginal on S_{-i} induced by μ_i^{n+1}, which by the induction hypothesis puts probability one on $S^n S_{-i}^+$. Hence, I conclude that $\bar{s}_i \in S^{n+1} S_i^+$.

For the left-hand inclusion, fix strategy $\bar{s}_i \in S^{n+1} S_i^+$ and pair of conjectures $\bar{\mu}_i^1$ and $\bar{\mu}_i^{n+1}$ such that: *(i)* $\bar{\mu}_i^1$ is as $\bar{\mu}_i$ in the initial step, *(ii)* $\bar{\mu}_i^{n+1}$ puts probability one on $S^n S_{-i}^+$ (and hence,

[30] I.e., in the closed ball with center $\bar{\mu}_i^1$ and radius ε.

so does it on $S^k S^+_{-i}$ for every $k = 1, \dots, n$) and (*iii*) $\bar{s}_i$ is a best-reply to both conjectures. Now, use these two conjectures to define:

- A set of beliefs $\bar{M}^1_i$ like $\bar{M}_i$ in the initial step.
- Belief μ^{n+1}_i that puts probability one on $X_{-i,n+1}$. To do so simply notice the induction hypothesis implies for any $s_{-i} \in S^n S^+_{-i}$ there exists some types $t_{-i}(s_{-i})$ such that $(s_{-i}, t_{-i}(s_{-i})) \in X_{-i,n+1}$. Then, for any measurable $E_{-i} \subseteq S_{-i} \times T_{-i}$ set:

$$\mu^{n+1}_i[E_{-i}] = \bar{\mu}^{n+1}_i \left[\{ s_{-i} \in S^n W_{-i} | \, (s_{-i}, t_{-i}(s_{-i})) \in E_{-i} \} \right].$$

  Notice that finiteness of S_{-i} guarantees that μ^{n+1}_i is well-defined, and that the fact that $\bar{\mu}^{n+1}_i$ puts probability one on $S^n S^+_{-i}$ ensures that μ^{n+1}_i puts probability one on $X_{-i,n+1}$. Furthermore, it follows from monotonicity of the belief operator that μ^{n+1}_i puts probability one on $X_{-i,k}$ for every $k = 0, \dots, n$.

Notice that $\bar{s}_i$ is a best-reply to all conjectures derived from $\bar{M}^1_i \cup \{\mu^{n+1}_i\}$. Then, let $\bar{M}_i$ denote the convex hull of $\bar{M}^1_i \cup \{\mu^{n+1}_i\}$ and pick (by completeness) type $\bar{t}_i \in T_i$ such that $M_i(\bar{t}_i) = \bar{M}_i$. The following three hold:

- $(\bar{s}_i, \bar{t}_i) \in CBRSC_{i,k}$ for any $k = 0, \dots, n+1$. To see it, simply notice that as seen above, μ^{n+1}_i, which is an element of $M_i(\bar{t}_i)$, puts probability one on every $X_{i,k}$.
- $(\bar{s}_i, \bar{t}_i) \in SC_i$. This is implied by $\bar{M}^1_i$, which is a subset of $M_i(\bar{t}_i)$ and has non-empty interior.

- $(\bar{s}_i, \bar{t}_i) \in R_i$. This follows immediately from the fact that $\bar{s}_i$ is a best-reply to the conjectures induced by all conjectures derived from $M_i(\bar{t}_i)$.

Thus, I conclude that $(\bar{s}_i, \bar{t}_i)$ is strategy-type pair in $R_i \cap SC_i \cap CBRSC_{i,n+1}$ that induces $\bar{s}_i$.

I prove next that, indeed:

$$\operatorname*{proj}_{S_i} (R_i \cap SC_i \cap CBRSC_i) \subseteq S^{\infty}S_i^{+},$$

and that the inclusion is an equality when the ambiguous type structure is complete. For the right-hand inclusion fix strategy-type pair $(\bar{s}_i, \bar{t}_i) \in R_i \cap SC_i \cap CBRSC_i$ and simply notice that since $(\bar{s}_i, \bar{t}_i) \in R_i \cap WC_i \cap CBRSC_{i,n}$ for any $n \geq 0$, the arguments above imply $\bar{s}_i \in S^n S_i^{+}$ for any $n \geq 0$. Thus, $\bar{s}_i \in S^{\infty} S_i^{+}$.

For the left-hand inclusion, fix strategy $\bar{s}_i \in S^{\infty}S_i^{+}$. Since, in particular, $\bar{s}_i \in S^n S_i^{+}$ for any $n \geq 0$, the arguments above imply that for any $n \geq 0$ there exists some type $t_i^n \in T_i$ such that $(\bar{s}_i, t_i^n) \in R_i \cap SC_i \cap CBRSC_{i,n}$. Now, let $\bar{M}_i$ denote the closure of the convex-hull of $\bigcup_{n \geq 0} M_i(t_i^n)$ and (using completeness) pick type $\bar{t}_i \in T_i$ such that $M_i(t_i) = \bar{M}_i$. Obviously, $\bar{s}_i$ is a best-reply is to every conjecture induced by the beliefs in $M_i(\bar{t}_i)$ and $\bar{t}_i$ is strongly cautious and is consistent with common belief in rationality and strong cautiousness. Thus, $(\bar{s}_i, \bar{t}_i) \in R_i \cap SC_i \cap CBRSC_i$ and hence, $\bar{s}_i \in \operatorname{proj}_{S_i}(R_i \cap SC_i \cap CBRSC_i)$. ■

The interpretation of parts (*i*) and (*ii*) of each theorem is analogous to that of the theorems in the previous section. However, the implications for the presence of Knightian uncertainty differ. In the case of the Dekel-Fudenberg procedure, if the iteration process requires more than one round, then the preferences of a weakly cautious and rational player that exhibits common belief of rationality and weak cautiousness must necessarily display Knightian uncertainty. This is due to the fact that weak cautiousness requires some belief with full-support in the whole set of states, and belief in the opponents' being rational and weakly cautious requires some belief that puts probability one on a proper subset of the set of states. This implies that the set of beliefs of the player must necessarily contain at least two beliefs, i.e. the Bayesian is not Bayesian. As discussed above, strongly cautious types always represent preferences displaying Knightian uncertainty. In consequence, Theorem 5 and Theorem 6 illustrate the deep connection between cautiousness and Knightian uncertainty: in order to guarantee that strategically sophisticated players choose according to cautious criteria, Knightian uncertainty is required to overcome the inclusion-exclusion problem.[31]

My characterization makes clear that the resolution of the problem requires the stronger notion of *rationality* instead of the (weaker) notion of *justifiability*. The reason is that as seen above multiplicity of beliefs is needed to overcome the problem. However, only under my notion of rationality all of these beliefs matter for the actual choice of the player. Justifiable strategies do not share this feature and therefore cautiousness does not provide any (higher-order) belief restrictions. The following corollary summarizes this observation.

Corollary 2. *Let G be a game. It holds that:*

[31] At least, in those contexts in which strategic sophistication has any bite.

(i) *For any ambiguous type structure, any player i, and any strategy-type pair* (s_i, t_i)*, if type* t_i *is consistent with weak cautiousness and common belief of justifiability and weak cautiousness and* s_i *is justifiable for* t_i*, then* s_i *is undominated; i.e.,*

$$\operatorname*{proj}_{S_i}(J_i \cap WC_i \cap CBJWC_i) \subseteq S_i^1.$$

(ii) *For any player i and any strategy* s_i*, if* s_i *is undominated then there exist a complete ambiguous type structure* $\mathcal{T}$ *and a type* t_i *consistent with weak cautiousness and common belief of justifiability and weak cautiousness for which* s_i *is justifiable; i.e.,*

$$S_i^1 \subseteq \operatorname*{proj}_{S_i}(J_i \cap WC_i \cap CBJWC_i).$$

The same holds if weak *is replaced with* strong *(and SC replaces WC) everywhere.*

Remark 1. *Without formally proving this corollary, it is easy to see that the first inclusion holds trivially. The second inclusion holds, because one can always add more beliefs (either a full support one or a set of beliefs with non-empty interior) while keeping the same strategy justifiable.*

1.5. Conclusions

In this chapter I provide reasoning-based characterizations of some solution concepts in game theory. Whereas most of the literature focuses on the case that a player in a strategic setting is Bayesian and maximizes his or her subjective expected utility, I extend this to the case of *Knightian uncertainty*. I formalize this idea by allowing players to have a set of beliefs instead of having a unique belief. Within this setting, I explore four natural extensions of the Bayesian case along two dimensions:

1. *Choice:* How does a player choose a strategy if she has a set of beliefs?

    (*i*) *Rationality:* Her choice needs to be a best-reply to *all* beliefs in her set.

    (*ii*) *Justifiability:* Her choice needs to be a best-reply to *at least one* beliefs in her set.

2. *Reasoning:* How does a player with a set of beliefs reason about the choice of other players?

    (*iii*) *Belief:* She believes an event E if *at least one* belief in her set assigns probability one to this event.

    (*iv*) *Full belief:* She fully believes an event E if *all* beliefs in her set assigns probability one to this event.

My main result combines these assumptions about choices and (higher-order) reasoning. Three of these combinations result in a characterization that is behaviorally equivalent to rationalizability: rationality and common belief of rationality (Theorem 1), rationality and common full belief of rationality (Theorem 3), and justifiability and common full belief of justifiability (Theorem 4). The behavioral implications of justifiability and common belief of justifiability are only undomindated strategies (Theorem 2). Thus, the behavioral implications of strategic Knightian uncertainty are standard solutions concepts in game theory. Since I consider the extreme cases of existential and universal quantifiers along both dimensions, there is an interesting open question of what the behavioral implications are for intermediate cases. I leave this question open for further research.

My framework of strategic Knightian uncertainty allows for much more flexibility than a Bayesian framework. I make use of this flexibility to overcome the inclusion-exclusion problem of Samuelson (1992), which states that strategic reasoning is in conflict with cautious behavior in a Bayesian setting. Since my framework allows for multiple beliefs and rationality requires that all of these beliefs matter, the problem does not arise. Using two similar, but distinct, notions of cautiousness I use Knightian uncertainty to provide a foundation for the Dekel-Fudenberg procedure (Theorem 5) and strict rationalizability (Theorem 6). Both of these characterizations are conceptually similar to Δ-rationalizability, an umbrella solution concept introduced by Battigalli and Siniscalchi (2003) in a Bayesian setting. Extending such a general characterization to my case of Knightian uncertainty is also left for future research.

CHAPTER 2

Strategic Cautiousness as an Expression of Robustness to Ambiguity[1]

"Zwey Seelen wohnen, ach! in meiner Brust"

Faust by Johann Wolfgang von Goethe (1841, p.56)

[1]This chapter was developed together with Peio Zuazo-Garin and the results were published in Ziegler and Zuazo-Garin (2020). As explained in Footnote 1 of Chapter 1 the grammatical singular form will be used throughout the dissertation.

2.1. Introduction

Economic modeling often invokes the avoidance of weakly dominated strategies as a criterion for equilibrium selection.[2] That is, players should choose *admissible* strategies. The rationale goes back to at least Wald (1939) where he addresses the foundations of statistics in terms of a single decision making problem. Soon after, economists started to use the same criterion, because as Arrow (1951, p. 429) states such rules are "extremely reasonable." In the context of game theory, Luce and Raiffa (1957) promote "[f]irst [...] discarding many of the inadmissible strategies [and then using] iterative procedures." Harsanyi (1962) is among the first to apply this reasoning to bargaining.[3] Further applications were considered by Moulin (1979). More recently Dekel and Fudenberg (1990, p. 245) use the "iterated deletion of weakly dominated strategies since it clearly incorporates certain intuitive objectives of rationality postulates." A formal argument for admissible strategies is due to Pearce (1984). Intuitively, Pearce's result says a decision maker would choose an admissible strategy if and only if she is *cautiousness*, which dictates that players favor strategies that, ceteris paribus, hedge against unexpected behavior. That is, she deems all eventualities as possible.

However, Samuelson's (1992) classic analysis illustrates that strategic reasoning is in conflict with the criterion of cautiousness. If players are modeled as subjective expected

[2]See, for example, Kohlberg and Mertens (1986), Palfrey and Srivastava (1991), Feddersen and Pesendorfer (1997), or Sobel (2017, 2019).

[3]Harsanyi describes the procedure as corresponding to mutual knowledge of admissibility, but it is clear from his exposition that he had in mind common knowledge of admissibility.

utility maximizers, the clash seems inescapable: Strategic reasoning requires each player i's beliefs to assign zero probability to some of the strategies of i's, while cautiousness requires player i's decision to be sensitive to these strategies that receive zero probability (and are therefore of negligible importance for the maximization problem). Thus, the seemingly mutually exclusive nature of strategic reasoning and cautiousness requires clarification. Such an understanding is desirable in particular in scenarios where behavior is likely to be reasoning-based and cautiousness plays a role.

This chapter proposes a new take on this longstanding problem by suggesting a novel theoretical foundation for the interplay between strategic reasoning and cautiousness. The analysis by Samuelson (1992) clearly shows that two ingredients are necessary to overcome this tension: First, multiple beliefs are needed to account for epistemic conditions that would be mutually excluding if required to be satisfied by a single belief. Second, the best-reply needs to be sensitive to all these beliefs. I achieve this within my framework by augmenting the underlying standard decision-theoretic foundation for each player by allowing for incomplete preferences à la Bewley (2002) where: (*i*) Each player's strategic uncertainty is represented by a possibly non-singleton set of beliefs thus allowing for ambiguity, and (*ii*) a rational player chooses a strategy that is a best-reply to every belief in her set, so that the resulting choice is robust to the possible ambiguity faced by the player.[4] Under this set-up, and inspired by Brandenburger et al. (2008), I say that a player *assumes* certain behavior by her opponents if *at least one* of the beliefs in her set has full-support

[4]Due to incompleteness such a strategy might not exist for a given set of beliefs. In such case I also say that the player is not rational; see Proposition 1 and the surrounding discussion.

on the collection of states representing such behavior.[5] Consequently, the introduction of ambiguity and the requirement of robustness give great flexibility: It is possible for a player to assume certain behavior and, simultaneously, assume certain more restrictive behavior. If the player is also rational, her choice needs to be a best-reply to both of these beliefs. Hence, in particular, the tension between strategic reasoning and cautiousness is solved: A player can be strategically sophisticated by having one belief that assigns zero probability to her opponents playing dominated strategies, and at the same time cautious by having another belief that assigns positive probability to every strategy of her opponents. Thus, my model overcomes the problem as identified by Samuelson (1992) since it allows precisely for the two necessary ingredients.

Based on the above, I build a framework that provides reasoning-based foundations for iterated admissibility—the iterated elimination of weakly dominated strategies. In Theorem 7 I show that, when type structures are belief-complete (roughly speaking, *rich* enough to capture any possible belief hierarchy), iterated admissibility characterizes the behavioral implications of rationality, cautiousness, and common assumption thereof. From my characterization, it is easy to see that the foundations of iterated admissibility necessarily require the presence of ambiguity whenever strategic reasoning has any bite. If the elimination procedure consists of multiple rounds, the set of ambiguous beliefs needs to contain a specific belief with full-support on the set of opponents' strategies that survive each round. Theorem 8 provides the analysis for the relaxation of belief-

[5]Thus, in addition to what I call *belief* in Chapter 1, here I add a full support restriction.

completeness and shows that, in this case, it is *self-admissible* sets à la Brandenburger et al. (2008) which characterize the behavioral implications of rationality, cautiousness and common assumption thereof.

The literature studying the conflict between strategic reasoning and cautiousness is epitomized by the seminal paper by Brandenburger et al. (2008), who shed light on the question by building upon the lexicographic probability system approach by Blume et al. (1991).[6] Lexicographic probability systems represent the uncertainty faced by a decision maker whose preferences depart from standard Bayesian preferences by allowing violations of the continuity axiom. In this setting, Brandenburger et al. (2008) provide reasoning-based foundations for finitely many iterations of weakly dominated strategy elimination based on rationality and finite-order assumption of rationality, but also present a celebrated impossibility result: under some standard technical conditions and generically in all games, common assumption of rationality cannot be satisfied. This negative result has spurred a line of research concerned with obtaining sound epistemic foundations for iterated admissibility. Keisler and Lee (2015) and Yang (2015) propose answers by tweaking topological properties of the modeling of higher-order beliefs and the notion of assumption, respectively, while Lee (2016) obtains foundations by proposing a modification in the definition of coherence.[7] Catonini and De Vito (2018) also provide foundations by introducing a weaker notion of the likeliness-ordering of events that characterizes the lexicographic probability system, and via an alternative definition of cautiousness that

[6] Early contributions along the same lines include, for example, Brandenburger (1992) and Stahl (1995).

[7] Similar to Epstein and Wang (1996), coherency is imposed on the preferences directly, not only on the beliefs that represent the preferences.

restricts attention to the payoff-relevant component of the states. In a slightly different direction, Heifetz, Meier, and Schipper (2019) propose a new solution concept, *comprehensive rationalizability*, that coincides with iterated admissibility in many settings and admits epistemic foundations. Within a standard Bayesian decision-theoretic model, Barelli and Galanis (2013) provide a characterization for iterated admissibility by introducing an exogenous 'tie breaking' criterion. Robustness to ambiguity is studied by Stauber (2011, 2014) with a different interpretation from ours.

My approach can be regarded as complementary to the lexicographic probability system approach as standard Bayesian preferences are also abandoned by dropping completeness instead of continuity. Both these relaxations allow for multiple beliefs, but while the former requires a specific order, my model drops the order altogether and allows for multiplicity directly. However, apart from the transparent link between cautiousness and robustness to ambiguity that my framework allows for, the *nice* structure of the sets of ambiguous beliefs representing incomplete preferences has some additional advantages. First, it is easy to show that rationality and common assumption of rationality is a non-empty event and thus, that iterated admissibility is properly founded for *all* games. Second, the definitions and formalism involved do not require departures from the canonical definition of the objects involved: (*i*) The modeling of higher-order beliefs (i.e., the type structures employed), including the definition of coherence, and the version of *assumption* that I rely on are natural extensions of their counterparts in the realm of standard Bayesian preferences; and (*ii*) the notion of cautiousness invoked in my theorems is not necessarily restricted to environments where the sets of states have a specific structure (e.g. games).[8] Finally, the presence of ambiguity via incomplete preferences has been shown to be empirically testable by recent work by Cettolin and Riedl (2019).

2.2. Non-technical overview

2.2.1. Examples

To illustrate the intuition behind the usual tension between rationality and cautious behavior and to show how my approach avoids this issue, I present two examples.

Example 1. Consider a two player game with the following payoff matrix:[9]

		Bob L	Bob R
Ann	T	1, 1	0, 1
Ann	D	0, 2	1, 0

Clearly, no action is strictly dominated for either player, so (standard) rationalizability predicts $\{T, D\} \times \{L, R\}$. However, R is weakly dominated by L. Deleting R will therefore make D strictly dominated in the reduced game. Thus, iterated admissibility has a unique prediction in this game: (T, L).

Now assume that one wishes to study how players themselves reason about this game. If Bob is rational and cautious he should play L. Suppose Ann is cautious as well.

[8] Though they are sensitive to topological specifications.

[9] This is the leading example of Brandenburger et al. (2008) and was introduced by Samuelson (1992).

Therefore her belief has to put positive probability on Bob playing L *and* on Bob playing R. However, if Ann believes that Bob is rational and cautious, then she should rule out Bob playing R. This is the inclusion-exclusion problem as identified by Samuelson (1992). On the one hand, Ann should include R in her belief because she is cautious. On the other hand, she should exclude R because she believes that Bob is rational and cautious. ◇

In my framework there is more flexibility because players are not Bayesian, but are allowed to have a (potentially non-singleton) set of beliefs. To see how this relaxation avoids the tension just described, I provide a slightly more elaborate example, which also explores the reasoning of the players more explicitly.

Example 2. Again, there are two players, Ann and Bob, who play the following game:

		Bob A	Bob B	Bob C	Bob D
Ann	H	1, 0	1, 4	1, 2	1, 0
	M	1, 4	1, 0	1, 2	1, 0
	L	0, 2	0, 2	1, 2	3, 2

Now suppose that each player faces ambiguity (as described by Bewley, 2002) about the strategy choice of their opponent. That is, neither player has a unique belief about the opponent's strategy choice, but rather each has a set of beliefs. In particular, suppose that Ann has a convex closed set of beliefs described by two extreme points. Her first belief is uniform across all of Bob's strategies, $\mu_A^1(s_B) = 1/4$ for $s_B = A, B, C, D$, and her second (extreme) belief is uniform across A, B, and C only, $\mu_A^2(s_B) = 1/3$ for $s_B = A, B, C$. Similarly, Bob faces uncertainty about Ann's choice. Consider the following set of beliefs for Bob, which also has two extreme beliefs. The first is uniform across all of Ann's strategies, $\mu_B^1(s_A) = 1/3$ for $s_A = H, M, L$, and the second belief assigns equal probability to H and M, $\mu_B^2(H) = \mu_B^2(M) = 1/2$.

What strategies are rational for each player given their beliefs? Preferences à la Bewley (2002) are incomplete, and for incomplete preferences there is no obvious definition of rationality: Optimality is a stronger requirement than maximality for incomplete orders. As stated in the introduction, the solution to the inclusion-exclusion problem requires that a best-reply to be sensitive to all beliefs. Thus, I identify rationality with optimality so that a rational strategy is a best-reply to *all* beliefs, i.e. the choice needs to be robust to the ambiguity faced by the player. In this example this implies that Ann will not rationally choose L since it is not a best-reply that is robust to the ambiguity that she faces. H and M, on the other hand, are best-replies to all beliefs and are therefore rational choices for Ann. For Bob, only D is not rational because it is not a best-reply to any of his beliefs. The three other strategies A, B, and C are rational as they are best-replies to all of his beliefs. Thus, with these sets of beliefs the prediction of the model would correspond to iterated admissibility. This is not a coincidence and foreshadows my results on the characterization of iterated admissibility, explained in more detail below, where the strategic reasoning is also made explicit. $\diamond$

2.2.2. Heuristic treatment of strategic reasoning

In the previous examplesit can be seen that a set of beliefs enables strategic reasoning and cautiousness to be incorporated. To study games in general, players need to be allowed to reason about the reasoning process of other players too. This necessitates the formalizing of infinite sequences of the following form:

a_1: Ann is rational and cautions	b_1: Bob is rational and cautions
a_2: a_1 holds and Ann assumes b_1	b_2: b_1 holds and Bob assumes a_1
a_3: a_1 holds and Ann assumes b_1 & b_2	b_2: b_1 holds and Bob assumes a_1 & a_2
$\cdots$	$\cdots$

If this infinite sequence holds, I say that there is *rationality, cautiousness, and common assumption thereof* (RCCARC).

To study these infinite sequences and to see which strategies are played if they hold, (epistemic) types need to be introduced for each player. Accordingly, consider T_A and T_B as type spaces for Ann and Bob, respectively. Usually, each of Ann's type $t_A \in T_A$ is associated with a belief about Bob's strategy *and* type, i.e. a probability distribution over $S_B \times T_B$. However, the idea here is to model players who face ambiguity, so each type is associated with a (closed, and convex) set of beliefs about $S_B \times T_B$. Thus, for a strategy-type pair of Ann (s_A, t_A), strategy s_A is said to be *rational* if s_A is a best-reply to *all* of the beliefs associated with t_A. Whether a player is cautious depends only on her beliefs: she thinks everything is possible. That is, one of her beliefs has full support on the full space of uncertainty. Thus, I say that Ann's type t_A is *cautious* if there exists a belief in the associated set of beliefs which has full support on $S_B \times T_B$.

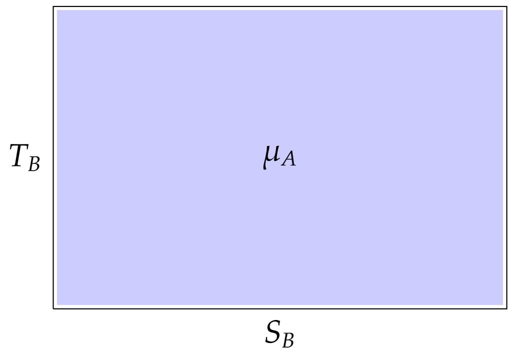

Figure 2.1: Cautiousness

For example, consider a type of Ann's, t_A, which has only a singleton set of beliefs $\{\mu_A\}$ with support as depicted in Figure 2.1. For such a cautious type, the question arises of which strategies are rational. Accordingly, consider the marginal of μ_A on Bob's strategy space S_B. This marginal has full support on S_B and if Ann is rational, her rational choice has to be a best-reply to this marginal. It then follows from Pearce (1984) that she must choose a strategy which is not weakly dominated.

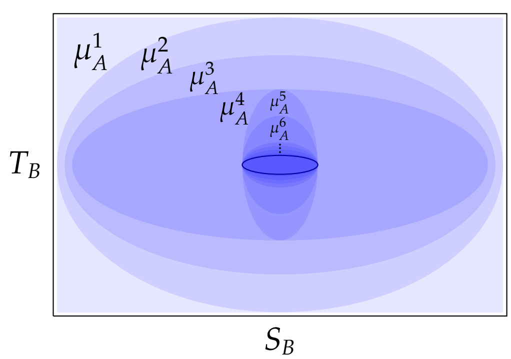

Figure 2.2: Rationality, cautiousness, and common assumption thereof.

Now, it is possible to study the infinite sequences described above. In this case the picture that emerges looks like Figure 2.2. Here the small area with solid boundary corresponds to all strategy-type combinations of Bob satisfying RCCARC. Now, set a strategy-type combination (s_A, t_A) for Ann. Does this type correspond to RCCARC for Ann? i.e. does the type satisfy the sequence $a_1, a_2, \ldots$? It is already known that if a_1 holds there needs to be a belief in the associated set of beliefs which has support as μ_A^1. Next, it is considered that Ann assumes b_1. This rules out some of Bob's strategy-type pairs, but also requires t_A to have a belief which has full support on the remaining pairs. Thus, in the associated set of beliefs there needs to a belief μ_A^2. In the next step, Ann is considered to assume b_1 and b_2. Similar reasoning applies and there needs to be a belief like μ_A^3 in the set of beliefs corresponding to t_A. This procedure can now be iterated (as indicated in the picture) to verify whether the type t_A corresponds to RCCARC for Ann. Only finite games are considered here so at some stage n this iteration no longer rules out any *strategies* for Bob. However, it might be the case that at *every* step there are still some types of Bob's that need to be ruled out. In the worst case there needs to be a different belief for each iteration as the support of each belief is changing over the course of the sequence. However, this does not cause a problem. For each type the set of beliefs could be potentially very large.[10] Since such large sets of beliefs are within the framework under consideration, the event RCCARC is not empty. Thus I do not get a negative result as Brandenburger et al. (2008) find in a different framework. To illustrate more specifically how this analysis works, types are added explicitly for the example considered above.

[10]That is, not finitely generated sets.

Example 2 (continuing from p. 80)**.** Consider the following type space $T_i = \left\{t_i^0, t_i^1, t_i^2, t_i^3\right\}$ for $i = A, B$ and define (with some abuse of notation) the following beliefs on $S_B \times T_B$:

$$\mu_A^1(s_B, t_B) = 1/16, \text{ for all } (s_B, t_B) \in S_B \times T_B,$$

$$\mu_A^2(s_B, t_B) = 1/9, \text{ for all } (s_B, t_B) \in \{A, B, C\} \times \{t_B^1, t_B^2, t_B^3\}, \text{ and}$$

$$\mu_A^3(s_B, t_B^3) = 1/3, \text{ for all } s_B \in \{A, B, C\}.$$

Similarly, define the following beliefs on $S_A \times T_A$:

$$\mu_B^1(s_A, t_A) = 1/12 \text{ for all } (s_A, t_A) \in S_A \times T_A,$$

$$\mu_B^2(s_A, t_A^1) = 1/6, \text{ for all } s_A \in S_A, \ \mu_B^2(s_A, t_A) = 1/8, \text{ for all } (s_A, t_A) \in \{H, M\} \times \{t_A^2, t_A^3\}, \text{ and}$$

$$\mu_B^3(s_A, t_A^3) = 1/2, \text{ for all } s_A \in \{H, M\}.$$

Given these beliefs, define the set of beliefs $M_i(t_i)$ for each type as follows: for $i = A, B$ set $M_i(t_i^0) = \left\{\mu_i^3\right\}$, $M_i(t_i^1) = \left\{\mu_i^1\right\}$, $M_i(t_i^2)$ is the convex hull of μ_i^1 and μ_i^2, and $M_i(t_i^3)$ is the convex hull of μ_i^1, μ_i^2, and μ_i^3

Now, it is possible to analyze the infinite sequences $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$ introduced above. a_1 is the event that Ann is rational and cautious, so I have to collect all strategy-type pairs which satisfy the full-support requirement (cautiousness) and the requirement that the strategy is a best-reply to all beliefs of the given type (rationality). Here, all types but t_i^0 have at least one belief with full support on $S_{-i} \times T_{-i}$. Together with rationality this gives that the following strategy-type pairs correspond to a_1: $S_A \times \{t_A^1\} \cup \{H, M\} \times \{t_A^2, t_A^3\}$. Similarly, b_1 corresponds to $\{A, B, C\} \times \{t_B^1, t_B^2, t_B^3\}$. So both a_1 and b_1 rule out some strategy-type pairs and in particular the weakly dominated strategy D is ruled out. Next, to get to a_2, I want to find all types of Ann that assume b_1. That is, all types of Ann that have at least one belief with full-support on $\{A, B, C\} \times \{t_B^1, t_B^2, t_B^3\}$. Only t_A^2 and t_A^3 satisfy this requirement, leaving $\{H, M\} \times \{t_A^2, t_A^3\}$ corresponding to a_2. For Bob, it emerges that b_2 corresponds to $\{A, B, C\} \times \{t_B^2, t_B^3\}$. Again, note that in this step the interactive reasoning leads to the ruling out of L, which is weakly dominated after elimination of D. In the next step (i.e. a_3 and b_3), types t_i^2 are ruled out, but no more strategies. This construction, however, would lead to the conclusion that a_4 and b_4 do not correspond to any strategy-type pairs. The solution, and this is the main idea of how to prove one direction of Theorem 7, is to add more types. For each iteration add another type with full support on the previous rounds (similar to types t_i^3). This gives an infinite (but countable) number of types and only the "limiting" type corresponds to RCCARC. This argument shows that the illustration in Figure 2.2 is accurate in the sense that for higher order iterations the supported strategies are constant, but only types are removed in each round.

Theorem 8 provides a direct (and hence different) way to construct finite type structures so that for strategy-type pairs satisfying RCCARC the strategies of iterated admissibility (or these of any other self-admissible set) are obtained. $\diamond$

2.3. Preliminaries

This section presents the main concepts used in my analyis. The object of study is the inclusion-exclusion problem inherent in the iterated elimination of weakly dominated strategies raised by Samuelson (1992). Thus, Subsection 2.3.1 recalls the formalization of strategic-form games, iterated admissibility, and self-admissible sets (Brandenburger et al., 2008). Subsection 2.3.2 introduces the relevant definitions of rationality and the epistemic objects.[11]

2.3.1. Games and iterated strategy elimination

A *game* consists of a tuple $G := \langle I, (S_i, u_i)_{i \in I} \rangle$ where I is a finite set of *players*, and for each player i there is a finite set of (pure) *strategies* S_i and a *utility function* $u_i : S \to \mathbb{R}$, where $S := \prod_{i \in I} S_i$ denotes the set of strategy *profiles*. For each player i a randomization of own strategies $\sigma_i \in \Delta(S_i)$ is referred to as a *mixed* strategy,[12] and a probability measure $\mu_i \in \Delta(S_{-i})$, where $S_{-i} := \prod_{j \neq i} S_j$, as a *conjecture*. When necessary, with some abuse of notation, I use s_i to refer to the degenerate mixed strategy that assigns probability one to s_i. Each conjecture μ_i and possibly mixed strategy σ_i naturally induce *expected* utility $U_i(\mu_i; \sigma_i)$ and based on this, each player i's *best-reply* correspondence is defined by assigning

[11] Since my analysis models players as individual decision makers whose beliefs may display ambiguity via incomplete preferences, Subsection 1.3.2 and Subsection 1.3.3 illustrate the underlying decision theory and how to envision games as decision problems, respectively.

[12] Recall that for any topological space X, as usual and as in Chapter 1, $\Delta(X)$ denotes the set of probability measures on the Borel σ-algebra of X.

to each conjecture μ_i the subset of pure strategies $BR_i(\mu_i)$ that maximize its corresponding expected utility.[13]

Following the duality results of Pearce (1984), I use the best-reply correspondence directly to define iterated admissibility whose foundations are then studied in Section 2.4. Strategy s_i is *iteratively admissible* if it survives the iterated elimination of weakly dominated strategies; i.e., if it is not weakly dominated given strategy profiles $S_{-i} \times S_i$, it is not weakly dominated given strategy profiles $W^1_{-i} \times W^1_i$ consisting only of strategies surviving the first elimination round, etc. Thus, formally, strategy s_i is iteratively admissible if $s_i \in W^\infty_i := \bigcap_{n \geq 0} W^n_i$, where $W^0_i := S_i$ and for any $n \in \mathbb{N}$,

$$W^n_i := \left\{ s_i \in W^{n-1}_i \;\middle|\; \begin{array}{l} \text{There exists some } \mu_i \in \Delta(S_{-i}) \text{ such that:} \\ (i) \quad \operatorname{supp} \mu_i = \prod_{j \neq i} W^{n-1}_j, \\ (ii) \quad s_i \in BR_i(\mu_i) \end{array} \right\}.$$

Finally, that set of strategy profiles $Q = \prod_{i \in I} Q_i$ is said to be a *self-admissible set* (SAS) if for every player i the following three conditions are satisfied:

(i) No $s_i \in Q_i$ is weakly dominated given $S_{-i} \times S_i$.

(ii) No $s_i \in Q_i$ is weakly dominated given $Q_{-i} \times S_i$.

(iii) For every $s_i \in Q_i$ and every mixed strategy σ_i such that $U_i(s_{-i}; \sigma_i) = U_i(s_{-i}; s_i)$ for every s_{-i}, it holds that $\operatorname{supp} \sigma_i \subseteq Q_i$.

[13] That is, given conjecture μ_i the expected utility is $U_i(\mu_i; \sigma_i) := \sum_{(s_{-i}; s_i) \in S} \mu_i[s_{-i}] \cdot \sigma_i[s_i] \cdot u_i(s_{-i}; s_i)$ for each possibly mixed strategy σ_i, and the set of best-replies is $BR_i(\mu_i) := \arg\max_{s_i \in S_i} U_i(\mu_i; s_i)$.

The connection between the notions of self-admissibility and iterated admissibility is immediately apparent: the set of iteratively admissible strategy profiles is a self-admissible set of game G, but in general there are other self-admissible sets. For details see Brandenburger and Friedenberg (2010), who also study properties of self-admissible sets for specific (classes of) games.

2.3.2. Behavioral and epistemic conditions

As in Subsection 1.3.4, I extend the definition of (standard Bayesian) type structures so to be able to deal with the possibility of ambiguity. Recall that an *ambiguous type structure* consists of a list $\mathcal{T} := \langle T_i, M_i \rangle_{i \in I}$ where for each player i there is:[14]

(i) A set of (*ambiguous*) *types* T_i.

(ii) An *ambiguous belief map* $M_i : T_i \to \mathscr{M}_i(S_{-i} \times T_{-i})$, where $T_{-i} := \prod_{j \neq i} T_j$, that associates each type with ambiguous beliefs on opponents' strategy-type pairs.

With such a type structure, the analysis of each player i's reasoning is focused on strategy-type pairs (s_i, t_i), which specify both player i's choice, and as described above, her ambiguous beliefs on her opponents' choices, her ambiguous beliefs on her opponents' first-order ambiguous beliefs, etc.

Next, I first recall the notion of rationality that I employ. Second, I introduce my formalization of cautiousness as a manifestation of ambiguity. Finally, I define the appropriate

[14]I assume each T_i to be compact and metrizable and each M_i, continuous. See Footnote 18 in Chapter 1.

tool to impose restrictions on higher-order beliefs, which is a generalization to Bewley preferences of the usual notion of full-support belief for standard Bayesian preferences.

Rationality

I say that strategy s_i is *rational* for type t_i if s_i is a best-reply to every first-order ambiguous belief induced by t_i; thus, the set of strategy-type pairs in which player i is rational is formalized as follows:

$$R_i := \left\{ (s_i, t_i) \in S_i \times T_i \,\middle|\, s_i \in \bigcap_{\mu_i \in M_i(t_i)} BR_i(\operatorname*{marg}_{S_{-i}} \mu_i) \right\}.$$

I discuss properties of this notion of rationality in more detail in Subsection 1.3.5. However, I want to stress that this notion of rationality also incorporates the requirement that each player does have a rational strategy available. If this is not the case, which is possible due to incompleteness, a player will be irrational.

Cautiousness and ambiguity

I next argue that cautiousness, intuitively thought of as the decision maker considering every state of the world when deciding which choice is best, can be interpreted as a product of ambiguity in the sense that types that exhibit cautiousness tend to represent preferences that also display ambiguity. I first formalize the notion of cautiousness that takes part in the characterizations result in Section 2.4 and then discuss its link to ambiguity.[15]

[15]This defintion is equivalent to what I call *weak* cautiousness in Definition 1. But since strong cautiousness does not play a role in this chapter, I omit the qualifier. I note that all of the following analyses could have

Definition 2 (Cautiousness)**.** *Let G be a game and $\mathcal{T}$, an ambiguous type structure. Then, for any player i and any type t_i I say that type t_i is* cautious *if at least one belief in $M_i(t_i)$ has full-support on $S_{-i} \times T_{-i}$. I denote the set of player i's strategy-type pairs in which the type is cautious by C_i.*

If at an intuitive level cautiousness is seen as the idea that a decision maker takes every possible contingency into account, then that is present in this definition. Cautiousness requires, loosely speaking, that every state is taken into account by the decision maker.[16] The link with ambiguity is easy to see. In principle, it is possible for a type to display cautiousness but not ambiguity. This is the case of every type whose set of ambiguous beliefs consists of a single belief with full-support on $S_{-i} \times T_{-i}$ as in Figure 2.1. However, if in addition to cautiousness the type also exhibits some form of strategic sophistication in the sense of having a (different) belief that rules out some proper subset of $S_{-i} \times T_{-i}$, then, necessarily, the type displays ambiguity: The corresponding ambiguous beliefs a fortiori contain at least two different beliefs. Hence, the introduction of ambiguity not only enables strategic reasoning and cautiousness to be made compatible, but is indeed, necessary when strategic reasoning has any bite.

been carried out employing a slightly weaker notion of cautiousness than the one introduced in Definition 2. In principle it would suffice to require full support on S_{-i} rather than $S_{-i} \times T_{-i}$. My reason for opting for the stronger notion is twofold: (*i*) It does not prevent my characterization from dispensing with impossibility issues à la Brandenburger et al. (2008) (see Subsection 2.4.3), so it is clear that it is not modifications in the notion of cautiousness that enable for this to be achieved; and (*ii*) since it does not apply only to state spaces with product structure, it has a more general decision-theoretic foundation.

[16]Cautiousness is also present in the analysis by Brandenburger et al. (2008). However, there it is incorporated into the definition of rationality. I find it more transparent to explicitly define the event when a player is cautious.

Assumption

Hereafter I refer to measurable subsets $E \subseteq S \times T$ as events. A standard Bayesian decision maker is said to *assume* event E when the unique subjective belief induced by her preference has full-support on E.[17] Some changes are in order if this idea is to be extended to Bewley preferences: The set of ambiguous beliefs may contain beliefs that have different supports. I say that a Bewleyian decision maker assumes event E when *at least one* belief in her set of ambiguous beliefs has full-support on E. Given the inclusion-exclusion problem, it is natural to consider such a weak version of assumption. As discussed in Section 2.1, it is necessary to have multiple beliefs which have potentially different supports to resolve the tension between strategic reasoning and cautiousness.

Definition 3 (Assumption)**.** *Let G be a game and $\mathcal{T}$, an ambiguous type structure. For any player i, any type t_i and any event $E_{-i} \subseteq S_{-i} \times T_{-i}$ I say that type t_i* assumes *E_{-i} if at least one belief in $M_i(t_i)$ has full-support on the topological closure of E_{-i}. I denote the set of player i's strategy-type pairs in which the type assumes E_{-i} by $A_i(E_{-i})$.*

Remark 2. *Cautiousness as defined in Definition 2 can be restated in terms of assumption: A type t_i is cautious if it assumes $S_{-i} \times T_{-i}$.*

[17] Technically, I am considering the collapse of the notion of assumption (see Brandenburger et al., 2008 and Dekel, Friedenberg, and Siniscalchi, 2016) under the lexicographic probability system when the preferences satisfy continuity and the corresponding lexicographic probability system thus collapses to a single belief.

2.4. Iterated admissibility and ambiguous types

This section presents the main results of the chapter. Based on the observation made in the previous section that the presence of ambiguity can reconcile strategic reasoning with cautiousness, I provide foundations for iterated admissibility and self-admissibility in terms of rationality, cautiousness, and certain higher-order assumption constraints. I provide these foundations in Subsection 2.4.1. Then, in Subsection 2.4.2, I discuss the link between iterated assumption and ambiguity to resolve the inclusion-exclusion problem. Finally, in Subsection 2.4.3 I review the seminal impossibility result due to Brandenburger et al. (2008) within the approach in terms of lexicographic probability systems, recall some of the responses in the related literature, and explore the connection with my results.

2.4.1. Epistemic foundation

As mentioned above, the epistemic foundation of iterated admissibility is to be formulated in terms of rationality, cautiousness, and higher-order assumption restrictions. The set of strategy-type pairs in which player i exhibits common assumption in rationality and cautiousness is given by $CARC_i := \bigcap_{n \geq 0} CARC_{i,n}$, where each $CARC_{i,n}$ is defined recursively by setting:

$$CARC_{i,0} := S_i \times T_i,$$

$$CARC_{i,n} := CARC_{i,n-1} \cap A_i(\prod_{j \neq i} R_j \cap C_j \cap CARC_{j,n-1}),$$

for every $n \in \mathbb{N}$. That is, $CARC_i$ brings together all the strategy-type pairs (s_i, t_i) where player i's type t_i assumes that every player $j \neq i$ is rational, cautious, and assumes that every player $j \neq i$ assumes that every player $k \neq j$ is rational, cautious, and so on. Based on the above:[18]

Theorem 7 (Foundation of iterated admissibility)**.** *Let G be a game. The following holds:*

(i) *For any complete ambiguous type structure, any player i and any strategy-type pair* (s_i, t_i)*, if type* t_i *is consistent with cautiousness and common assumption of rationality and cautiousness and* s_i *is rational for* t_i*, then* s_i *is iteratively admissible; i.e.,*

$$\operatorname*{proj}_{S_i}(R_i \cap C_i \cap CARC_i) \subseteq W_i^{\infty}.$$

(ii) *For any player i and any strategy* s_i*, if* s_i *is iteratively admissible then there exist a complete ambiguous type structure* $\mathcal{T}$ *and a type* t_i *consistent with cautiousness and common*

[18] The theorem is stated and holds only for a complete type structure because the assumption operator is not monotone. This is similar to, for example, assumption in Brandenburger et al. (2008) or strong belief of Battigalli and Siniscalchi (2002). An example showing why completeness is needed is available upon request.

assumption of rationality and cautiousness for which s_i is rational; i.e.,

$$W_i^\infty \subseteq \operatorname*{proj}_{S_i}(R_i \cap C_i \cap CARC_i).$$

As an intermediate step, which is of interest by itself I first prove the result for every finite iteration.

Proposition 2. *Let G be a game and $\mathcal{T}$ a complete ambiguous type structure. For any $n \in \mathbb{N}$ and every player i the following holds:*

$$\operatorname*{proj}_{S_i}(R_i \cap C_i \cap CARC_{i,n}) = W_i^{n+1}.$$

Proof. For the sake of convenience, for each player i define $X_{i,0} := S_i \times T_i$ and for any $n \in \mathbb{N}$, $X_{i,n} := R_i \cap C_i \cap CARC_{i,n-1}$. Now, I proceed by induction on n:

Initial Step ($n = 0$). For the right-hand inclusion, set strategy-type pair $(\bar{s}_i, \bar{t}_i) \in R_i \cap C_i$ and denote $\bar{M}_i = M_i(\bar{t}_i)$. Then, since $\bar{t}_i$ is cautious there exists a belief $\mu_i^1 \in \bar{M}_i$ whose support is $S_{-i} \times T_{-i}$, and since $(\bar{s}_i, \bar{t}_i) \in R_i$ it follows that $\bar{s}_i$ is a best-reply for $\operatorname{marg}_{S_{-i}} \mu_i^1$. Thus, $\bar{\mu}_i^1 := \operatorname{marg}_{S_{-i}} \mu_i^1$ is a conjecture in with full-support on S_{-i} for which $\bar{s}_i$ is a best-reply. Hence, $\bar{s}_i \in W_i^1$.

For the left-hand inclusion, set strategy $\bar{s}_i \in W_i^1$ and conjecture $\bar{\mu}_i$ with full-support on S_{-i} for which $\bar{s}_i$ is a best-reply. Then, take arbitrary full-support belief $\eta_i \in \Delta(T_{-i})$ and set $\mu_i^1 := \bar{\mu}_i \times \eta_i$ and $\bar{M}_i := \{\mu_i^1\}$. Since $\mathcal{T}$ is complete there exists a type $\bar{t}_i \in T_i$ such that

$M_i(\bar{t}_i) = \bar{M}_i$. Since μ_i^1 has full-support on $S_{-i} \times T_{-i}$ it holds that $\bar{t}_i$ is cautious, and hence, that $(\bar{s}_i, \bar{t}_i) \in C_i$, and as $\bar{s}_i$ is a best-reply to the marginal on S_{-i} induced by the unique belief in $M_i(\bar{t}_i)$ it follows that $(\bar{s}_i, \bar{t}_i) \in R_i$. Thus, it can be concluded that $(\bar{s}_i, \bar{t}_i)$ is a strategy-type pair in $R_i \cap C_i$ that induces $\bar{s}_i$.

Inductive Step. Suppose that $n \geq 0$ is such that the claim holds. Next, I verify it holds for $n+1$. For the right-hand inclusion, set strategy-type pair $(\bar{s}_i, \bar{t}_i) \in R_i \cap C_i \cap CARC_{i,n+1}$ and denote $\bar{M}_i := M_i(\bar{t}_i)$. Then, since $(\bar{s}_i, \bar{t}_i) \in R_i \cap C_i \cap CARC_{i,n}$ it is known from the induction hypothesis that $\bar{s}_i \in W_i^{n+1}$, and since $(\bar{s}_i, \bar{t}_i) \in R_i \cap CARC_{i,n+1}$ there must exist a belief $\mu_i^{n+1} \in \bar{M}_i$ whose support is the closure of $X_{-i,n+1} := \prod_{j \neq i} X_{j,n+1}$ and whose marginal on S_{-i} admits $\bar{s}_i$ as a best-reply. It follows from the induction hypothesis and completeness that the support of $\bar{\mu}_i^{n+1} := \text{marg}_{S_{-i}} \mu_i^{n+1}$ is W_{-i}^{n+1} and hence, it can be concluded that $\bar{s}_i \in W_i^{n+2}$.

For the left-hand inclusion, set strategy $\bar{s}_i \in W_i^{n+2}$ and family of conjectures $\{\bar{\mu}_i^k\}_{k=1}^{n+2}$ such that for each $k = 1, \ldots, n+2$: (*i*) $\bar{\mu}_i^k$ has full-support on W_{-i}^{k-1}, and (*ii*) $\bar{s}_i$ is a best-reply to $\bar{\mu}_i^k$. Now, set arbitrary $k = 0, \ldots, n+1$ and for any player $j \neq i$ and any strategy $s_j \in W_j^k$ define:

$$Y_{j,k}(s_j) := \operatorname*{proj}_{T_j} \left(\{s_j\} \times T_j \cap X_{j,k} \right),$$

which is known from the induction hypothesis to be non-empty. It is also known from the induction hypothesis that $\{Y_{j,k}(s_j) | s_j \in W_j^k\}$ is a finite cover of $\text{proj}_{T_j}(X_{j,k})$. Now, for each $s_{-i} \in W_{-i}^k$ pick arbitrary belief $\eta_i^k(s_{-i}) \in \Delta(\prod_{j \neq i} Y_{j,k}(s_j))$ whose support is the closure of

$\prod_{j\neq i} Y_{j,k}(s_j)$, and define belief μ_i^k in $\Delta(S_{-i} \times T_{-i})$ as follows:

$$\mu_i^k[E] := \sum_{s_{-i}\in W_{-i}^{k-1}} \bar{\mu}_i^{k+1}[s_{-i}] \cdot \eta_i^k(s_{-i}) \left[E \cap \prod_{j\neq i} \{s_j\} \times Y_{j,k}(s_j) \right].$$

Obviously, μ_i^k is well-defined and its support is exactly the closure of $X_{-i,k} := \prod_{j\neq i} X_{j,k}$.[19] Notice in addition that the marginal of μ_i^k on S_{-i} is precisely $\bar{\mu}_i^{k+1}$ and therefore $\bar{s}_i$ is a best-reply to μ_i^k. Then, let $\bar{M}_i$ be the convex hull of $\{\mu_i^k\}_{k=0}^{n+1}$ and pick type $\bar{t}_i \in T_i$ such that $M_i(\bar{t}_i) = \bar{M}_i$. Clearly, the following two hold:

- $(\bar{s}_i, \bar{t}_i) \in C_i \cap CARC_{i,k}$ for any $k = 0, \dots, n+1$. To see this, simply note that for any $k = 0, \dots, n+1$, it holds that $\mu_i^k \in M_i(\bar{t}_i) = \bar{M}_i$. Then, the claim is proven since (as seen above) the support of μ_i^k is exactly the closure of $X_{-i,k}$.

- $(\bar{s}_i, \bar{t}_i) \in R_i$. This follows immediately from—as seen above—$\bar{s}_i$ being a best-reply to the conjecture induced by each belief in $\{\mu_i^k\}_{k=0}^{n+1}$ and thus, also to each belief in $M_i(\bar{t}_i)$.

Thus, it can be concluded that $(\bar{s}_i, \bar{t}_i)$ is a strategy-type pair in $R_i \cap C_i \cap CARC_{i,n+1}$ that induces $\bar{s}_i$. ■

Now, the proof makes use of Proposition 2.

Proof of Theorem 7. For the right-hand inclusion set strategy-type pair $(\bar{s}_i, \bar{t}_i) \in R_i \cap C_i \cap CARC_i$ and simply notice that since $(\bar{s}_i, \bar{t}_i) \in R_i \cap C_i \cap CARC_{i,n}$ for any $n \geq 0$, Proposition 2 reveals that $\bar{s}_i \in W_i^n$ for any $n \geq 1$. Thus, $\bar{s}_i \in W_i^\infty$.

[19] For the latter, simply note that for any (s_{-i}, t_{-i}), $\mu_i^k[N] > 0$ for any neighborhood N of (s_{-i}, t_{-i}) if and only if $\eta_i^k(s_{-i})[N] > 0$ for any neighborhood N of (s_{-i}, t_{-i}).

For the left-hand inclusion, set strategy $\bar{s}_i \in W_i^{\infty}$. Since, in particular, $\bar{s}_i \in W_i^{n+1}$ for any $n \geq 0$, it is known from Proposition 2 that for any $n \geq 0$ there exists a type $t_i^n \in T_i$ such that $(\bar{s}_i, t_i^n) \in R_i \cap C_i \cap CARC_{i,n}$. Now, let $\bar{M}_i$ denote the closure of the convex-hull of $\bigcup_{n\geq 0} M_i(t_i^n)$ and pick type $\bar{t}_i \in T_i$ such that $M_i(t_i) = \bar{M}_i$. Obviously, $\bar{s}_i$ is a best-reply is to every conjecture induced by the beliefs in $M_i(\bar{t}_i)$ and $\bar{t}_i$ is cautious and is consistent with common assumption in rationality and cautiousness. Thus, $(\bar{s}_i, \bar{t}_i) \in R_i \cap C_i \cap CARC_i$ and hence, $\bar{s}_i \in \mathrm{proj}_{S_i}(R_i \cap C_i \cap CARC_i)$. ■

Thus, Theorem 7 provides a complete characterization of iterated admissibility. Part (*i*) is a sufficiency result. It shows that whenever a player chooses in a robust way that maximizes with respect to higher-order assumptions that represent common assumption in rationality and cautiousness, then the resulting strategy is necessarily iteratively admissible. Part (*ii*) is, partially, the necessity counterpart: while it is not true that every time an iteratively admissible strategy is chosen this is due to the player being rational, cautious, and best-replying to the higher-order assumption restrictions that represent common assumption in rationality and cautiousness, it is true that every iteratively admissible strategy is a rational choice for a type that is consistent with common assumption in rationality and cautiousness. Note that the proof of Theorem 7 and Proposition 2 reveals from a conceptual perspective that whenever the elimination procedure involves more than one round, satisfying the epistemic conditions above requires players' preferences to display ambiguity. As the next theorem shows, if the requirement of completeness of the type structure is dropped then the behavioral consequences of rationality, cautiousness and common assumption thereof are captured by self-admissibility:

Theorem 8 (Foundation of self-admissibility)**.** *Let G be a game. Then:*

(i) *For any ambiguous type structure $\mathcal{T}$ the set of strategies consistent with rationality, cautiousness and common assumption of rationality and cautiousness is a self-admissible set; i.e., the following set is self-admissible:*

$$\prod_{i \in I} \operatorname{proj}_{S_i}(R_i \cap C_i \cap CARC_i).$$

(ii) *For any self-admissible set Q there exists a finite ambiguous type structure $\mathcal{T}$ for which Q characterizes the behavioral implications of rationality, cautiousness and common assumption of rationality and cautiousness; i.e., such that:*

$$\prod_{i \in I} \operatorname{proj}_{S_i}(R_i \cap C_i \cap CARC_i) = Q.$$

The interpretation is analogous to that of Theorem Theorem 7. Part (i) states that given an arbitrary ambiguous type structure, not necessarily complete, the set of strategy profiles that are consistent with rationality, cautiousness and common assumption of rationality and cautiousness is a self-admissible set. Part (ii) offers the partial converse: For any given self-admissible set Q there exists an ambiguous type structure $\mathcal{T}$, notably, finite, such that Q is exactly the set of strategy profiles that are consistent with rationality, cautiousness and common assumption of rationality and cautiousness within $\mathcal{T}$. Theorem 7 and Theorem 8 are clearly connected because the set of iteratively admissible strategy profiles is itself self-admissible. In particular, for a fixed game this reveals that the set of iteratively admissible strategies can be understood as strategies obtained not only in a very *large* complete type structure, but also under a *smaller* finite one in which, as shown in the proof of Theorem 8, each player i only has as many types as there are iteratively admissible strategies plus one additional dummy type. The proof of Theorem 8 proceeds in a way very similar to the one by Brandenburger et al. (2008) of their characterization result for self-admissible sets (Theorem 8.1). In particular, I need exactly the same number of types for each player.

For the characterization of self-admissible sets a simple preliminary observations is needed first: the reasoning process about *strategies only* stops after finitely many rounds.

Lemma 6. *Let G be a game. Consider an ambiguous type structure $\mathcal{T}$. There exists a $N \in \mathbb{N}$ such that for all $n \geq N$,*

$$\prod_{i \in I} \operatorname*{proj}_{S_i} (R_i \cap C_i \cap CARC_{i,n}) = \prod_{i \in I} \operatorname*{proj}_{S_i} (R_i \cap C_i \cap CARC_{i,N}) .$$

Proof. By definition $CARC_{i,n+1} \subseteq CARC_{i,n}$, so that it also holds that $R_i \cap C_i \cap CARC_{i,n+1} \subseteq R_i \cap C_i \cap CARC_{i,n}$. Since S_i is finite there has to be an $N_i \in \mathbb{N}$ such that $n \geq N_i$

$$\operatorname*{proj}_{S_i} (R_i \cap C_i \cap CARC_{i,n}) = \operatorname*{proj}_{S_i} (R_i \cap C_i \cap CARC_{i,N_i}) .$$

Take $N = \max_i N_i$. ∎

Proof of Theorem 8. For the first part consider an ambiguous type structure $\mathcal{T}$ and consider

$$Q := \prod_{i \in I} \operatorname*{proj}_{S_i} (R_i \cap C_i \cap CARC_i)\,.$$

If $Q = \emptyset$ then Q is a SAS. So assume it is non-empty. Set $s_i \in \operatorname{proj}_{S_i}(R_i \cap C_i \cap CARC_i)$; then there exists a t_i such that $(s_i, t_i) \in R_i \cap C_i \cap CARC_i$. Thus $(s_i, t_i) \in R_i \cap C_i$ implies that condition (i) of SAS is satisfied. Furthermore, with N from Lemma 6 and because $(s_i, t_i) \in CARC_i \subseteq CARC_{i,N+1}$ there must exist a $\mu_i \in M_i(t_i)$ such that $\operatorname{supp} \mu_i = \prod_{j \neq i} R_j \cap C_j \cap CARC_{j,N}$. Then, $\bar{\mu}_i := \operatorname{marg}_{S_{-i}} \mu_i$ is a conjecture with full-support on Q_{-i} (again using Lemma 6) for which s_i is a best-reply. Hence, condition (ii) of SAS is satisfied. Lastly, consider mixed strategy σ_i such that $U_i(s_{-i}; \sigma_i) = U_i(s_{-i}; s_i)$ for every s_{-i}. Then, by Lemma D.2 of Brandenburger et al. (2008) $\operatorname{supp} \sigma_i \subseteq BR_i(\operatorname{marg}_{S_{-i}} \mu_i)$ for every $\mu_i \in M_i(t_i)$ giving $(r_i, t_i) \in R_i$ for all $r_i \in \operatorname{supp} \sigma_i$. Then it also holds that $(r_i, t_i) \in R_i \cap C_i \cap CARC_{i,n}$ for every $n \geq 1$, so that $r_i \in \operatorname{proj}_{S_i}(R_i \cap C_i \cap CARC_i)$ and thus, condition (iii) of SAS is satisfied too.

For the second part set SAS Q. By definition of SAS (and Pearce, 1984), for each $s_i \in Q_i$, there exist a $\mu_i^1(s_i), \mu_i^2(s_i) \in \Delta(S_{-i})$ such that $\operatorname{supp} \mu_i^1(s_i) = S_{-i}$ and $\operatorname{supp} \mu_i^2(s_i) = Q_{-i}$. By Lemma D.4 of Brandenburger et al. (2008) choose $\mu_i^1(s_i)$ so $r_i \in BR_i(\mu_i^1(s_i))$ if and only if $r_i \in \operatorname{supp} \sigma_i$ for a mixed strategy σ_i with $U_i(s_{-i}; \sigma_i) = U_i(s_{-i}; s_i)$ for every s_{-i}.

Now, consider the set of types $T_i := \{t_i(s_i) | s_i \in Q_i\} \cup \{\star_i\}$; to get an ambiguous type structure define $M_i(\star_i) \subseteq \Delta(S{-}i \times T_{-i})$ such that there is no $\eta_i \in M_i(\star_i)$ with $\operatorname{supp} \eta_i =$

$S_{-i} \times T_{-i}$. For $s_i \in Q_i$, define,

$$\begin{aligned} Y_i(s_i) := \{(r_i, t_i(s_i)) : &\text{ either } r_i = s_i \text{ or} \\ &\exists \sigma_i \in \Delta(S_i), \text{ such that } r_i \in \operatorname{supp} \sigma_i \text{ and } U_i(s_{-i}; \sigma_i) = U_i(s_{-i}; s_i) \text{ for all } s_{-i} \in S_{-i}\} \end{aligned}$$

and then define two beliefs $\eta_i^1(s_i), \eta_i^2(s_i) \in \Delta(S_{-i} \times T_{-i})$ such that

$$\begin{aligned} &\operatorname{supp} \eta_i^1(s_i) = S_{-i} \times T_{-i} && \text{and} && \operatorname*{marg}_{S_{-i}} \eta_i^1(s_i) = \mu_i^1(s_i), \\ &\operatorname{supp} \eta_i^2(s_i) = \prod_{j \neq i} \cup_{s_j \in Q_j} Y_j(s_j) \cap R_j && \text{and} && \operatorname*{marg}_{S_{-i}} \eta_i^2(s_i) = \mu_i^2(s_i). \end{aligned}$$

To complete the description of the type structure, set $M_i(t_i(s_i))$ to be the convex hull of $\eta_i^1(s_i)$ and $\eta_i^2(s_i)$. Note that R_i only depends on the marginal beliefs on the strategies, so for $\eta_i^2(s_i)$ to be well-defined the following is required:

Claim 1: $\operatorname{proj}_{S_i} \bigcup_{s_i \in Q_i} Y_i(s_i) \cap R_i = Q_i$. If $s_i \in Q_i$, then $(s_i, t_i(s_i)) \in Y_i(s_i)$ and by construction also $(s_i, t_i(s_i)) \in R_i$. Conversely, set $r_i \in \operatorname{proj}_{S_i} \bigcup_{s_i \in Q_i} Y_i(s_i) \cap R_i$. So there exists a $s_i \in Q_i$ such that $(r_i, t_i(s_i)) \in Y_i(s_i) \cap R_i$. If $r_i = s_i$, the proof is complete. If not, then by property (*iii*) of self-admissible sets (and definition of $Y_i(s_i)$), it also follows that $r_i \in Q_i$. ★

Next I prove that the type structure satisfies that:

$$Q = \prod_{i \in I} \operatorname*{proj}_{S_i} (R_i \cap C_i \cap CARC_i).$$

Claim 2: $\bigcup_{s_i \in Q_i} Y_i(s_i) \cap R_i = R_i \cap C_i$. Consider $(r_i, t_i) \in \bigcup_{s_i \in Q_i} Y_i(s_i) \cap R_i$. Then for some $s_i \in Q_i$ it holds that $\eta_i^1(s_i)$ and thus $t_i(s_i)$ is cautious. Conversely, for $(r_i, t_i) \in R_i \cap C_i$ it is needed that $t_i \neq \star_i$ since $\star_i$ is not cautious. Thus, there exists a $s_i \in Q_i$ such that $t_i = t_i(s_i)$. If $s_i = r_i$, the proof is complete. If not, then R_i requires that $r_i \in BR_i(\mu_i^1(s_i))$, which holds if and only if (see above) $r_i \in \operatorname{supp} \sigma_i$ for a mixed strategy σ_i with $U_i(s_{-i}; \sigma_i) = U_i(s_{-i}; s_i)$ for every s_{-i}. Thus, in either case it holds that $(r_i, t_i(s_i)) \in Y_i(s_i)$. ★

Claim 3: $\bigcup_{s_i \in Q_i} Y_i(s_i) \cap R_i = R_i \cap C_i \cap CARC_{i,1}$. $(r_i, t_i) \in \cup_{s_i \in Q_i} Y_i(s_i) \cap R_i$, that is $(r_i, t_i(s_i)) \in Y_i(s_i) \cap R_i$ for some $s_i \in Q_i$. Then, $(r_i, t_i(s_i)) \in CARC_{i,1}$ due to $\eta_i^2(s_i)$. The converse follows from Claim 2. ★

Induction concludes the proof. ■

2.4.2. Iterated assumption and ambiguity

The main distinctive feature of assumption with respect to the usual belief for Bayesian agents, and as in the assumption operator of Brandenburger et al. (2008), is the failure of monotonicity.[20] Whenever a Bayesian agent believes in event E, she also believes in every event F such that $E \subseteq F$: The (Bayesian) belief μ_i that assigns probability one to E assigns probability one to F. This is not the case with my notion of assumption. Type t_i might assume event E via some belief $\mu_i \in M_i(t_i)$ that has full-support on E, but she may fail to assume an event F such that $E \subseteq F$;[21] even if t_i assumed such F, it certainly, could not be via μ_i. Thus, when considering a sequence of nested events such as the finite iterations in

[20] This is also reminiscent of strong belief as defined and studied by Battigalli and Siniscalchi (1999, 2002).
[21] I am implicitly assuming that the topological closure of F contains that of E.

the common assumption events defined above, a single belief can assign probability one to all the events in the sequence simultaneously, but different beliefs are required in order to assume each of them at the same time. This is exactly why the inclusion-exclusion problem arises within a standard Bayesian framework, but it can be resolved within my framework.

In principle there is no reason to consider that the assumption of an event is an expression of cautiousness; for every type there exists *always* an event that is assumed and this simply relates to which specific states play some role in how preference ranks acts. However, the assumption of different nested events is a non-trivial feature that reveals a cautious attitude: Whenever a type assumes two nested events E and F, the preference represented is crucially sensitive to comparisons at every state in E but also to comparisons at every state in the larger event F, in particular to these states outside E. Of course, as mentioned above, the simultaneous assumption of different events necessarily requires belief multiplicity.

2.4.3. (Non-)Emptiness of common assumption of rationality and cautiousness

The canonical epistemic foundation of iterated admissibility in the literature is due to Brandenburger et al. (2008). Their seminal result shows that m rounds of elimination of non-admissible strategies characterize the behavioral implications of rationality and m^{th}-order mutual assumption of rationality for finite m in a model where players' uncertainty is formalized by type structures where types are mapped to lexicographic probability systems. As shown by Blume et al. (1991), lexicographic probability systems arise under a variation of Anscombe and Aumann's (1963) preferences in which the axiom of continuity is relaxed (rather than that of completeness, as in Bewley's (2002) variant). However, Brandenburger et al. (2008) also reveal a vexatious feature of the common assumption case: Their celebrated impossibility result shows that for every generic game, if the type structure is complete and maps types continuously, then common assumption in rationality is empty. Below I also discuss the work by Keisler and Lee (2015), Yang (2015), Lee (2016) and Catonini and De Vito (2018), who propose changes in the formalism that allow for sound epistemic foundations, and compare their results to ours.

Notice first that within my set-up, and for every game G, common assumption in rationality and cautiousness is never empty in complete ambiguous type structures. The intuition behind the claim is easy to see: For each iteration in player i's reasoning process set a belief $\mu_i^n \in \Delta(S_{-i} \times T_{-i})$ that has full-support on the topological closure of $\prod_{j \neq i} R_j \cap CARC_{j,n}$ (these collections of strategy-type pairs are clearly never empty; thus, the belief μ_i^n always exists). Then, define M_i as the topological closure of the convex hull of $\{\mu_i^n\}_{n \in \mathbb{N}}$, and by virtue of the ambiguous type structure being complete, pick type t_i with ambiguous beliefs M_i.[22] By construction, t_i is a type representing common assumption of rationality and cautiousness and hence, $CARC_i$ is non-empty.

[22]As shown by Ahn (2007), this assignment can take place in an ambiguous type structure that maps types to ambiguous beliefs continuously.

Furthermore, as briefly mentioned in Section 2.1, the non-emptiness of rationality and common assumption thereof does not follow from specific alterations in the formalism (beyond the different decision-theoretic model underlying the approach). This is easier to visualize by direct comparison with other studies that also provide sound foundations for iterated admissibility. Keisler and Lee (2015) obtain their result by dropping the requirement that types are mapped continuously, Yang (2015) considers a weaker version of assumption than that in Brandenburger et al. (2008), and Lee (2016) explicitly imposes coherence on the preferences, which is usually only checked for the beliefs that represent the preferences. For lexicographic probability systems, which he builds on, this makes a difference. As said, I do not require any of these modifications: My type structures map types continuously, my notion of assumption is a direct adaption of that in Brandenburger et al. (2008) and Dekel et al. (2016),[23] and the coherence requirement implicit in my type structures resembles the standard one in literature due to Brandenburger and Dekel (1993).[24] Finally, Catonini and De Vito (2018) consider a weaker version of the likeliness-ordering of events that characterizes the lexicographic probability system and an alternative version of cautiousness where only the payoff-relevant aspect of the states of the world play any role. Again (and despite Theorem 7 and Theorem 8 would remain unchanged under this alternative notion of cautiousness), I obtain my non-emptiness result with a standard, purely decision-theoretic notion of cautiousness that does not require any specific structure of the set of states.

[23] See also Footnote 17.

[24] The requirement is explicit in the construction by Ahn (2007).

To end this section, I present a comparison between lexicographic probability systems and ambiguous beliefs that provides some understanding of the differences between the two approaches with respect to the presence of ambiguity. Remember that a lexicographic probability system consists of a finite sequence beliefs $\{\mu^k\}_{k=1}^n \subseteq \Delta(\Theta)$,[25] where the order of the sequence represents the epistemic priority attached to each element: μ^1 is the decision maker's 'primary' hypothesis, μ^2 is the 'secondary' hypothesis, and so on. This is reflected by the lexicographic consideration, i.e. if act f is better than g for belief μ^1, then the comparison between the two acts for the rest of the beliefs in the sequence is immaterial and the decision maker prefers f to g. The main distinction between lexicographic probability systems and ambiguous beliefs is then clear: Both are composed of multiple beliefs, but the former incorporates a hierarchy in terms of epistemic priority and hence removes any trace of ambiguity. However, as shown above, this hierarchy is not important to overcome the inclusion-exclusion problem; what is important is the *multiplicity* of beliefs.

2.5. Conclusions

Cautiousness in games is intuitively understood as the idea that even when a player deems some of her opponents' strategies to be completely unlikely (typically on the basis of strategic reasoning), she still prefers to choose strategies that are immune to deviations towards such unexpected strategies. This is at odds with the strategically sophisticated

[25] Brandenburger et al. (2008) use lexicographic conditional probability systems, but their result extends to more general lexicographic probability systems as shown by Dekel et al. (2016).

expected utility maximization process representing a standard Bayesian rational decision maker who believes her opponent to be rational too: Every suboptimal strategy of the latter is assigned zero probability by the subjective belief of the former, and cannot therefore affect the decision process.

This chapter proposes a new theoretic understanding of cautiousness in interactive settings that reconciles it with strategic sophistication. I interpret cautiousness under strategic sophistication as a manifestation of robustness to ambiguity, which renders more choices as non-optimal. Then I show that the resulting behavioral implications can be obtained as a consequence of rationality and related higher-order assumption constraints. Specifically:

(*i*) I introduce the possibility of ambiguity in beliefs by allowing players' preferences to be incomplete. This is done by replacing the standard Anscombe and Aumann (1963) decision-theoretic framework behind each player with a model of (possibly) incomplete preferences à la Bewley (2002) so that each player's uncertainty about her opponents' behavior is represented by a possibly non-singleton set of beliefs that reflects the decision maker's possibly ambiguous uncertainty. My main result implies that for choices that are iteratively admissible the justifying set of beliefs has to be non-singleton for non-trivial games.

(*ii*) I apply the framework described above to study the epistemic (i.e. reasoning-based) foundations of iterated admissibility in belief-complete type structures and find that it characterizes the behavioral implications of rationality, cautiousness, and common assumption thereof (Theorem 7). For non-complete type structures I find that it is self-admissible sets that characterize the behavioral implications of such an event (Theorem 8).

Thus, the main insight is immediately apparent: The inclusion-exclusion problem of Samuelson (1992) can be resolved not only by relaxing continuity of preferences (i.e. through lexicographic probability systems), but also by relaxing completeness (while maintaining continuity). Notably, this enables me to provide a sound epistemic foundation of iterated admissibility—a challenging task within the framework of lexicographic probability systems. Using my approach, it is easy to see that the event of rationality, cautiousness, and common assumption thereof is non-empty across all games—unlike, for instance, the foundations for iterated admissibility under lexicographic probability systems, as found by Brandenburger et al. (2008), and the instruments involved in my characterization (type structures and assumption operators) are straightforward generalizations of the instruments in the realm of standard Bayesian preferences. In addition, the suggested link between ambiguity via incomplete preferences and the presence of cautiousness is potentially testable by applying techniques for the identification of incompleteness of preferences recently developed in the literature on experimental economics (see Cettolin and Riedl, 2019).

CHAPTER 3

Adversarial Bilateral Informational Design

"Simplicity is prerequisite for reliability."

Edsger W. Dijkstra (1975)

3.1. Introduction

Information provision and bilateral contracting are ubiquitous in today's economy. For example, contract research organizations (CROs) provide information to downstream firms (called sponsors), which are typically pharmaceutical or biotechnology companies. CROs do so mainly by conducting clinical trials, but also by utilizing their internal healthcare data in combination with data science. By providing this information, CROs are an integral part of the pharmaceutical and biotechnology industry. The global CRO market was valued at almost $35 billion in 2018 and is projected to reach about $55 billion in 2025. (Grand View Research, 2019)

Sponsors, such as pharmaceutical companies, engage with CROs to outsource part of the drug development. If an agreement is reached, the contract specifies which trials the CRO will conduct for the given sponsor, but not which trials are performed for other sponsors. This is a typical example of bilateral contracting: the contract is contingent only on events that can be verified by *both* of the involved parties. The largest CROs generate most of the revenue of the industry, so it is common for sponsors of the same CRO to be direct competitors. For example, Pfizer and Novartis, are clients of the same CRO, even as they seek to develop similar products. (Ibid.)

Leaving aside details of specific industries, three considerations are crucial for any information provision organization determining what information to provide to clients. First, the provider effectively *commits* to deliver specific information to a given client in a contract. For example, a contract will specify exactly which medical tests will be conducted. Second, the *bilateral* nature of *contracting* excludes commitment to a grand information structure shared with all clients. That is, a contract will only state which tests will be conducted for a specific sponsor and will not state which tests will be performed for other sponsors.[1] Third, the receivers' use of the information is determined within an interactive setting. Therefore, a receiver faces *strategic uncertainty* and needs to reason about what information other receivers get. Crucially, the details of this reasoning process are usually unknown to the information provider. For example, the decision for one sponsor to conduct further research on a drug depends on whether the sponsor believes its competitors are also developing a competing drug and, if so, what information the sponsor believes its competitors are receiving.

In this chapter, I provide a general, yet tractable, method for examining how an information provider determines which information to supply bilaterally to multiple receivers, taking into consideration each of the three aspects outlines above. In particular, motivated by the severity of strategic uncertainty, I take an adversarial approach which ensures robustness to details of receivers' strategic reasoning and is tractable. That is, the information provided to one receiver is required to be optimal for the designer no matter how that receiver thinks about the information other receivers may get. The adversarial approach adopted here ensures that the supplied information is optimal even if nature "chooses" the receiver's reasoning that is least advantageous to the provider.

First, I formalize the issue of robustness to the receivers' reasoning. From a CRO's point of view, I provide a precise answer to the following question: given that a pharmaceutical

[1] Contracts do not specify such details for several reasons. First, CROs have reputational concerns. If CROs disclose which trials they were conducting for a sponsor's competitor, the CRO might reveal the competitor's private information, undermining the CRO's relationship with the competitor. Second, a contract that is contingent on every trial conducted for every sponsor is complex and lacks enforceability. These reasons are broadly applicable and do not only affect CROs. In particular, the second point was raised by McAfee and Schwartz (1994) regarding any supplier that deals with multiple downstream firms.

sponsor gets some information about their drug, how does the pharmaceutical sponsor decide whether to bring the drug to the market or, for example, drop the project altogether? As noted above, sponsors face strategic uncertainty because they do not know what information their competitors have access to. This section's primary contribution is to provide a solution concept that captures this kind of uncertainty. The key insight is that the reasoning about the competitors' information can be sidestepped: to form a best-reply the competitors' information is not relevant, but only the beliefs about the state of nature and the competitor's action matter. For this, a characterization of "rational" competitor's action for any information structure is needed: all *belief-free rationalizable* actions. Furthermore, I demonstrate that this solution concept depends only upon players' first-order beliefs about the payoff state. For a CRO, this means that the solution concept depends only on the information a sponsor receives about their own drug, but not on how a sponsor thinks about the information its competitors have.

Second, I contribute to the foundations of information design with multiple receivers. Mathevet, Perego, and Taneva (2020, p.2) describe information design as "an exercise in belief manipulation;" therefore, it is crucial to characterize which beliefs can be induced by a designer. If there is only one receiver, it is well known that there is only one restriction on the distribution of beliefs about the state of nature. The average belief under this distribution is equal to the prior—a requirement deemed *Bayes plausibility* by Kamenica and Gentzkow (2011). This chapter extends this characterization to multiple receivers (cf. Theorem 9). Furthermore, I provide necessary bounds on the dependence of beliefs if there are two receivers (cf. Proposition 6). These bounds are reminiscent of, but usually tighter than, the Fréchet-Hoeffding bounds known from copulas in probability theory and statistics.[2] Moreover, these bounds are novel not only for information-design and the economics literature more generally, but—to the best of my knowledge—to probability theory as well and provide tractability because they are related to the supermodular stochastic ordering. Even more tractability is gained when more assumptions are put on the primitives, which in particular include supermodular games. I illustrate this in a stylized version of the problem faced by a CRO.

The remainder of the chapter is organized as follows: the next subsections elaborate on related literature and provide the setting for the stylized model of a CRO, which will be used as a running example throughout the chapter. Section 3.2 develops the solution concept. The main representation theorem for the general design problem is formalized in Section 3.3. Section 3.4 studies the case of pure persuasion, which includes the derivation of the belief-dependence bounds and the solution to the CRO model. In Section 3.5, I discuss some extensions and highlight issues related to interpretations of the model. Section 3.6 concludes.

[2] A standard reference for copulas including the Fréchet-Hoeffding bounds is Nelsen (2006).

3.1.1. Related Literature

This chapter is related to several strands of the literature: a solution concept capturing a notion of robustness, general information design, and adversarial and bilateral design. In this section, I discuss these three strands in detail.

Informational Robustness and Solution Concepts

Harsanyi's (1967) theory of games with incomplete information is partially motivated by the possibility that players' information may not be common knowledge. The solution concept I develop in this chapter is directly inspired by the literature on informational robustness which later extended Harsanyi's insights to allow uncertainty about the information structure itself from an outside oberserver's perspective. Early pioneers in this area include Aumann (1987), Brandenburger and Dekel (1987), and Forges (1993, 2006). Bergemann and Morris (2013, 2016) recently exploited the full power of informational robustness to provide robust predictions in economic environments with uncertainty. Within this subset of the literature, the conceptual idea of my proposed solution concept is closest to that of Bergemann and Morris (2017). Proposition 4 is directly inspired by their Section 4.5.[3] However, for the actual solution concept and its foundation in Proposition 5 there is major conceptual difference: Bergemann and Morris are interested in robustness over all information structures from the perspective of an outside observer,[4] while the

[3]Battigalli (1999, 2003) and Battigalli and Siniscalchi (2003) obtain simlar results from an interim perspective.

[4]Some of these ideas are fruitfully applied to the theory of robust mechanism design as initiated by Bergemann and Morris (2005, 2009, 2011). Relatedly, Artemov, Kunimoto, and Serrano (2013) study robust

proposed solution concept in this chapter focuses on the notion of robustness from a player's perspective. This allows sharper predictions because a player considers parts of the information structure that an outside observer does not know. In this vein, Börgers and Li (2019) use a related solution concept to define strategic simplicity. However, these authors do not assume common belief in rationality and also do not provide a foundation for their solution concept.[5]

Other papers dealing with related ideas about robustness (usually from an outside observer) include Battigalli (1999, 2003), Battigalli and Siniscalchi (2003), Dekel et al. (2007), Liu (2015), Tang (2015), and Germano and Zuazo-Garin (2017). In each of these papers players have symmetric knowledge about the information structure. Either the full information structure is commonly known, or no (common) knowledge about the information structure is assumed at all. In my case, there is no assumption about common knowledge of the information structure, but each player knows *her* own information structure. As discussed in more detail in Subsection 3.5.2, the main solution in this chapter concept can be given an epistemic foundation by simply modifying the arguments introduced by Battigalli and Siniscalchi (2003, 2007) and developed further in Battigalli, Di Tillio, Grillo, and Penta (2011).

---

mechanism design when the designer knows that the first-order beliefs belong to a specific set of beliefs. In contrast to my appraoch, the (sets of) first-order beliefs are common knowledge among the players in their setting. A similar approach was considered by Ollár and Penta (2017).

[5]Taking this individual perspective is rather uncommon in recent developments in game theory. Notable exceptions are recent developments in epistemic game theory. Dekel and Siniscalchi (2015) provide a modern overview, whereas Perea (2012) gives a textbook treatment highlighting the indiviudal perspective explicitly. Furthermore, Aumann and Dreze (2008) explicitly study the implications of common knowledge of rationality and a common prior in a complete information game from an individual perspective.

Information Design

The literature on information design originated from contributions of Calzolari and Pavan (2006), Bergemann and Pesendorfer (2007), Brocas and Carrillo (2007), and Eső and Szentes (2007). Since then the literature has grown rapdily. The interested reader is referred to two recent reviews by Bergemann and Morris (2019) and Kamenica (2019). Here I highlight papers that are more closely related to this one, which provide general methods to analyze information design as this chapter does. The seminal paper pertaining to a single receiver is Kamenica and Gentzkow (2011) which illustrates the usefulness of the concavification approach for information design. Regarding multiple receivers, Taneva (2019) uses a Myersonian approach, exploiting a version of the revelation principle, which can be interpreted as a akin to partial implementation known from mechanism design.

The closest work on information design is the recent article by Mathevet et al. (2020). Like Taneva (2019), Mathevet et al. consider information design in cases when the designer has the power to commit to the provision of a grand information structure. However, for a given grand information structure, they allow for the case of adversarial equilibrium selection. Thus, their approach is reminiscent of full implementation in mechanism design. They show that attaining robustness to equilibrium selection requires constructing the full *hierarchy of beliefs* for each receiver.[6] My approach is complementary to theirs. In my setting, strategic uncertainty arises from the bilateral contracting environment which

[6] Similar to the full implementation literature the revelation does not apply in Mathevet et al.'s (2020) setting either.

excludes commitment to a grand information structure. Therefore, in my case the designer is not only concerned about equilibrium selection, but also about strategic uncertainty. My proposed solution concept reflects this more general robustness concern. In addition, I show that my robust solution concept depends only on induced first-order beliefs. Therefore, it is not necessary to induce a full hierarchy of beliefs, but it suffices to look at first-order beliefs only. Thus, the approach I propose is closer in spirit to Kamenica and Gentzkow (2011): since they consider a single receiver, by definition only first-order beliefs matter. However, in the present chapter there are multiple receivers and therefore a new characterization in terms of distributions of first-order beliefs is needed. This is the main result of Section 3.3.

Recent and independent work by Arieli, Babichenko, Sandomirskiy, and Tamuz (2020) studies the question of which distributions over (first-order) beliefs can be induced by information structures in the case of binary states of nature.[7] They provide a full characterization of these distributions for two receivers and extend this characterization to multiple receivers in spirit of No Trade Theorems. In Section 3.4, I provide bounds on the dependence structure across two receivers for these distributions, which are necessary, but not sufficient. Under more stringent assumptions (which include the CRO model of Subsection 3.1.2) my bounds become equivalent to the conditions of Arieli et al. (2020) and are therefore also sufficient.

[7] Arieli et al. also consider applications to social persuasion, i.e. persuading multiple receivers in a non-strategic setting.

Adversarial and Bilateral Design

A few recent studies employ an adversarial approach to information design:[8] Carroll (2016), Goldstein and Huang (2016), Inostroza and Pavan (2018), Hoshino (2019),[9] and ans Daisuke Oyama and Takahashi (2020). All apply the adversarial selection for a solution concept that relies on a grand information structure. In this chapter the adversarial selection is more severe because of the additional robustness coming from the bilateral contracting environment. Recently, Dworczak and Pavan (2020) study adversarial information design using a reduced-form approach that could arise from robustness to equilibrium selection or, like in this chapter, from bilateral contracting. Bilateral information design with or without adversarial robustness is, to the best of my knowledge, new to this chapter.

In a recent review, Carroll (2019) discusses adversarial selection aspects in mechanism design. Bilateral contracting has a long history in economics and has been studied extensively in industrial organization.[10] The relevant paper from this body of literature is Dequiedt and Martimort (2015). Dequiedt and Martimort examine bilateral mechanism design when the designer cannot commit to a grand mechanism. My analysis shares the motivation for analyzing a setting with limited commitment with Dequiedt and Martimort. They overcome the limited commitment by imposing appropriate ex-post incentive constraints on side of the principal. In equilibrium, these ex-post constraints determine all

[8] That is, in addition to Mathevet et al. (2020) as mentioned above.

[9] I thank Nageeb Ali for making me aware of Hoshino's paper.

[10] The interested reader is referred to two handbook chapters: Bresnahan and Levin (2012) and Segal and Whinston (2012).

beliefs of the agents including how they think about other agents' contracts. My approach resolves the limited-commitment issue in a different way. In my model, the designer does not assume that all beliefs are in equilibrium and therefore needs consider the reasoning of the receivers. By taking an adversarial approach, the designer circumvents these issues and seeks an information structure that is robust to the reasoning of the receivers.[11]

3.1.2. Leading Example: A Stylized CRO Model[12]

Consider a situation where a CRO conducts medical trials for two pharmaceutical companies called Pfizr (P) and Novarty (N).[13] Both work on developing similar breast cancer drugs. For simplicity, suppose that each drug could be either effective, or ineffective, and one drug is effective if and only if the other drug is effective. Thus, there are two states of nature, i.e. $\Theta = \{0, 1\}$ representing an ineffective drug and an effective drug, respectively. Furthermore, there are two possible actions the pharmaceutical companies can take: either conduct further research (R), or drop the project (D). Profits (i.e. payoffs) are such that, if firms knew the effectiveness of the drug, they would like to conduct research if and only if the drug is actually effective. However, if a pharmaceutical company decides to conduct further research, its payoff will be lower if the competitor also conducts further research. The reduction in payoffs could be caused by lower expected profits in the future, because

[11] In this sense, the literature on mechanism design without or with limited commitment is also related.

[12] For readers familiar with information design, this section can be skipped. However, in the main analysis I will refer back to this example to illustrate some of the results.

[13] These companies and names are purely fictional.

the competitor's drug is likely to be on the market. The following payoff tables represent such a situation.

		Novarty R	Novarty D
Pfizr	R	1, 1	2, 0
	D	0, 2	0, 0

$\theta = 1$ (effective)

		Novarty R	Novarty D
Pfizr	R	-2, -2	-1, 0
	D	0, -1	0, 0

$\theta = 0$ (ineffective)

For any belief (about the state of nature) that puts probability greater than $2/3$ on the state in which the drug is effective ($\theta = 1$),[14] R is the dominant action. Similarly, for any belief less than $1/3$, the dominant action becomes D. For intermediate beliefs about θ, the best action depends beliefs about competitors' actions. Formal analysis in this chapter shows that these predictions are exactly these which are robust to the reasoning about the information of the competitor. For example, if Pfizr assigns probability close to one to $\theta = 1$, then it does not matter what information Novarty gets and Pfizr should conduct further research. However, if the probability of $\theta = 1$ is $1/2$, Novarty's information matters. To see this, consider the Novarty medical trials, conducted by a CRO, that reveal with high probability that the drug is ineffective. In such a case, Novarty will drop the project with high probability too. This implies that Pfizr should conduct more research (given their belief about θ). On the other hand, if the medical trials for Novarty are such that there is a high likelihood of revealing that the drug is effective, then Novarty is likely conducting research and Pfizr should drop the project (again given their belief about θ). Thus, Pfizr's

[14] Henceforth, I will always associate *belief* with the probability of the state being $\theta = 1$.

beliefs about Novarty's information matter. Therefore, if robustness is a concern, the CRO should take both actions, R and D, into account.

By providing information to the pharmaceutical companies, the CRO can effectively influence the actions taken by the pharmaceutical companies. For example, a natural assumption is that the CRO prefers further research rather than dropping the project, because of the likelihood that further research will include subsequent trials for the CRO to conduct. The goal of this chapter is to provide a tractable method for solving for the optimal provision of information in such settings. In the remainder of this subsection, I highlight some specific information structures that are part of the CRO's choice set.

Suppose that both pharmaceutical companies have a prior belief that assigns probability $1/3$ to the drugs being effective. A trivial choice of the CRO would be to provide no information. In this case and similar to the explanation above, $\{R, D\}$ is the robust prediction for both receivers. Thus, under adversarial selection, the CRO expects both companies to drop the project, which would be the worst possible outcome from the CRO's perspective. Another possibility would be for the CRO to provide full information to each pharmaceutical company. In this case, each company will conduct further research if and only if their drug is effective. Overall, there will be further research (by both firms) with probability equal to the prior, i.e. slightly above 33%. However, the CRO could increase the probability of further research by providing information that does not fully reveal the effectiveness of the drugs.

For illustrative purposes, consider first a case where the CRO can actually commit to a grand information structure and therefore does not have to worry about what conjectures the receivers form about their competitor.[15] This problem can be analyzed with tools provided by Bergemann and Morris (2016) and Taneva (2019) and the solution provides an upper bound for the CRO under the bilateral-contracting assumptions of interest.[16] Consider the following information structure, where both companies get one of two possible reports: either the trial reveals that the drug is ineffective (bad news, b) or the trial suggests the drug is effective but without fully proving the drug's efficacy (good news, g). The reports are generated according to the distribution shown in Table 3.1.[17]

Table 3.1: Optimal Information with Full Commitment.

		Report for Novarty			
		$\theta = 1$		$\theta = 0$	
		b	g	b	g
Report for Pfizr	b	0	0	0	1/2
	g	0	1	1/2	0

For example, when getting the good news, Pfizr will update its belief to get a posterior of 1/2, but since the designer committed to the grand information structure Pfizr knows even more: Novarty will get bad news with probability 1/3, which is higher than the ex-ante probability of bad news, equal to 1/6. Furthermore, Pfizr also knows how the

[15]With commitment to a grand information structure, Pfizr would exactly know what information Novarty gets. That is, not the exact realization (i.e. the result of the trial), but the information structure overall (i.e. which trials will be conducted).

[16]Applying the more robust method akin to full implementation of Mathevet et al. (2020) yields the same result for this example.

[17]The information structure in Table 3.1 is optimal for a designer with symmetric, increasing, and submodular preferences, i.e. $v(R, D) = v(D, R)$, $v(R, \cdot) \geq v(D, \cdot)$, and $v(R, R) + v(D, D) \leq v(D, R) + v(R, D)$.

state describing the effectiveness of the drugs correlates with the Novarty reports. This reasoning about Novarty's reports is crucial because under these assumptions, a unique Bayes-Nash equilibrium exists,[18] where the receivers conduct further research if and only if they receive good news. Thus, with full commitment to a grand information structure the designer can ensure that at least one company will conduct further research with certainty, while both will conduct research with probability equal to the prior belief of $1/3$.

However, the CRO cannot actually commit to the grand information structure. Due to the bilateral-contracting assumption, the CRO can only commit to the marginal distributions and the receivers have to reason about the competitors' information. For example, if the CRO adopts the above information structure, Pfizr could nevertheless conjecture that Novarty does not obtain any useful information from the CRO. For the information structure based on this conjecture, a Bayes-Nash equilibrium exists wherein Pfizr will drop the project given *either* report.[19] Novarty could reason similarly. If the CRO is concerned about adversarial selection, then the CRO's worst-case scenario results in both pharmaceutical companies dropping the project. The question then becomes, is there a way to get these companies to conduct further research given that only bilateral contracting is possible and the designer is concerned about adversarial selection?[20]

[18]The equilibrium action profile is also the unique interim-correlated rationalizable profile.

[19]In this conjectured equilibrium, Novarty would conduct research, but this does not matter for the rest of the analysis.

[20]The arguments in this paragraph relate to a foundation I give in Subsection 3.2.2 for the solution concept developed in Section 3.2.

Table 3.2: Optimal Information for Adversarial Bilateral design.

		Report for Novarty			
		$\theta = 1$		$\theta = 0$	
		b	g	b	g
Report for Pfizr	b	0	0	1/2	1/4
	g	0	1	1/4	0

A positive answer is provided by the robust information structure described in Table 3.2.[21] This information structure reduces the overall probability of the good report from 2/3 to 1/2. Now, after receiving the good report the posterior is 2/3, which makes R a dominant action. Thus, each report now has a unique dominant action[22] and the conjecture about the competitor's information no longer plays a role. The optimal information structure exactly balances the *trade-off* between inducing posteriors that are robust to receivers' conjecture about the information of their competitor and making further research as likely as possible. However, to achieve this, the proposed robust information structure reduces the probability of at least one receiver conducting further research to 2/3.[23] Therefore, the CRO suffers a loss of about 33 percent that at least one company will conduct further research relative to the optimal full commitment information structure. This is the loss due to the constraints of bilateral contracting.

[21] As before, this information structure is optimal for the same preferences as stated in Footnote 17.

[22] With the exact posterior of 2/3 both actions are still undominated. Therefore, the induced posterior should be $2/3 + \varepsilon$ for some small $\varepsilon > 0$. This example ignores this tie-breaking issue here. The full theory presented below does account for this.

[23] Even with this robust information structure both receivers will conduct further research with probability of 1/3.

3.2. A Robust Solution Concept

This section develops a solution concept that delivers predictions that are robust in the sense that they depend on what information the player receives about the economic fundamental, but do not depend on how the player reasons about information other players might receive. I refer to these predictions as *individual robust predictions* and the corresponding solution concept is developed in two stages. The first stage builds on the concept of belief-free rationalizability (see Battigalli et al., 2011).[24] This version of rationalizability is robust to any information any player might get. Thus, this stage corresponds to robustness across information structure from an outside observer. For the purposes of this chapter, this solution concept is too extreme since it does not take into account any information that a player gets about the state of nature, which describes, for example, the effectiveness of a drug. The second stage of the solution concept adds exactly this information, therefore refining belief-free rationalizability. I argue that this new solution concept reflects the robust prediction given that a player knows his/her information about the state of nature.

There are two players $i \in N := \{1, 2\}$, who will be also called receivers.[25] Each player has a finite set of actions A_i and as usual $A = A_1 \times A_2$ denotes the set of action profiles.[26]

[24] Battigalli (1999, 2003) and Battigalli and Siniscalchi (2003) introduce a more general class of versions of rationalizability. One instance corresponds to belief-free rationalizability.

[25] This section is concerned only with the predictions of receivers' actions for the given information structure. The sender/designer does not play a role and will be introduced later.

[26] I follow the standard notation that for a fixed player i, A_{-i} denotes the set of actions for the other player $3 - i$. More generally, I use this notation for any player-specific sets.

Uncertainty is modeled via a finite set of *states of nature* denoted by Θ. Each agent's preferences are represented by a utility function $u_i : A \times \Theta \to \mathbb{R}$. All these components form an *economic environment* $\mathcal{E} = \langle \Theta, (A_i, u_i)_{i \in N} \rangle$,[27] which is assumed to be common knowledge.

Example 3. The economic environment for the CRO example is succinctly described by the two payoff tables specified in Subsection 3.1.2. ◇

The economic environment does not specify any information the players might have. Most solution concepts need a specification of the information structure. However, Battigalli et al. (2011) provide a solution concept—*belief-free rationalizability*—that depends only on the economic environment, capturing the exact behavioral implications of (correct) common belief in rationality.[28] Formally, action a_i is *belief-free rationalizable* if $a_i \in BFR_i := \bigcap_{n \geq 0} BFR_i^n$, where $BFR_i^0 := A_i$ and inductively for any $n \in \mathbb{N}$

$$BFR_i^n(\theta_i) := \left\{ a_i \in BFR_i^{n-1} \;\middle|\; \begin{array}{ll} \text{There exists some } \mu_i \in \Delta(\Theta \times A_{-i}) \text{ such that:} & \\ (i) \quad \operatorname{supp} \mu_i \subseteq \Theta \times BFR_{-i}^{n-1}, & \\ (ii) \quad a_i \in BR_i(\mu_i) & \end{array} \right\}. \qquad (BFR^n)$$

According to the usual arguments (e.g. Wald, 1949; Pearce, 1984), this procedure is the same as deleting ex-post dominated actions iteratively. An action $a_i \in A_i$ is ex-post

[27] This is different from a basic game which is widely used in information design (see e.g. Bergemann and Morris, 2013; Mathevet et al., 2020). The difference is that a basic game also specifies a common prior on the states of nature.

[28] Bergemann and Morris (2017) also mention this solution concept, but they call it ex post rationalizability. They also define a notion of belief-free rationalizability, which is stronger than the version used here.

dominated (relative to $X_{-i} \subseteq A_{-i}$), if there exists $\alpha_i \in \Delta(A_i)$ such that

$$\sum_{a_i'} \alpha(a_i') u_i(a_i', a_{-i}, \theta) > u_i(a_i, a_{-i}, \theta), \quad \text{for all } (a_{-i}, \theta) \in X_{-i} \times \Theta.$$

Example 4. In the CRO example from Subsection 3.1.2 it is easy to see that no action is ex-post dominated; hence $BFR_i = A_i$. ◇

As mentioned at the beginning of this section, belief-free rationalizability only takes the economic environment and rationality as primitive objects. In the current situation, a player has some information about the state of nature which affects his/her *individual robust predictions*.[29] Thus, Player 1 is assumed have a prior $\pi_1 \in \Delta(\Theta)$ and gets some information about the state of nature, which is described by a marginal information structure.[30]

Definition 4. *Fix an economic environment $\mathcal{E}$. A* marginal information structure (for $\mathcal{E}$) *is $I_1 = \langle S_1, \psi_1 \rangle$, where*

1. *S_1 is a finite set of signals, and*

2. *$\psi_1 : \Theta \to \Delta(S_1)$ is a conditional signal distribution.*

This marginal information structure does not specify any possible signals for the other player, nor does it it specify the signal distribution for the other player. Thus, this marginal

[29] The remainder of this section describes the perspective of Player 1. To apply it to Player 2, switch the player indicies.

[30] A marginal information structure is equivalent to a statistical experiment as introduced by Blackwell (1951, 1953). The restriction to finite signals might not be without loss, but it remains an open question whether it is. In this section, I also assume that each signal realization $s_1 \in S_1$ has (ex-ante) positive probability. This can be relaxed at the cost of more cumbersome notation. See Subsection 3.2.2.

information structure provides information *only* about the state of nature. The solution concept depends only on the marginal information structure.[31] This solution concept will be a set of pure strategies denoted by $R_1(I_1, \pi_1) \subseteq A_1^{S_1}$ and is formally defined as follows.

Each signal realization $s_1 \in S_1$ induces a posterior belief[32] $\mu_{s_1} \in \Delta(\Theta)$ by Bayesian updating:

$$\mu_{s_1}(\theta) := \frac{\psi_1(s_1|\theta)\pi(\theta)}{\sum_{\theta'} \psi_1(s_1|\theta')\pi(\theta')}. \tag{3.1}$$

Since these signals only induce a belief about the state of nature θ, these beliefs are not rich enough to form a best-reply in an interactive setting. To form a best-reply, beliefs about the actions of the other player are also needed. A *rational-extended* belief incorporates this additional requirement by assigning positive probability only to the belief-free rationalizable actions of the other player.

Definition 5. *Fix an economic environment $\mathcal{E}$, a prior $\pi_1 \in \Delta(\Theta)$ and a marginal information structure I_1. A* rational-extended belief *for $s_1 \in S_1$ is a belief $\tilde{\mu}_1 \in \Delta(\Theta \times A_2)$ such that* (*i*) $\operatorname{marg}_\Theta \tilde{\mu}_1 = \mu_{s_1}$ *as given by Equation 3.1 and* (*ii*) $\operatorname{supp} \tilde{\mu}_1 \subseteq \Theta \times BFR_2$. *Let $\mathcal{M}_1 : S_1 \rightrightarrows \Delta(\Theta \times A_2)$ denote the set of rational-extended beliefs for each $s_1 \in S_1$, i.e.*

$$\mathcal{M}_1(s_1) = \{\tilde{\mu} \in \Delta(\Theta \times A_2) : \tilde{\mu} \textit{ is a rational-extended belief for } s_1\}.$$

[31] Similar to before, the solution concept also depends on the economic environment, but this dependence will be implicit.

[32] To save on notation, the player's index is kept implicit by using the signals' index.

Finally, these rational-extended beliefs allow me to define the individual robust prediction.

Definition 6. *Fix an economic environment $\mathcal{E}$, a prior $\pi_1 \in \Delta(\Theta)$, and a marginal information structure I_1. A pure strategy $b : S_1 \to A_1$ is* conceivable *for (π_1, I_1) if b is optimal for at least one selection of $\mathcal{M}_1$, i.e. b is optimal given μ_1, i.e. for each $s_1 \in S_1$, there exists $\tilde{\mu}_1 \in \mathcal{M}_i(s_1)$ such that*

$$b(s_1) \in \arg\max_{a_1' \in A_1} \sum_{\theta, a_2} \tilde{\mu}_1(\theta, a_2) u_1(a_1', a_2, \theta).$$

The individual robust prediction *is the set of all conceivable strategies and is denoted by $R_1(I_1, \pi_1)$.*

A foundation in terms of explicit epistemic assumptions is discussed Subsection 3.5.2: the individual robust prediction corresponds to the behavioral implications of *common* knowledge of the economic environment, *common* belief in rationality, and *knowledge* of the marginal information structure. Thus, the prediction does not rely on implicit or explicit *common* knowledge assumptions about the marginal information structure. This is relevant for later questions about information design. The nature of bilateral contracting allows the designer to only commit to a marginal information structure. The receiver understands this marginal information, but needs to reason about what actions their opponent chooses. This reasoning process is not transparent to the designer. Thus, all actions the designer can rule out are exactly these strategies that are not part of the individual robust prediction. This is the essence of Definition 6.

In Subsection 3.2.2, I provide another foundation of this solution concept in terms of informational robustness and Bayes-Nash equilibirum analyis similar in spirit to Bergemann and Morris (2013, 2016, 2017). This foundation relies on a theory of how player's resolve uncertainty about the grand information structure: each player conjectures a grand information structure consistent with their marginal information structure. Given this conjecture, each player chooses a strategy as predicted by a Bayes-Nash equilibrium. The individual robust predictions correspond to the union across all such conjectures and all corresponding equilibria. Independently of the foundations, the robust predictions are often simple to calculate as the following example shows.

Example 5. Table 3.3 shows the marginal information for Pfizr induced by the full commitment optimal information structure described in Table 3.1. The bad report leads to a

Table 3.3: Pfizr's marginal information derived from the information structure of Table 3.1.

		$\theta = 1$	$\theta = 0$
Report for Pfizer	b	0	$1/2$
	g	1	$1/2$

posterior[33] of zero, whereas the good report induces a posterior belief of $1/2$. Example 4 established that all actions are belief-free rationalizable. Thus, the sets of rational-extended

[33]Recall that within this example beliefs correspond to the likelihood of the state of the drug being effective ($\theta = 1$).

beliefs for each signal are given by:

$$\mathcal{M}_P(b) = \{\tilde{\mu} \in \Delta(\Theta \times A_N) : \tilde{\mu}(1, R) + \tilde{\mu}(1, D) = 0\}, \text{ and}$$
$$\mathcal{M}_P(g) = \{\tilde{\mu} \in \Delta(\Theta \times A_N) : \tilde{\mu}(1, R) + \tilde{\mu}(1, D) = 1/2\}.$$

Since Research (R) is a dominated action if the drug is ineffective, R cannot be part of the individual robust prediction for the bad report. However, for the good report both actions are conceivable. For example, D is a best-reply to $\mu(1, R) = 1 - \mu(0, R) = 1/2$, whereas R is a best-reply $\mu(1, D) = 1 - \mu(0, D) = 1/2$. Both beliefs are valid rational-extended belief for the good signal. Thus, the individual robust prediction for Pfizr is

$$R_P(\text{Table 3.3}, 1/3) = \{(D, D), (D, R)\},$$

where the first coordinate indicates the action after the bad report, and the second coordinate corresponds to the good report. ◇

3.2.1. The interim perspective

Thus far the solution concept has been stated from an ex-ante perspective, which is relevant for later questions about information design question. However, it will also be useful to have the solution concept in an interim form. This is done by defining a correspondence

$R_1(\cdot|I_1, \pi_1) : S_1 \rightrightarrows A_1$ as

$$R_1(s_1|I_1, \pi_1) := \{a_1 \in A_1 : \exists b \in R_1(I_1, \pi_1) \text{ s.t. } a_1 = b(s_1)\}.$$

The interim individual robust prediction relies only on the belief about the state of nature that is induced by the signal. Thus, the solution concept does not depend on the (marginal) information structure it is defined for, but only on the posteriors it generates. Moreover, the robust predictions can be strategically distinguished by changing the economic environment. The following proposition formalizes these simple observations, which will be useful to address the information-design question.

Proposition 3. *Fix a set of states of nature Θ. Consider an economic environment $\mathcal{E}$ (with states of nature given by Θ), two priors $\pi_1, \pi_1' \in \Delta(\Theta)$ and two marginal information structures $I_1 = \langle S_1, \psi_1 \rangle$ and $I_1' = \langle S_1', \psi_1' \rangle$. For all $(s_1, s_1') \in S_1 \times S_1'$, if $\mu_{s_1} = \mu_{s_1'}$, then $R_1(s_1|I_1, \pi_1) = R_1(s_1'|I_1', \pi_1')$.*

Conversely, consider two priors $\pi_1, \pi_1' \in \Delta(\Theta)$ and two marginal information structures $I_1 = \langle S_1, \psi_1 \rangle$ and $I_1' = \langle S_1', \psi_1' \rangle$. If there exists $(s_1, s_1') \in S_1 \times S_1'$ and $\theta \in \Theta$ such that $\mu_{s_1}(\theta) \neq \mu_{s_1'}(\theta)$ then there exists a (finite) economic environment (holding Θ fixed) such that $R_1(s_1|I_1, \pi_1) \cap R_1(s_1'|I_1', \pi_1') = \emptyset$.

Proof. The statement is trivial if $|\Theta| = 1$, so suppose $|\Theta| > 1$.

The first part follows directly from the definition, since BFR_i depends only on the economic environment and the rational-extended beliefs exactly capture only the beliefs about the states of nature, which are the same by assumption.

For the second part, fix $\theta' \in \Theta$ such that

$$\mu := \frac{\psi_1(s_1|\theta')\pi_1(\theta')}{\sum_\theta \psi_1(s_1|\theta)\pi_1(\theta)} \neq \frac{\psi_1'(s_1'|\theta')\pi_1'(\theta')}{\sum_\theta \psi_1'(s_1'|\theta)\pi_1'(\theta)} =: \mu'.$$

Consider the following economic environment: $A_i = \{\mu, \mu'\}$ and payoffs are given by

$$u_i(a_i, a_{-i}, \theta) = (a_i - 1[\theta = \theta'])^2.$$

By construction only the belief about the state matters for best-replies, so the difference between the induced belief on Θ and an rational-extended belief does not matter. Now, note that μ (as action) is the unique best-reply to μ (as belief). Then, by construction $R_1(s_1|I_1, \pi_1) = \{\mu\}$ and $R_1(s_1'|I_1', \pi_1') = \{\mu'\}$ and the conclusion follows. ■

With Proposition 3 in mind,[34] I abuse notation for the interim version of the solution concept and write it as a correspondence defined on belief space, i.e. $R_1 : \Delta(\Theta) \rightrightarrows A_1$. Thus, R_1 denotes the ex-ante version, whereas $R_1(\mu_1)$ denotes the interim version. The interim notion is illustrated by applying it to the CRO example.

Example 6. Due to the binary state space, the interim individual robust predictions (defined on belief space) can be illustrated by means of a simple diagram. Figure 3.1 shows these predictions for both companies, where, a belief corresponds to the probability of the drug being effective. It was already argued in the introduction, that for beliefs greater than

[34]As stated the proposition requires that every signal happens with positive probability. If any signals have zero ex-ante probability, then the proposition needs to be adjusted to condition on positive probability signals only.

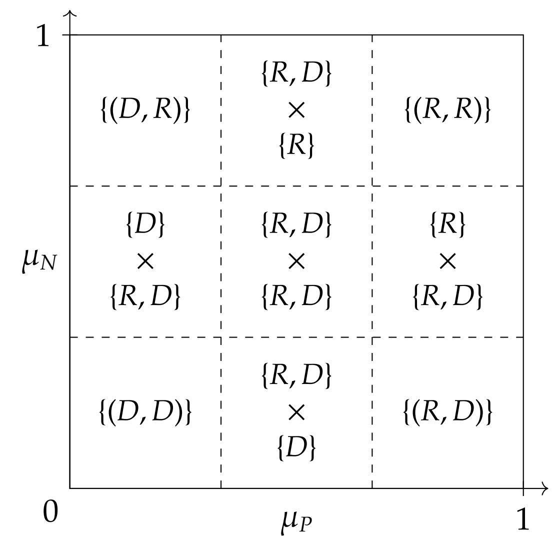

Figure 3.1: Individual robust predictions of the CRO game.

$^{2}/_{3}$ R is uniquely undominated, whereas for beliefs lower than $^{1}/_{3}$ D is the only dominant action. For all intermediate beliefs, a similar argument as in the previous example can establish that both actions are the individual robust prediction. ◇

3.2.2. A Foundation for the Individual Robust Predictions

In this section, I provide a foundation for the individual robust predictions. To achieve this from an ex-ante perspective zero-probability events have to be explicitly be accounted for. This changes some of the previous definitions slightly at the cost of more burdensome notation. Whenever zero-probability events can be ruled out, all the following definitions reduce to the previous definitions.

Robustness for an Outside Observer

Starting with an economic environment, a Bayesian game is obtained by adding priors for each player $\pi_i \in \Delta(\Theta)$ and specifying a (grand) information structure with possible heterogeneous signal functions.

Definition 7. *Fix an economic environment $\mathcal{E}$. A (grand)* generalized information structure (for $\mathcal{E}$) *is $I = \langle (S_i, \Psi_i)_{i \in N} \rangle$, where for each player $i \in N$,*

1. *S_i is a finite set of signals, and*
2. *$\Psi_i : \Theta \to \Delta(S_1 \times S_2)$ is a conditional signal distribution.*

A Bayesian game *$G = \langle \mathcal{E}, I, (\pi_i)_{i \in N} \rangle$ is given by (i) an economic environment $\mathcal{E}$, (ii) a generalized information structure I, and (iii) a prior $\pi_i \in \Delta(\Theta)$ for each player $i \in N$.*

A generlized information structure together with the two priors gives rises to a standard type structure á la Harsanyi (1967) but without a common prior.[35] Without common priors and signal distributions the definition of equilibrium needs to account for zero probability events. For complete information games, Brandenburger and Dekel (1987) introduced *a posteriori equilibrium* to rule out the play of dominated actions after a zero probability events. The definition of equilibrium in this chapter will be an extension to incorporate uncertainty about the states of nature. But first, I need to introduce a tool to define beliefs even in case of zero probability events.

[35]Recently, Piermont and Zuazo-Garin (2020) allow for even more disagreement by allowing for lack of common knowledge of the Harsanyi type structure and the states of nature.

Definition 8. *Fix an economic environment $\mathcal{E}$, a player i, a prior $\pi_i \in \Delta(\Theta)$ and a generalized information structure I. A* conditional probability system (CPS) *for (π_i, I) is a mapping $\mu_i : S_i \to \Delta(\Theta \times S_{-i})$ such that for every $(\theta, s_i, s_{-i}) \in \Theta \times S_1 \times S_2$,*

$$\mu_i(\theta, s_{-i}|s_i)\left[\sum_{\theta', s'_{-i}} \pi_i(\theta')\Psi_i(s_i, s'_{-i}|\theta')\right] = \pi_i(\theta)\Psi_i(s_i, s_{-i}|\theta).$$

That is, a CPS defines beliefs about the state of nature and the opponent's signal realization for every signal relation of the given player. In addition, the beliefs have to be updated via Bayes' rule whenever possible. To formally state the appropriate version of equilibrium, it only remains to define strategies. A (behavioral) strategy for player i in a Bayesian Game G is a mapping $\beta_i : S_i \to \Delta(A_i)$.

Definition 9. *Fix an economic environment $\mathcal{E}$, priors $\pi_i \in \Delta(\Theta)$ for each player $i \in N$, and an information structure I. A* Bayes-Nash equilibrium (BNE) *for (π_1, π_2, I) is a tuple (β_i, μ_i) for each player $i \in I$ such that*

1. *β_i is a strategy,*

2. *μ_i is a CPS for (π_i, I), and*

3. *β_i is optimal (given μ_i and β_{-i}), i.e. for each $s_i \in S_i$*

$$a_i \in \operatorname{supp} \beta_i(\cdot|s_i) \implies a_i \in \arg\max_{a'_i} \sum_{\theta, s_{-i}, a_{-i}} \mu_i(\theta, s_{-i}|s_i)\beta_{-i}(a_{-i}|s_{-i})u_i(a'_i, a_{-i}, \theta).$$

Let $BNE(\pi_1, \pi_2, I)$ be the set of all BNEs for (π_1, π_2, I).[36]

For some of the formal proofs below it will be useful to have an equivalent fixed-point definition of belief-free rationalizability (Equation BFR^n). Usual arguments can be used to show the following equivalent fixed-point definition:[37] for every player i consider a set of actions $F_i \subseteq A_i$ with the following fixed-point property

$$F_i := \left\{ a_i \in A_i \;\middle|\; \begin{array}{l} \text{There exists some } \mu_i \in \Delta(\Theta \times A_{-i}) \text{ such that:} \\ \\ (i) \quad \operatorname{supp} \mu_i \subseteq \Theta \times F_{-i}, \\ \\ (ii) \quad a_i \in BR_i(\mu_i) \end{array} \right\}. \tag{BFR_{FP}}$$

Then the pair $(BFR_i)_{i \in I}$ understood as correspondences is equal to the pair of correspondences $(F_i)_{i \in I}$ satisfying the fixed-point property and are largest by set inclusion.

The first result states that belief-free rationalizability characterizes all actions that can be played in any Bayes-Nash equilibrium for any information structure (and any prior beliefs). Thus, without making any assumptions about the information structure an outside observer can not make any prediction that is a refinement of belief-free rationalizability. In this sense, belief-free rationalizability is robust to the specification of the (generalized) information structure.

[36]The dependence on the economic environment is suppressed in this notation since it will be fixed throughout. Furthermore, I will slightly abuse notation and write $\beta = (\beta_1, \beta_2) \in BNE(\pi_1, \pi_2, I)$ if there exists CPS' $\mu = (\mu_1, \mu_2)$ such that $(\beta, \mu) \in BNE(\pi_1, \pi_2, I)$. Similarly, I will write $\beta_i \in BNE_i(\pi_1, \pi_2, I)$ if there exists β_{-i} and μ such that $(\beta_1, \mu_1, \beta_2, \mu_2) \in BNE(\pi_1, \pi_2, I)$.

[37]Battigalli (2003, Proposition 8.1) establishes this equivalence in a more general setting.

Proposition 4. *Fix an economic environment $\mathcal{E}$. For every player i, $a_i \in BFR_i$ iff there exists priors (π_1, π_2), an information structure I and a signal $s_i \in S_i$ such that $a_i \in \operatorname{supp} \beta_i(\cdot|s_i)$ for some $\beta_i \in BNE_i(\pi_1, \pi_2, I)$.*

Proof. For given priors (π_1, π_2), information structure I, consider a signal s_i such that $a_i \in \operatorname{supp} \beta_i(\cdot|s_i)$ for some $(\beta_i, \hat{\mu}_i, \beta_{-i}, \hat{\mu}_{-i}) \in BNE(\pi_1, \pi_2, I)$. I show that $a_i \in BFR_i$ by induction, i.e. $a_i \in BFR_i^n$ for every n. The statement is trivial for $n = 0$. So assume the statement is true for $n \geq 0$. Consider the following belief $\mu_i \in \Delta(\Theta \times S_{-i} \times A_{-i})$ defined by

$$\mu_i(\theta, s_{-i}, a_{-i}) = \hat{\mu}_i(\theta, s_{-i}|s_i)\beta_{-i}(a_{-i}|s_{-i}),$$

Note that a_i is a best-reply to μ_i by the definition of BNE.

Let $m_i = \operatorname{marg}_{\Theta \times A_{-i}} \mu_i$, then

$$m_i(\theta, a_{-i}) > 0 \implies \mu_i(\theta, s_{-i}, a_{-i}) > 0 \text{ for some } s_{-i} \text{ such that } \beta_{-i}(a_{-i}|s_{-i}) > 0,$$

and by the induction hypothesis $a_{-i} \in BFR_{-i}^n$. Hence, $\operatorname{supp} \mu_i \subseteq \Theta \times BFR_{-i}^n$. Since, a_i is a best-reply to μ_i, $a_i \in BFR_i^{n+1}$.

Conversely, for every $a_i \in BFR_i$, there is a justifying belief $\mu_i^{a_i}$ satisfying (1) and (2) from BFR.[38] Then define a prior by

$$\pi_i(\theta) = \sum_{a_i \in BFR_i} \frac{\sum_{a_{-i}} \mu_i^{a_i}(\theta, a_{-i})}{|BFR_i|}$$

and consider the following information structure: $S_i = BFR_i$ and

$$\Psi_i(a_i, a_{-i}|\theta) = \frac{\mu_i^{a_i}(\theta, a_{-i})}{\pi_i(\theta)} |BFR_i|^{-1},$$

if $\pi_i(\theta) > 0$ and arbitrary otherwise. Note that for every $a_i \in BFR_i$,

$$\sum_{a_{-i},\theta} \pi_i(\theta)\Psi_i(a_i, a_{-i}|\theta) = |BFR_i|^{-1} > 0,$$

so that the CPS is entirely determined by Bayesian updating.

Now, fix $a_i \in BFR$ and consider the obedient strategies, i.e. $\beta_i(a_i|s_i) = 1[s_i = a_i]$. Then,

$$\begin{aligned}
a_i &\in \arg\max_{a_i' \in A_i} \sum_{\theta, a_{-i}} \mu_i^{a_i}(\theta, a_{-i}) u_i(a_i', a_{-i}, \theta) \\
&\in \arg\max_{a_i' \in A_i} \sum_{\theta, a_{-i}} \Psi_i(a_i, a_{-i}|\theta)\pi_i(\theta) u_i(a_i', a_{-i}, \theta) \\
&\in \arg\max_{a_i' \in A_i} \sum_{\theta, a_{-i}, s_{-i}} \pi_i(\theta)\Psi_i(a_i, s_{-i}|\theta)\beta_{-i}(a_{-i}|s_{-i}) u_i(a_i', a_{-i}, \theta),
\end{aligned}$$

[38] Here, the equivalent fixed-point definition of belief-free rationalizability stated as in Equation BFR_{FP} (without any belief restrictions) is used.

so that the obedient strategy of i is indeed a best-reply to the obedient strategy of the other player (given the information structure). That is, β (and the CPS derived from Bayesian updating) constitute a BNE. ■

Robustness from the Player's Perspective[39]

Now, I add back the marginal information structure of Player 1 (see Definition 4). Here as well, zero probability events need to be taken care of and therefore rational-extended beliefs are not appropriate anymore. A version of a conditional probability system is needed again. Although, now it should only capture beliefs about the state of nature.

Definition 10. *Fix an economic environment $\mathcal{E}$, a prior $\pi_1 \in \Delta(\Theta)$ and a marginal information structure I_1. A* marginal conditional probability system (mCPS) *for (π_1, I_1) is a mapping $\mu_1 : S_1 \to \Delta(\Theta)$ such that for every $(\theta, s_1) \in \Theta \times S_1$,*

$$\mu_1(\theta|s_1)\left[\sum_{\theta'} \pi_1(\theta')\psi_1(s_1|\theta')\right] = \pi_1(\theta)\psi_1(s_1|\theta).$$

Similar to rational-extended beliefs, mCPS need to be extended as well.

Definition 11. *Fix an economic environment $\mathcal{E}$, a prior $\pi_1 \in \Delta(\Theta)$ and a marginal information structure I_1. A* rational-extended conditional probability system (rCPS) *for (π_1, I_1) is a mapping $\mu_1 : S_1 \to \Delta(\Theta \times A_2)$ such that*

[39] This subsection is described from the perspective of player 1. It applies verbatim to player 2 by switching the player indices.

1. $\tilde{\mu}_1 = (\mu_1(\cdot|s_1))_{s_1\in S_1}$ *is a mCPS for* (π_1, I_1), *where* $\tilde{\mu}_1(\cdot|s_1) = \mathrm{marg}_\Theta\, \mu_1(\cdot|s_1)$ *for all* $s_1 \in S_1$, *and*

2. *for all* $s_1 \in S_1$, $\mathrm{supp}\, \mu_1(\cdot|s_1) \subseteq \Theta \times BFR_2$.

Finally, these rCPS' allow to define the individual robust prediction even with zero probability events.

Definition 12. *Fix an economic environment* $\mathcal{E}$, *a prior* $\pi_1 \in \Delta(\Theta)$, *and a marginal information structure* I_1. *A pure strategy* $b : S_1 \to A_1$ *is* conceivable *for* (π_1, I_1) *if there exists a rCPS* μ_1 *for* (π_1, I_1) *such that* b *is optimal given* μ_1, *i.e. for each* $s_1 \in S_1$,

$$b(s_1) \in \arg\max_{a_1'} \sum_{\theta, a_2} \mu_1(\theta, a_2|s_1) u_1(a_1', a_2, \theta).$$

The individual robust prediction *is the set of all conceivable strategies and is denoted by* $R_1(I_1, \pi_1)$.

The goal of this section is to provide a foundation of the individual robust predictions. That is, it should capture they idea of informational robustness across all information structures of the opponent (fixing the marginal information structure of the player). This leads to the idea of an extended information structure.

Definition 13. *Fix an economic environment* $\mathcal{E}$ *and a marginal information structure* $I_1 = \langle S_1, \psi_1 \rangle$. *An* extended information structure *(for* I_1*) is* $I = \langle (\hat{S}_i, \Psi_i)_{i\in N} \rangle$ *such that*

1. *I is a generalized information structure,*

2. $S_1 \subseteq \hat{S}_1$, *and*

3. $\operatorname{marg}_{S_1} \Psi_1(\cdot|\theta) = \psi_1(\cdot|\theta)$, *for all* $\theta \in \Theta$.

Let $\mathcal{I}(I_1)$ *be the set of extending information structures for* I_1.

Condition (1) ensures that an extended information structure is indeed a generalized information structure, whereas conditions (2) and (3) make sure that the extended information structure incorporates the marginal information structure of Player 1. A natural interpretation of this definition is that Player 1 conjectures a grand information structure for given economic environment so that she can analyze the resulting Bayesian game. However, since she knows exactly what information she gets about the state of nature, she uses this knowledge to rule out information structures which do not align with her marginal information structure. Indeed, the individual robust prediction correspond to all strategies that are conceivable across all such conjectures. This means that for each conceivable strategy there is an extending information structure (and a conjectured prior for the opponent)[40] and a corresponding Bayes-Nash equilibrium where this strategy is played.

Proposition 5. *Fix an economic environment* $\mathcal{E}$, *prior* $\pi_1 \in \Delta(\Theta)$, *and a marginal information structure* I_1. $b \in R_1(I_1, \pi_1)$ *if and only if there exists an extending information structure* $I \in \mathcal{I}(I_1)$, *a prior* $\pi_2 \in \Delta(\Theta)$, *and a corresponding BNE* β_i *such that* $b(s_i) \in \operatorname{supp} \beta_i(\cdot|s_i)$ *for all* $s_i \in S_i$.

Proof. Fix an economic environment $\mathcal{E}$, prior $\pi_1 \in \Delta(\Theta)$, and a marginal information structure I_1.

[40] Recall that the economic environment does not specify priors.

For a given extending information structure $I \in \mathcal{I}(I_1)$, a prior π_2, and a corresponding BNE $(\beta, \hat{\mu})$ consider any selection $b(s_1) \in \operatorname{supp} \beta_i(\cdot|s_1)$ for all $s_1 \in S_1$. For every $s_2 \in \hat{S}_2$ and every $a_2 \in \operatorname{supp} \beta_2(\cdot|s_2)$, $a_2 \in BFR_2$ by Proposition 4. For each $s_1 \in S_1$ consider beliefs $\mu_1(\cdot|s_1) \in \Delta(\Theta \times A_2)$ defined by

$$\mu_1(\theta, a_2|s_1) = \sum_{\hat{s}_2} \hat{\mu}_1(\theta, \hat{s}_2|s_1)\beta_2(a_2|\hat{s}_2).$$

Then $\mu_1(\theta, a_2|s_1) > 0$ implies that there exists $s_2 \in \hat{S}_2$ such that $\beta_2(a_2|s_2) > 0$, which implies that $a_2 \in BFR_2$. Hence, $\operatorname{supp} \mu_1(\cdot|s_1) \subseteq \Theta \times BFR_2$ for every $s_1 \in S_1$. Furthermore, let $\tilde{\mu}_1(\cdot|s_1) = \sum_{a_2} \mu_1(\cdot, a_2|s_1)$ for every $s_1 \in S_1$, then

$$\begin{aligned}
\tilde{\mu}_1(\theta|s_1)\left[\sum_{\theta'} \pi_1(\theta')\psi_1(s_1|\theta')\right] &= \sum_{a_2,\hat{s}_2} \hat{\mu}_1(\theta, \hat{s}_2|s_1)\beta_2(a_2|\hat{s}_2)\left[\sum_{\theta'} \pi_1(\theta')\psi_1(s_1|\theta')\right] \\
&= \sum_{\hat{s}_2} \hat{\mu}_1(\theta, \hat{s}_2|s_1)\left[\sum_{\theta'} \pi_1(\theta')\psi_1(s_1|\theta')\right] \\
&= \sum_{\hat{s}_2} \hat{\mu}_1(\theta, \hat{s}_2|s_1)\left[\sum_{\theta', \hat{s}_2'} \pi_1(\theta')\Psi_1(s_1, \hat{s}_2'|\theta')\right] \\
&= \sum_{\hat{s}_2} \pi_1(\theta)\Psi_1(s_1, \hat{s}_2|\theta) = \pi_1(\theta)\psi_1(s_1|\theta),
\end{aligned}$$

where the third and last equality use property 3 of an extending information structure (Definition 13). The fourth equality follows from $\hat{\mu}_1$ being a CPS for (π_1, I) (see Definition 8). Thus, μ_1 is a rCPS and by construction $b(s_i)$ is a best-reply to $\mu_1(\cdot|s_1)$ for each $s_1 \in S_1$. This proves that b is conceivable.

Conversely, consider $b \in R_1(I_1, \pi_1)$. By definition of R_1 there exists a rCPS μ_1 such that b is optimal given μ_1. Define $BFR_1^- = BFR_1 \setminus \cup_{s_1 \in S_1}\{b(s_1)\}$ and set $\hat{S}_1 = S_1 \cup BFR_1^-$ and $\hat{S}_2 = BFR_2$.

For player 1, define a conditional signal distribution as follows.

$$\Psi_1(s_1, \hat{s}_2|\theta) = \frac{\mu_1(\hat{s}_2, \theta|s_1)}{\pi_1(\theta)} \sum_{\tilde{\theta}} \pi_1(\tilde{\theta})\psi_1(s_1|\tilde{\theta}), \quad \text{for all } s_1 \in S_1, \text{ and}$$

$$\Psi_1(a_1, \hat{s}_2|\theta) = 0 \quad \text{for all } a_1 \in BFR_1^-,$$

if $\pi_i(\theta) > 0$ and arbitrary otherwise. Since the marginal of μ_1 on θ is a mCPS, it holds that $\operatorname{marg}_{S_1} \Psi_1 = \psi_1$.

Since $\hat{S}_2 \subseteq BFR_2$, there is a belief $\mu_2^{a_2}$ satisfying (1) and (2) from BFR[41] for each $a_2 \in \hat{S}_2$. Then define a prior by

$$\pi_2(\theta) = \sum_{a_2 \in \hat{S}_2} \frac{\sum_{a_1} \mu_2^{a_2}(\theta, a_1)}{|\hat{S}_2|}$$

and consider the following conditional signal distribution for player 2.

$$\Psi_2(s_1, a_2|\theta) = \frac{1}{|\hat{S}_2|} \frac{1}{|b^{-1}(b(s_1))|} \frac{\mu_2^{a_2}(b(s_1), \theta)}{\pi_2(\theta)}, \quad \text{for all } s_1 \in S_1, \text{ and}$$

$$\Psi_2(a_1, a_2|\theta) = \frac{1}{|\hat{S}_2|} \frac{\mu_2^{a_2}(a_1, \theta)}{\pi_2(\theta)}, \quad \text{for all } a_1 \in BFR_1^-,$$

[41] Again, the equivalent fixed-point definition of BFR stated in Equation BFR_{FP} (without any belief restriction) is used.

if $\pi_2(\theta) > 0$ and arbitrary otherwise.

Since $\sum_{\hat{s}_1,\theta} \Psi_2(\hat{s}_1, s_2|\theta)\pi_2(\theta) = |\hat{S}_2|^{-1} > 0$ for all $s_2 \in \hat{S}_2$, the CPS for player 2 is determined by Bayesian updating. For player 1, consider the CPS that is defined by Bayesian updating if $\sum_{\tilde{\theta},\hat{s}_2} \pi_1(\tilde{\theta})\Psi_1(s_1, \hat{s}_2|\tilde{\theta}) = \sum_{\tilde{\theta}} \pi_1(\tilde{\theta})\psi_1(s_1|\tilde{\theta}) > 0$ and in the other case for $s_1 \in S_1$ define

$$\hat{\mu}_1(\theta, \hat{s}_2|s_1) = \sum_{a_2} \mu_1(\theta, a_2|s_1)1[a_2 = \hat{s}_2].$$

For $a_1 \in BFR_1^-$ there exists a justifying BFR belief $\mu_1^{a_1} \in \Delta(\Theta \times A_2)$, so take as a CPS belief[42]

$$\hat{\mu}_1(\theta, \hat{s}_2|s_1) = \sum_{a_2} \mu_1^{a_1}(\theta, a_2)1[a_2 = \hat{s}_2].$$

Now, consider the obedient strategies $\beta_1(b(s_1)|s_1) = 1$ if $s_1 \in S_1$, $\beta_1(a_1|a_1) = 1$ if $a_1 \in BFR_1^-$, and $\beta_2(a_2|a_2) = 1$ for every $a_2 \in \hat{S}_2$. It remains to verify that these strategies are optimal given the CPS (and the strategy of the opponent).

Player 1 For every $s_1 \in S_1$ with $\sum_{\tilde{\theta},\hat{s}_2} \pi_1(\tilde{\theta})\Psi_1(s_1, \hat{s}_2|\tilde{\theta}) > 0$

$$\begin{aligned}
b(s_1) &\in \arg\max_{a_1' \in A_1} \sum_{\theta,a_2} \mu_1(\theta, a_2|s_1)u_1(a_1', a_2, \theta) \\
&\in \arg\max_{a_i' \in A_i} \sum_{\theta,a_2} \Psi_1(s_1, a_2|\theta)\pi_1(\theta)u_1(a_1', a_2, \theta) \\
&\in \arg\max_{a_i' \in A_i} \sum_{\theta,a_2,\hat{s}_2} \pi_1(\theta)\Psi_1(s_1, \hat{s}_2|\theta)\beta_2(a_2|\hat{s}_2)u_1(a_1', a_2, \theta),
\end{aligned}$$

[42]By construction, these a_1 have zero probability under the signal distributions of player 1.

where the second line uses the definition of the signal distribution and the belief in the last line is (equivalent to) the updated belief together with belief in the strategy of the other player.

For every $s_i \in S_i$ with $\sum_{\tilde{\theta},\hat{s}_2} \pi_1(\tilde{\theta})\Psi_1(s_1,\hat{s}_2|\tilde{\theta}) = 0$,

$$\begin{aligned}
b(s_1) &\in \arg\max_{a'_1\in A_1} \sum_{\theta,a_2} \mu_1(\theta,a_2|s_1)u_1(a'_1,a_2,\theta) \\
&\in \arg\max_{a'_i\in A_i} \sum_{\theta,a_2,\hat{s}_2} \mu_1(\theta,a_2|s_1)1[a_2=\hat{s}_2]u_1(a'_1,\hat{s}_2,\theta) \\
&\in \arg\max_{a'_i\in A_i} \sum_{\theta,\hat{s}_2,a_2} \hat{\mu}_1(\theta,\hat{s}_2|s_1)\beta_2(a_2|\hat{s}_2)u_1(a'_1,a_2,\theta).
\end{aligned}$$

Like in the last case, for every $a_1 \in BFR_1^-$ a_1 is a best-reply to $\hat{\mu}_1$ and β_2.

Player 2 For every $a_2 \in \hat{S}_2$

$$\begin{aligned}
a_2 &\in \arg\max_{a'_2\in A_2} \sum_{\theta,a_1} \mu_2^{a_2}(\theta,a_1)u_2(a_1,a'_2,\theta) \\
&\in \arg\max_{a'_2\in A_2} \sum_{\theta} \left[\sum_{a_1\in\{b(s_1)\}_{s_1}} \mu_2^{a_2}(\theta,a_1)u_2(a_1,a'_2,\theta) + \sum_{a_1\in BFR_1^-} \mu_2^{a_2}(\theta,a_1)u_i(a_1,a'_2,\theta) \right] \\
&\in \arg\max_{a'_2\in A_2} \sum_{\theta} \left[\sum_{s_1\in S_1} \frac{\mu_2^{a_2}(\theta,b(s_1))}{|b^{-1}(b(s_1))|} u_2(b(s_1),a'_2,\theta) + \sum_{a_1\in BFR_1^-} \mu_2^{a_2}(\theta,a_1)u_2(a_1,a'_2,\theta) \right] \\
&\in \arg\max_{a'_2\in A_2} \sum_{\theta} \sum_{\hat{s}_1\in\hat{S}_1,a_1} \pi_2(\theta)\psi_2(\hat{s}_1,a_2|\theta)\beta_1(a_1|\hat{s}_1)u_2(a_1,a'_2,\theta).
\end{aligned}$$

So that β (together with the constructed CPS) is indeed a BNE. ∎

Proposition 5 constitutes the main result of this section, because it provides an informational robustness foundation for the individual robust predictions.

3.3. Adversarial Bilateral Information Design

3.3.1. The General Problem

The previous section prepared the stage to address the question of information design with bilateral contracting. Due to the nature of bilateral contracting, receivers' behavior is not uniquely predicted and the information designer is concerned about robustness to this uncertainty. For this, the previous section introduced a solution concept that captures robust predictions of receivers' actions. Crucially, this solution concept depends only on the receiver's belief about the states of nature. This feature produces a general representation theorem for information design with an adversarial and bilateral aspect.

To formally address the design question, the economic environment $\mathcal{E}$ needs to be appended with the preferences of the designer (she) $v : A \times \Theta \to \mathbb{R}$, which describes the utility she gets if the receivers take actions $a = (a_1, a_2)$. Furthermore, I assume that she knows the receivers' priors, and that these priors are the same as her prior, i.e. $\pi_1 = \pi_2 = \pi \in \Delta(\Theta)$.[43] Given this assumption, it is without loss to assume that the prior

[43] The assumption says the designer knows the prior of the receivers, which happens to be the same prior. It does not state that players know the prior of their opponent, i.e. there is no *common* prior. Relaxing the assumption of the designer knowing the receivers' priors is active research even for the single receiver case. See, for example, Beauchêne, Li, and Li (2019), Kosterina (2019), and Pahlke (2019). Heterogeneous priors with the same support can be incorporated along the lines of Alonso and Câmara (2016). If priors with different supports are allowed, an extension is not straightforward. Galperti (2019) addresses some of the

has full support. Together these components form a *design environment* $\mathcal{D} = \langle \mathcal{E}, \pi, v \rangle$. In such an environment, the designer chooses (grand) information structures, which specifies signals and distributions over signals for both receivers:

Definition 14. *Fix an economic environment $\mathcal{E}$. A (grand)* information structure (for $\mathcal{E}$) *is $I = \langle (S_1, S_2), \Psi \rangle$, where for each player $i \in N$,*

1. *S_i is a finite set of signals, and*

2. *$\Psi_i : \Theta \to \Delta(S_1 \times S_2)$ is a conditional signal distribution.*

Let $\mathcal{I}$ denote the set of information structures (for $\mathcal{E}$).

As before, I assume that each signal happens with positive probability.[44] Additionally, a given information structure I induces a marginal information structure, denoted by $\operatorname{marg}_i I$ (or sometimes just I_i—no confusion should arise), by marginalization. That is,

$$\psi_i(\cdot|\theta) = \operatorname*{marg}_{S_i} \Psi(\cdot|\theta), \text{ for all } \theta \in \Theta,$$

which justifies the naming.

The timeline of the overall design game is as follows and schematically shown in Figure 3.2.

Step 1: Designer chooses an information structure $I \in \mathcal{I}$.

subsequent issues in the case of a single receiver. Applying Galperti's approach to the multiple receivers setting of this chapter seems interesting for future research.

[44]This is without loss in this section.

Step 2: Receivers learn their respective marginal information structure I_i.

Step 3: The state of nature θ realizes and signals (s_1, s_2) are sent according to $\Psi(\cdot|\theta)$.

Step 4: For each signal (s_1, s_2), Nature recommends a conceivable action for each receiver to minimize the payoff of the designer.

Step 5: Each receiver plays as recommended by Nature.

Step 6: Payoffs are realized.

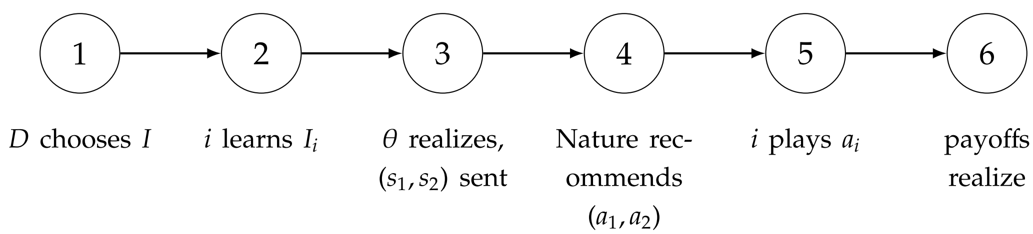

Figure 3.2: Timeline of the design game.

The bilateral contracting assumption is reflected in Step 2: a contract only specifies the marginal information structure for each player. Step 4 corresponds to the adversarial selection of the receivers' actions. Due to bilateral contracts, there might be multiple conceivable actions for each receiver, giving rise to uncertainty as to which actions will be played. Here, the designer is assumed to be very sensitive to this uncertainty and she considers a worst-case scenario.

3.3.2. The Representation

With this timing in mind, the information-design problem can be stated formally as

$$\sup_{I \in \mathcal{I}} V(I),$$

where

$$V(I) := \sum_{\theta \in \Theta, s \in S} \pi(\theta)\psi(s|\theta) \min_{(a_i \in R_i(s_i|I_i,\pi))_{i \in N}} v(a_1, a_2, \theta), \tag{3.2}$$

and recall that I_i is the marginal information structure derived from I.[45] If a maximizer exists,[46] then the resulting information structure captures robustness in the following sense: the optimal information structure performs well no matter how Nature chooses and coordinates the receivers' conceivable actions.

Given the structure of the problem, a natural approach would be to try to use a version of the revelation principle. However, the standard revelation principle argument á la Myerson (1982) does not apply here: this approach requires tie-breaking in favor of the designer. Instead, adversarial selection, by definition, selects actions that are incentive-compatible for the agents and bad for the principal. The following example illustrates

[45] See the discussion after Definition 4.

[46] In general, a maximizer might not exist. The adversarial approach includes tie-breaking against the designer's favor. This can lead to a failure of upper semicontinuity of the objective function.

that such an approach is bound to fail and shows that the problem is even more subtle than the tie-breaking issue. [47]

Example 7. Let $\Theta = \{0, 1\}$ and consider an economic environment, where player 2 has two actions (x and y) and is indifferent between them. Thus, $R_2(\mu_2) = \{x, y\} = A_2$ for any $\mu_2 \in \Delta(\Theta)$. Player 1 has three actions a, b, c and payoffs are given by Table 3.4.

Table 3.4: Payoffs for Player 1.

		Player 2's action			
		$\theta = 1$		$\theta = 0$	
		x	y	x	y
Player 1's action	a	2	0	0	2
	b	3	0	1	0
	c	0	1	0	3

First, b is conceivable for any belief: b is a best-reply if Player 1 is certain that player 2 chooses x. Similarly, c is also always conceivable. For beliefs close to certainty of either state, a is dominated by a mixture of b and c (e.g. in state $\theta = 1$ almost all the weight of the mixture will be on b). However, beliefs around $1/2$ about θ makes a conceivable. For example, suppose the belief about θ is exactly $1/2$, then consider the following rational-extended belief: $\tilde{\mu}(1, x) = \tilde{\mu}(0, y) = 1/2$. For this belief, a is a best-reply. It can be verified that for any belief $\mu \in \Delta(\Theta)$ such that $\mu \in [1/4, 3/4]$ a is conceivable.

Now, consider a designer who only cares about Player 1's action. In particular, assume her (state-independent) preferences are given by $a \prec b \prec c$. Figure 3.3 shows the robust

[47] I am indebted to Marciano Siniscalchi for providing this simple, yet elucidative, example. Inostroza and Pavan (2018, Example 1) illustrate a similar issue when the designer has full commitment.

predictions for Player 1 in belief space and the implied worst-case selection for the designer. For any prior $\pi \in \Delta(\Theta)$ the designer can get her (constrained) best outcome (b) by

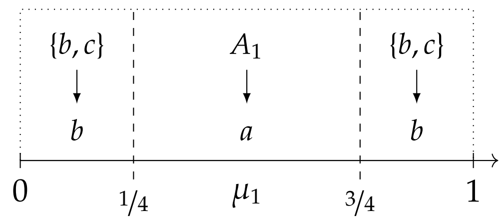

Figure 3.3: Robust Predictions for Player 1 and implied designer's worst-case realization.

fully revealing the state. This optimal payoff cannot be attained with recommendation in general. For example, consider a prior belief of $\pi = 1/2$. A recommendation would send b with certainty. However, this signal does not provide information beyond the prior and therefore the worst-case prediction will be a rather than b as recommended.

The crucial failure is that a revelation principle with some sort of recommendations usually works by pooling signals together. This gives rise to a posterior that is a convex combination of the posteriors derived from each of the pooled signals. However, it is not true that a best-reply to the convex combination is also a best-reply to one of the original posteriors. For example, here, a is a best-reply to a convex combination of beliefs that are certain about a state. For each of these extreme beliefs, a is dominated by either b or c. $\diamond$

Example 7 illustrates that there is no obvious simplification in signal space available that does not use some specific structure of the underlying economic environment. Since the individual robust prediction R_i depends only on the belief induced by the signal (see Proposition 3), the problem can be simplified by working with beliefs directly similarly to the single-receiver case of Kamenica and Gentzkow (2011). However, for multiple receivers, their approach does not readily extend itself because the designer has to address the full hierarchy of beliefs. This approach has been studied by Mathevet et al. (2020).

In the present chapter the information designer can only commit to the marginal information structures because of the bilateral contracting assumption. In this setting, the players know what information they will receiver about the state of nature, but they do not know what information their opponent receives. The individual robust prediction corresponds to such an environment. Thus, the current setting raises a question about which distribution over beliefs can be induced by an information structure.[48]

Like Equation 3.1, Bayesian updating gives rise to receiver i's posterior belief about the state of nature:[49]

$$\mu_{s_i}(\theta) := \frac{\sum_{s_{-i}} \Psi(s_i, s_{-i}|\theta)\pi(\theta)}{\sum_{s_{-i},\theta'} \Psi(s_i, s_{-i}|\theta')\pi(\theta')}. \tag{3.3}$$

Thus, the information structure gives rise to a distribution over beliefs and the state of nature, i.e. an element of $\Delta(\Delta(\Theta) \times \Delta(\Theta) \times \Theta)$. Formally, this distribution τ is given by

$$\lambda(\mu_1, \mu_2, \theta) = \sum_{i \in N} \sum_{s_i : \mu_{s_i} = \mu_i} \pi(\theta)\Psi(s_1, s_2|\theta). \tag{3.4}$$

[48]Indeed, this is an open question in the literature. Ely (2017, p. 47) raises this concern quite directly by stating that "[...] there is no useful generalization for the multi-agent case". However, see also Arieli et al. (2020).

[49]Mechanically, Bayesian updating gives rise to a belief about the other receiver's signals as well. However, as mentioned above only beliefs about the states matter.

Say a distribution over beliefs τ is induced by some information structure, if there exists an information structure such that τ can be derived from the information structure by applying Equation 3.3 and 3.4. Using Proposition 3, the objective from Equation 3.2 can be rewritten as follows:

$$\begin{aligned} V(I) &= \sum_{\theta \in \Theta, s \in S} \pi(\theta)\psi(s|\theta) \min_{\left(a_i \in R_i(\mu_{s_i})\right)_{i \in N}} v(a_1, a_2, \theta) \\ &= \sum_{\mu_1, \mu_2, \theta} \lambda(\mu_1, \mu_2, \theta) \min_{\left(a_i \in R_i(\mu_i)\right)_{i \in N}} v(a_1, a_2, \theta), \end{aligned}$$

where λ corresponds to the distribution over beliefs induced by I. Now, the objective is stated purely in terms of beliefs and the actual information structure no longer plays a role. However, a simplification of the design problem calls for a characterization of a subset of $\Delta(\Delta(\Theta) \times \Delta(\Theta) \times \Theta)$ so that every element of this subset is induced by some information structure.

Obviously, consistency with the prior π requires the marginal of λ on the state space to coincide with π, i.e. $\text{marg}_\Theta \lambda = \pi$. Furthermore, it is well known that another requirement that needs to be satisfied for any distribution over beliefs is that the belief of each player averages out to the prior, i.e. for each $i \in N$

$$\sum_{\mu_1, \mu_2, \theta} \mu_i \lambda(\mu_1, \mu_2, \theta) = \pi. \tag{3.5}$$

Kamenica and Gentzkow (2011) show that this condition is also sufficient to characterize the marginal distribution over beliefs for each player. However, these martingale properties on the marginals are not enough to characterize the possible joint distributions. Intuitively, what is missing are constraints linking together co-movement of beliefs across players.

Table 3.5: λ not induced by any information structure.

λ		μ_2			
		$\theta = 1$		$\theta = 0$	
		0	1	0	1
μ_1	0	0	0	0	1/2
	1	1/2	0	0	0

Example 8. Let $\Theta = \{0, 1\}$ and consider a uniform prior. Table 3.5 states a candidate distribution λ, which satisfies consistency with the prior and satisfies the martingale property for each player. However, no information structure induces such a distribution over beliefs. Intuitively, why no information structure can give rise to such a posterior distribution is easily seen: the extreme posteriors reflect the idea that the information structure fully reveals the state to the receivers. But if this is the case, there is no way to reveal one state to Player 1 and, at the same time, reveal the other state to Player 2. ◇

The following representation theorem takes care of the restrictions across players and follows from a direct-revelation argument in belief space:[50]

[50]Here and henceforth, I will differ slightly from the standard notation by denoting the set of finite support probability measures with $\Delta(X)$ even for the case when X is infinite. This is not restrictive beyond the maintained assumption of finite information structures.

Theorem 9 (Representation Theorem). *Fix a design environment $\mathcal{D}$ and define $\nu(\mu_1, \mu_2, \theta) := \min_{\left(a_i \in R_i(\mu_i)\right)_{i \in N}} v(a_1, a_2, \theta)$. The designer's problem can be represented as*[51]

$$\sup_{I \in \mathcal{I}} V(I) = \sup_{\lambda \in \Delta(\Delta(\Theta)^2 \times \theta)} \sum_{\mu_1, \mu_2, \theta} \lambda(\mu_1, \mu_2, \theta) \nu(\mu_1, \mu_2, \theta)$$

$$\text{s.t.} \quad (1) \ \operatorname*{marg}_{\Theta} \lambda = \pi,$$

$$(2) \ \mathbb{E}_\lambda[\delta_\theta | \mu_i] = \frac{\sum_{\mu_{-i}} \lambda(\mu_i, \mu_{-i}, \cdot)}{\sum_{\mu_{-i}, \theta} \lambda(\mu_i, \mu_{-i}, \theta)} = \mu_i, \quad \text{for every } \mu_i \in \operatorname{supp} \lambda.$$

Furthermore, this restated problem is is a linear program.

Proof. I only proof (1)+(2). The rest is obvious or follows from the previous discussion.

Fix an information structure $I \in \mathcal{I}$ and let $\lambda \in \Delta(\Delta(\Theta) \times \Delta(\Theta) \times \Theta)$ be the induced distribution. Then (1) is satisfied, because

$$\sum_{\mu_1, \mu_2} \lambda(\mu_1, \mu_2, \theta) = \sum_{\mu_1, \mu_2} \left[\sum_{i \in N} \sum_{s_i : \mu_{s_i} = \mu_i} \pi(\theta) \Psi(s_1, s_2 | \theta) \right]$$

$$= \pi(\theta) \sum_{s_1, s_2} \Psi(s_1, s_2 | \theta) = \pi(\theta).$$

[51] Henceforth, for a given set X and any $x \in X$, $\delta_x \in \Delta(X)$ denotes the Dirac measure concentrated at x.

For (2), consider $\mu_1 \in \Delta(\Theta)$. Then,

$$\begin{aligned}
\sum_{\mu_2} \lambda(\mu_1, \mu_2, \theta) &= \sum_{\mu_2} \left[\sum_{i \in N} \sum_{s_i : \mu_{s_i} = \mu_i} \pi(\theta) \Psi(s_1, s_2 | \theta) \right] \\
&= \sum_{s_1 : \mu_{s_1} = \mu_1} \sum_{s_2} \pi(\theta) \Psi(s_1, s_2 | \theta) \\
&= \sum_{s_1 : \mu_{s_1} = \mu_1} \left[\mu_{s_1}(\theta) \sum_{s_2, \theta'} \pi(\theta') \Psi(s_1, s_2 | \theta') \right] \\
&= \mu_1(\theta) \sum_{s_1 : \mu_{s_1} = \mu_1} \left[\sum_{s_2, \theta'} \pi(\theta') \Psi(s_1, s_2 | \theta') \right] \\
&= \mu_1(\theta) \sum_{\mu_2, \theta'} \lambda(\mu_1, \mu_2, \theta').
\end{aligned}$$

The argument for player 2 is the same.

Conversely, suppose there exists λ with conditions (1) and (2), I will construct an information structure which induces λ. For this let $S_1 = \operatorname{supp} \operatorname{marg}_1 \tau$ and $S_2 = \operatorname{supp} \operatorname{marg}_2 \tau$ and define the conditional signal distribution as

$$\Psi(\mu_1, \mu_2 | \theta) = \frac{\lambda(\mu_1, \mu_2, \theta)}{\pi(\theta)}.$$

Note that condition (1) implies that Ψ gives rises to valid distributions.

Furthermore, for signals which happen with positive probability condition (2) gives

$$\mu_{\mu_i}(\theta) = \frac{\sum_{\mu_{-i}} \lambda(\mu_i, \mu_{-i}, \theta)}{\sum_{\mu_{-i}, \tilde{\theta}} \lambda(\mu_i, \mu_{-i}, \theta)} = \mu_i(\theta).$$

Hence,

$$\sum_{i \in N} \sum_{s_i : \mu_{s_i} = \mu_i} \pi(\theta)\Psi(s_1, s_2|\theta) = \pi(\theta)\Psi(\mu_1, \mu_2|\theta) = \lambda(\mu_1, \mu_2, \theta)$$

so that the constructed information structure induces λ. ■

Example 8 (continuing from p. 156)**.** The proposed distribution of Table 3.5 does not satisfy condition (2) of the program states in Theorem 9. To see this, consider the case where Player 2 is certain of $\theta = 0$ (i.e. $\mu_2 = 0$), so that

$$\frac{\sum_{\mu_1} \lambda(\mu_1, 0, 0)}{\sum_{\mu_1, \theta} \lambda(\mu_1, 0, \theta)} = 0 \neq \mu_2(0) = 1 - \mu_2 = 1.$$

This formally illustrates that λ is not induced by any information structure as was intuitively explained before. ◇

The characterization of the distributions over belief in the representation theorem does not make use of the martingale properties (Equation 3.5). Indeed, the two conditions in the theorem imply the martingale condition, because

$$\sum_{\mu_i, \mu_{-i}, \theta} \tau(\mu_i, \mu_{-i}, \theta)\mu_i = \sum_{\mu_i} \mu_i \sum_{\mu_{-i}, \theta} \tau(\mu_i, \mu_{-i}, \theta) \overset{(2)}{=} \sum_{\mu_i} \sum_{\mu_{-i}} \tau(\mu_i, \mu_{-i}, \cdot) \overset{(1)}{=} \pi. \tag{3.6}$$

Furthermore, this characterization is a direct extension of the single-receiver characterization of Kamenica and Gentzkow (2011), i.e. if one receiver does not get any information the martingale condition (Equation 3.5) remains the only constraint for the other receiver.

Corollary 3 (Kamenica and Gentzkow, 2011)**.** *Fix an economic environment $\mathcal{E}$ and a full-support prior $\pi \in \Delta(\Theta)$. Consider $\lambda \in \Delta(\Delta(\Theta) \times \Delta(\Theta) \times \Theta)$ with marginals (on $\Delta(\Theta)$) τ_1 and τ_2 and suppose that (i) $\tau_2 = \delta_\pi$ and (ii) $\mathrm{marg}_\Theta\, \lambda = \pi$. Then, λ is induced by an information structure if and only if $\sum_{\mu_1} \mathbb{E}_{\tau_1}[\mu_1] = \pi$.*

3.4. The Case of Pure Persuasion

So far the general problem allowed for the designer having intrinsic preference on how the information is provided to the receivers. This section will address the case where the designer only cares about the receivers' actions, but does not care about the state of nature herself. Henceforth, with a slight abuse of notation, the designer's preferences are given by $v : A \to \mathbb{R}$. Thus the objective of the designer becomes to maximize

$$\sum_{\theta \in \Theta, s \in S} \pi(\theta)\psi(s|\theta) \min_{(a_i \in R_i(s_i|I_i,\pi))_{i \in N}} v(a_1, a_2).$$

In this case the belief-space approach simplifies the problem even further, because now the distribution over two beliefs of the receivers are a sufficient to calculate the expected value for the designer. For a given a information structure I and the induced distribution over beliefs *and* states $\lambda \in \Delta(\Delta(\Theta) \times \Delta(\Theta) \times \Theta)$, consider the marginal distribution over beliefs alone, i.e. $marg_{1,2}\lambda =: \tau \in \Delta(\Delta(\Theta) \times \Delta(\Theta))$. Similar to above this allows to exploit

Proposition 3 to rewrite the objective in belief space as

$$\sum_{\mu_1,\mu_2} \tau(\mu_1,\mu_2)\nu(\mu_1,\mu_2),$$

where $\nu(\mu_1,\mu_2) := \min_{\left(a_i \in R_i(\mu_i)\right)_{i\in I}} v(a_1,a_2)$. However, working in belief space requires again a characterization of the choice set of the designer; that is a characterization of distributions over receivers' beliefs that can be induced by information structures. However, this is an open problem in the literature. Very recently and independently from my work, Arieli et al. (2020) provide such a characterization for the case of binary states. Their characterization requires a quantification over all subsets of the support of the distributions over beliefs and therefore does not readily yield a simplification of the design problem. Instead, I follow a different route by providing bounds on how dependent the beliefs can be across the two receivers. Although, these bounds turn out to be only a necessary condition for distributions over beliefs to be induced by any information structure, they are tractable and work for any (finite) number of states of nature.[52] Furthermore, these bounds are sufficient under more assumptions about the design environment. One set of such assumptions will be presented later.

3.4.1. Measuring Dependence of Random Variables

A bit more notation is needed to introduce the relevant measure of dependence for random variables that is also relevant when realizations are beliefs. Let $\mathbf{X}$ and $\mathbf{Y}$ be real-valued

[52] Arieli et al. (2020, Appendix B) discuss why my bounds are only necessary.

random variables.[53] distributed according to cumulative distribution functions (CDFs) F_X and F_Y, respectively. Then the Fréchet class $\mathcal{F}(F_X, F_Y)$ is the set of all joint CDFs with marginals given by F_X and F_Y.

Definition 15 (Joe, 1997, Section 2.2.1)**.** *Fix two univariate CDFs F_1 and F_2. Consider $F, F' \in \mathcal{F}(F_1, F_2)$. F' is said to be* more concordant than *F (denoted by $F \precsim F'$) if*[54]

$$F(x, y) \leq F'(x, y),$$

for all $(x, y) \in \mathbb{R}^2$.

Intuitively, this stochastic ordering formalizes the idea that large values happen more often together (across both dimensions) under F' than under F. Furthermore, the Fréchet class $\mathcal{F}$ can be bounded according to this stochastic ordering. That is, for given univariate CDFs F_1 and F_2, for every $F \in \mathcal{F}(F_1, F_1)$, $\underline{F} \precsim F \precsim \overline{F}$, where

$$\underline{F}(x, y) := \max\{0, F_1(x) + F_2(y) - 1\}, \text{ and} \tag{3.7}$$

$$\overline{F}(x, y) := \min\{F_1(x), F_2(y)\}. \tag{3.8}$$

These bounds are often called Fréchet-Hoeffding bounds[55] and they correspond to extremal dependence across the two dimensions. The lower bound corresponds to coun-

[53] The definition readily extends to random variables taking values in a totally ordered set.

[54] This stochastic order is also known as the positive quadrant dependent (PQD) ordering. See, e.g., Shaked and Shanthikumar (2007, Chapter 9).

[55] They were discoverd by Hoeffding (1940) and Fréchet (1951). They play an important role in Copula theory. For more see, for example, Nelsen (2006).

termonotonic random variables (i.e. low realizations in one dimension happen only with high realizations in the other dimension), whereas the upper bound describes comonotonic random variables (i.e. perfect positive dependence). These bounds also describe the extremal dependence for information structures.[56] This is illustrated with the help of the information structures from the CRO example next.

Example 9. Consider the information structures described by Table 3.1 and Table 3.2. Table 3.6 shows their corresponding CDFs.[57] Both CDFs correspond to the lower Fréchet-Hoeffding bound (given their respective marginal distributions).

Table 3.6: CDFs corresponding to the information structures from Subsection 3.1.2.

		Report for Novarty			
		$\theta = 1$		$\theta = 0$	
		b	g	b	g
Report for Pfizr	b	0	0	0	1/2
	g	0	1	1/2	1

CDF for Table 3.1

	Report for Novarty			
	$\theta = 1$		$\theta = 0$	
	b	g	b	g
b	0	0	1/2	3/4
g	0	1	3/4	1

CDF for Table 3.2

With the information from Table 3.6 the upper bound can be obtained by using Equation 3.8. The resulting CDFs are shown in Table 3.7. With these CDFs, the signals are perfectly aligned. For example, the distribution coniditonal on the state $\theta = 0$ defined by the left side of Table 3.7 corresponds to sending the bad report to both receivers with probability 1/2 and sending the good report with the remaining probability of 1/2 to both

[56] For this, the set of individual signals needs be endowed with *any* total order. Recall that information structures are distributions over signals conditional on the state of nature, see Definition 14. If all conditional distributions are equal to their (upper or lower) Fréchet-Hoeffding bound (fixing the conditional marginal distributions), then I say the information structures attains its bound.

[57] Signals are ordered so that g is assumed to be greater than b.

companies. Therefore, both companies will always get the exact same report in the case the drug is ineffective. This is also true for the distribution described on the right side, but in this case the probabilities differ. ◇

Table 3.7: $\overline{F}$ for same marginals as in Table 3.6.

		Report for Novarty					Report for Novarty			
		$\theta = 1$		$\theta = 0$			$\theta = 1$		$\theta = 0$	
		b	g	b	g		b	g	b	g
Report for Pfizr	b	0	0	1/2	1/2	b	0	0	3/4	3/4
	g	0	1	1/2	1	g	0	1	3/4	1
		$\overline{F}$ for Table 3.1					$\overline{F}$ for Table 3.2			

For distributions over beliefs more restrictive bounds can be established. In general, the Fréchet-Hoeffding bounds are too wide for distributions over beliefs. Examples best illustrate this issue. First, Example 10 shows that, although the information structure from Table 3.1 attains[58] the lower Fréchet-Hoeffding bound, the induced belief distribution does not attain the Fréchet-Hoeffding bound. Second, Example 8 shows a belief distribution that attains the lower Fréchet-Hoeffding bounds. However, this belief distributions cannot be induced by an information structure, meaning that the usual Fréchet-Hoeffding bounds can be tightened to bound the distributions of beliefs induced by any information structure.

Example 10. As before, consider a binary state case with prior $\pi = 1/2$. Suppose the information structure in Table 3.8 is given. The information structure attains the Fréchet-Hoeffding lower bound and induces two posteriors: $1/4$ and $3/4$.

[58]See Footnote 56.

Table 3.8: Non-revealing symmetric information structure.

		Report for Player 2			
		$\theta = 1$		$\theta = 0$	
		b	g	b	g
Report for Player 1	b	$1/2$	$1/4$	0	$1/4$
	g	$1/4$	0	$1/4$	$1/2$

The induced distribution over beliefs is given in Table 3.9: the bivariate uniform distribution. This clearly differs the Fréchet-Hoeffding lower bound, which is shown on the right. Thus, although the information structure attains the lower bound, the induced distribution over beliefs does not attain the Fréchet-Hoeffding lower bound. As shown later, the belief distribution on the right (i.e. the Fréchet-Hoeffding lower bound) cannot be induced by any information structure. Indeed, any distribution that shows more negative dependence than the actual belief distribution (i.e. the distribution on the left) cannot be a belief distribution induced by any information structure. Therefore, the usual Fréchet-Hoeffding bounds are not tight enough. This becomes even more transparent in the next example. ◇

Table 3.9: Belief distribution induced by the information structure of Table 3.8.

		Belief of Player 2					
		PMF		CDF		$\underline{F}$	
		$1/4$	$3/4$	$1/4$	$3/4$	$1/4$	$3/4$
Belief of Player 1	$1/4$	$1/4$	$1/4$	$1/4$	$1/2$	0	$1/2$
	$3/4$	$1/4$	$1/4$	$1/2$	1	$1/2$	1

Example 8 (continuing from p. 156)**.** Consider the distributions only over beliefs derived from Table 3.5, which is stated in Table 3.10. Example 8 already established that the

distribution cannot be induced by any information structure. However, the distribution of Table 3.10 corresponds to the lower Fréchet-Hoeffding bound. ◇

Table 3.10: Pure belief distribution not induced by any information structure.

		Belief of Player 2	
		PMF	
		0	1
Belief of Player 1	0	0	1/2
	1	1/2	0

The previous examples established that the usual Fréchet-Hoeffding bounds can be tightened to provide necessary conditions for distributions over beliefs induced by information structures. In this section, I introduce and discuss such bounds that are useful for the information design question at hand. Since these bounds concern CDFs defined on beliefs, the space of beliefs needs to be ordered. Although the proposed bounds hold for *any* total order, it is convenient to take a linear extension of the first-order stochastic dominance order.[59] To do this, endow the state of nature Θ with a total order, i.e. $\Theta = \{\theta_1, \ldots, \theta_K\}$ for some finite $K < \infty$ and the order corresponds to the indexing set. Then endow $\Delta(\Theta)$ with a completion of first-order stochastic dominance giving rise to a lattice structure. Given $\mu, \mu' \in \Delta(\Theta)$, a sufficient condition for $\mu \geq \mu'$ is μ first-order stochastic dominating μ', i.e. for every $L = 1, \ldots, K$,

$$\sum_{k=1}^{L} \mu(\theta_k) \leq \sum_{k=1}^{L} \mu'(\theta_k).$$

[59]Such a completion always exists due to Szpilrajn's extension theorem. See Aliprantis and Border (2006, Theorem 1.9).

Given this order, define CDFs over beliefs analogously to the case of CDFs of real-valued random variables. That is, for a given distribution $\tau \in \Delta(\Delta(\Theta))$, define the associated CDF by $T(\mu) = \sum_{\mu' \leq \mu} \tau(\mu')$. Similarly, $\Delta(\Theta) \times \Delta(\Theta)$ is endowed with the product order derived from the order on each dimension. Then, for any joint distribution $\tau \in \Delta(\Delta(\Theta) \times \Delta(\Theta))$ the associated (joint) CDF is given by

$$T(\mu) = T(\mu_1, \mu_2) = \sum_{\mu_1' \leq \mu_1, \mu_2' \leq \mu_2} \tau(\mu_1', \mu_2').$$

With these definitions in hand, the belief-dependence bounds can be defined. Similar to the Fréchet-Hoeffding bounds, these bounds are defined for given marginal distributions.

Definition 16. *Fix two univariate distributions over beliefs τ_1, $\tau_2 \in \Delta(\Delta(\Theta))$ and a prior $\pi \in \Delta(\Theta)$. The* lower belief-dependence bound *is defined as*

$$\underline{T}(\mu_1, \mu_2) = \max_{0 \leq L \leq K} \max \left\{ \underline{T}_1(\mu_1, \mu_2; L), \underline{T}_2(\mu_1, \mu_2; L) \right\}, \tag{3.9}$$

where for each[60] $L = 0, \dots, K$,

$$\underline{T}_1(\mu_1, \mu_2; L) = \sum_{\mu_1' \leq \mu_1} \tau_1(\mu_1') \sum_{k=1}^{L} \mu_1'(\theta_k) + \sum_{\mu_2' \leq \mu_2} \tau_2(\mu_2') \sum_{k=1}^{L} \mu_2'(\theta_k) - \sum_{k=1}^{L} \pi(\theta_k),$$

and (3.10)

$$\underline{T}_2(\mu_1, \mu_2; L) = \sum_{\mu_1' \leq \mu_1} \tau_1(\mu_1') \sum_{k=L+1}^{K} \mu_1'(\theta_k) + \sum_{\mu_2' \leq \mu_2} \tau_2(\mu_2') \sum_{k=L+1}^{K} \mu_2'(\theta_k) - \sum_{k=L+1}^{K} \pi(\theta_k).$$

The upper belief-dependence bound *is defined as*

$$\overline{T}(\mu_1, \mu_2) = \min_{1 \leq L \leq K} \min \left\{ \overline{T}_1(\mu_1, \mu_2; L), \overline{T}_2(\mu_1, \mu_2; L) \right\}, \tag{3.11}$$

where for each[61] $L = 0, \dots, K$,

$$\overline{T}_1(\mu_1, \mu_2; L) = \sum_{\mu_1' \leq \mu_1} \tau_1(\mu_1') \sum_{k=1}^{L} \mu_1'(\theta_k) + \sum_{\mu_2' \leq \mu_2} \tau_2(\mu_2') \sum_{k=L+1}^{K} \mu_2'(\theta_k),$$

and (3.12)

$$\overline{T}_2(\mu_1, \mu_2; L) = \sum_{\mu_1' \leq \mu_1} \tau_1(\mu_1') \sum_{k=L+1}^{K} \mu_1'(\theta_k) + \sum_{\mu_2' \leq \mu_2} \tau_2(\mu_2') \sum_{k=1}^{L} \mu_2'(\theta_k).$$

A few observation are in order. First, as argued in the previous section, the goal is to tighten the usual Fréchet-Hoeffding bounds using the restrictions imposed by the actual information structures and Bayesian updating. Thus, the belief-dependence bounds should be tighter, which is indeed the case. Formally, for the lower bound it

[60] By convention, empty sums are defined to be zero.
[61] Again, by convention, empty sums are defined to be zero.

holds that $\underline{F}(\mu_1, \mu_2) \leq \underline{T}(\mu_1, \mu_2)$ since $\underline{F}(\mu_1, \mu_2) = \max_{L \in \{0,K\}} \underline{T}(\mu_1, \mu_2; L) \leq \underline{T}(\mu_1, \mu_2)$. For the upper bound the reversed inequality, $\overline{F}(\mu_1, \mu_2) \geq \overline{T}(\mu_1, \mu_2)$, holds because $\overline{F}(\mu_1, \mu_2) = \min\left\{\overline{T}_1(\mu_1, \mu_2; K), \overline{T}_2(\mu_1, \mu_2; K)\right\} \geq \overline{T}(\mu_1, \mu_2)$. Second, if the marginal distributions are equal, i.e. $\tau_1 = \tau_2$, then the upper belief-dependence bound is actually the same as the upper Fréchet-Hoeffding bound. Formally:

Lemma 7. *Fix an economic environment $\mathcal{E}$ and a full-support prior $\pi \in \Delta(\Theta)$. Consider two univariate distributions $\tau_1, \tau_2 \in \Delta(\Delta(\Theta))$ such that $\tau_1 = \tau_2$ and suppose that $\mathbb{E}_{\tau_1}[\mu_1] = \pi$. Then, the upper belief-dependence bound is the usual upper Fréchet-Hoeffding bound, i.e. $\overline{T} = \overline{F}$.*

Proof. For notation, see the proof of Proposition 6 below. By Symmetry $T_1 = T_2$ and similar for M_i. Thus, I will drop the indices. Without loss say $\mu_1 \leq \mu_2$, then $T(\mu_1) \leq T(\mu_2)$. Fix any L and then $\overline{T}_1(\mu_1, \mu_2; L) = M(\mu_1) + T(\mu_2) - M(\mu_2) \geq M(\mu_1) + T(\mu_1) - M(\mu_1) = T(\mu_1)$. Similarly, $\overline{T}_2(\mu_1, \mu_2; L) = T(\mu_1) - M(\mu_1) + M(\mu_2) \geq T(\mu_1) - M(\mu_1) + M(\mu_1) = T(\mu_1)$. ■

Furthermore, these bounds are indeed necessary conditions for distributions over beliefs to be induced by any information structure.

Proposition 6. *Fix an economic environment $\mathcal{E}$ and a full-support prior $\pi \in \Delta(\Theta)$. $\tau \in \Delta(\Delta(\Theta) \times \Delta(\Theta))$ is induced by an information structure only if*[62]

1. *$\sum_{\mu_1,\mu_2} \tau(\mu_1, \mu_2)\mu_1 = \sum_{\mu_1,\mu_2} \tau(\mu_1, \mu_2)\mu_2 = \pi$, and*

2. *$\underline{T} \precsim T \precsim \overline{T}$.*

[62]Here, a slight abuse of notation appears: the belief bounds are formally only defined for two marginal beliefs. In the statement there is only the joint distribution τ. The belief bounds correspond to the bounds defined by using the two marginals distributions derived from τ.

The formal proof of Proposition 6 is stated in below, but requires some cumbersome notation. Thus, I provide a sketch of the main steps of the proof first.

Step 1 [Characterization of state-dependent distributions over beliefs]: Theorem 9 gives a characterization of the distributions $\lambda \in \Delta(\Delta(\Theta)\times\Delta(\Theta)\times\Theta)$ that can arise from any information structure, where the third dimension corresponds to the actual state of nature.

Step 2 [From τ to marginals of λ]: Consider τ induced by an information structure. By Step 1, there exists λ with marginal τ satisfying the properties of Theorem 9. Equation 3.6 shows condition (1) in Proposition 6 has to hold.

Using property (2) of Theorem 9 it can be verified that the two other bivariate marginals of λ are given by $\lambda_{i,\theta}(\mu_i,\theta) = \mu_i(\theta)\sum_{\mu_{-i}}\tau(\mu_i,\mu_{-i})$. It remains to show necessesity of (2) as stated in Proposition 6 using the marginals of λ.

Step 3 [Higher-order Fréchet-Hoeffding bounds]: Joe (1997, Theorem 3.11) extends the usual Fréchet-Hoeffding bounds to trivariate distribution with given bivariate marginals. For the distribution here, these bounds say that the desired λ exists only if

$$\underline{\Gamma}\left(\tau,\lambda_{1,\theta},\lambda_{2,\theta}\right) \leq \overline{\Gamma}\left(\tau,\lambda_{1,\theta},\lambda_{2,\theta}\right), \tag{3.13}$$

where $\underline{\Gamma}$ and $\overline{\Gamma}$ are functionals mapping to CDFs of trivariate distributions. By contradiction, suppose that τ does not satisfy the bounds of Proposition 6. Since both $\lambda_{i,\theta}$'s depend only on τ, simple alegbra shows that the inequality of Joe (1997) described in Equation 3.13 is violated, which implies that λ does not exists. Contradiction. QED.

Proof. For the proof consider the following definitions. Given $\tau_1, \tau_2 \in \Delta(\Delta(\Theta))$ and a prior $\pi \in \Delta(\theta)$ define

$$\begin{aligned}
\Pi(L) &= \sum_{k=1}^{L} \pi(\theta_k) \\
T_1(\mu_1) &= \sum_{\mu_1' \leq \mu_1} \tau_1(\mu_1') \quad \text{and} \quad T_2(\mu_2) = \sum_{\mu_2' \leq \mu_2} \tau_2(\mu_2') \\
T(\mu_1, \mu_2) &= \sum_{\mu_1' \leq \mu_1} \sum_{\mu_2' \leq \mu_2} \tau(\mu_1', \mu_2') \\
M_1(\mu_1, L) &= \sum_{\mu_1' \leq \mu_1} \tau_1(\mu_1') \sum_{k=1}^{L} \mu_1'(\theta_k) \\
M_2(\mu_2, L) &= \sum_{\mu_2' \leq \mu_2} \tau_2(\mu_2') \sum_{k=1}^{L} \mu_2'(\theta_k).
\end{aligned}$$

With these definitions, the elementary functions of the belief-dependence bounds can be restated as

$$\begin{aligned}
\underline{T}_1(\mu_1, \mu_2; L) &= M_1(\mu_1, L) + M_2(\mu_2, L) - \Pi(L) \\
\underline{T}_2(\mu_1, \mu_2; L) &= T_1(\mu_1) - M_1(\mu_1, L) + T_2(\mu_2) - M_2(\mu_2, L) - [1 - \Pi(L)] \\
\overline{T}_1(\mu_1, \mu_2; L) &= M_1(\mu_1, L) + T_2(\mu_2) - M_2(\mu_2, L) \\
\overline{T}_2(\mu_1, \mu_2; L) &= T_1(\mu_2) - M_1(\mu_2, L) + M_2(\mu_1, L),
\end{aligned}$$

for every $L = 0, \ldots, K$.[63]

Consider τ that is induced by an information structure. Since it is induced by an information structure, there exists $\lambda \in \Delta(\Delta(\Theta) \times \Delta(\Theta) \times \Theta)$ with marginal distribution on $\Delta(\Theta) \times \Delta(\Theta)$ given by τ and properties (1) and (2) as stated in Theorem 9. By Equation 3.6 the marginal conditions (1) of Proposition 6 are satisfied. Now, define

$$\lambda_1(\mu_1, \theta) = \mu_1(\theta) \sum_{\mu_2} \tau(\mu_1, \mu_2),$$
$$\lambda_2(\mu_2, \theta) = \mu_2(\theta) \sum_{\mu_1} \tau(\mu_1, \mu_2).$$

Since τ is a (bivariate) marginal of λ and due to (2) of Theorem 9, λ_1 and λ_2 are the two other bivariate marginals of λ.

Now, by Joe (1997, Theorem 3.11), λ with the given bivariate marginals exists only if for every $L = 0, \ldots, K$ and every $\mu_1, \mu_2 \in \Delta(\Theta)$,

$$\begin{aligned}
&\max\{0, T(\mu_1,\mu_2) - [T_1(\mu_1) - M_1(\mu_1, L)], T(\mu_1,\mu_2) - [T_2(\mu_2) - M_2(\mu_2, L)], \\
&\quad M_1(\mu_1, L) + M_2(\mu_2, L) - \Pi(L)\} \\
&\quad \leq \\
&\min\{T(\mu_1,\mu_2), M_1(\mu_1, L), M_2(\mu_2, L), \\
&\quad T(\mu_1,\mu_2) + [1 - \Pi(L)] - [T_1(\mu_1) - M_1(\mu_1, L)] - [T_2(\mu_2) - M_2(\mu_2, L)]\}.
\end{aligned} \tag{3.14}$$

[63] Recall that a summation over an empty set is, by definition, zero.

Now, it remains to prove that existence of λ and Equation 3.14 imply the bounds $\overline{T} \gtrsim T \gtrsim \underline{T}$. By way of contradiction, suppose the bounds are violated, then (at least) one of the following cases has to apply for some $L = 0, \ldots, K$ and some $\mu_1, \mu_2 \in \Delta(\Theta)$

- If $T(\mu_1, \mu_2) > \overline{T}_1(\mu_1, \mu_2; L) = M_1(\mu_1, L) + T_2(\mu_2) - M_2(\mu_2, L)$, then

$$T(\mu_1, \mu_2) - [T_2(\mu_1) - M_2(\mu_1, L)] > M_1(\mu_2, L).$$

- If $T(\mu_1, \mu_2) > \overline{T}_2(\mu_1, \mu_2; L) = T_1(\mu_2) - M_1(\mu_2, L) + M_2(\mu_1, L)$, then

$$T(\mu_1, \mu_2) - [T_1(\mu_1) - M_1(\mu_1, L)] > M_2(\mu_2, L).$$

In either case, Equation 3.14 is violated. Similarly, if $T(\mu_1, \mu_2) < \underline{T}_1(\mu_1, \mu_2; L) = M_1(\mu_1, L) + M_2(\mu_2, L) - \Pi(L)$ or $T(\mu_1, \mu_2) < \underline{T}_2(\mu_1, \mu_2; L) = T_1(\mu_1) - M_1(\mu_1, L) + T_2(\mu_2) - M_2(\mu_2, L) - [1 - \Pi(L)]$ then Equation 3.14 is violated.

Thus, if the bounds are not satisfied at any point, Equation 3.14 is violated. This means that there is no trivariate distribution with the marginals given by τ, λ_1, and λ_2. However, this is in contradiction with the existence of λ. ■

Thus, these bounds gives rise to a problem for finding an upper bound of the pure-persuasion design problem:

Corollary 4. *Fix a design environment $\mathcal{D}$. Then,*

$$\sup_{I \in \mathcal{I}} V(I) \leq \overline{V}(\pi) := \sup_{\tau \in \Delta(\Delta(\Theta)^2)} \sum_{\mu_1, \mu_2} \tau(\mu_1, \mu_2) v(\mu_1, \mu_2)$$

$$s.t. \sum_{\mu_1, \mu_2} \tau(\mu_1, \mu_2) \mu_1 = \pi,$$

$$\sum_{\mu_1, \mu_2} \tau(\mu_1, \mu_2) \mu_2 = \pi,$$

$$and \;\; \underline{T} \precsim T \precsim \overline{T}.$$

This corollary shows that the designer solves the (relaxed) problem as if she chooses marginal belief distributions for each receiver subject to the familiar Bayes plausibility conditions. Moreover, the beliefs across the two receivers cannot be too dependent so that the joint distribution satisfies the belief-dependence bounds. The constraints on the distributions of beliefs are tractable, especially if the designer utility v (as a function on belief space) has special properties.

For two-dimensional real-vectors it is well known[64] that the stochastic order $\precsim$ (recall Definition 15) has a dual characterization in terms of utility functions. In particular,

$$F \precsim G \iff \mathbb{E}_F[w(x, y)] \leq \mathbb{E}_G[w(x, y)],$$

for all Bernoulli utility functions $w : \mathbb{R}^2 \to \mathbb{R}$ that are supermodular. Meyer and Strulovici (2015) extend this result to distribution over a finite, n-dimensional lattice. Since the order

[64] In probability theory, this is known at least since Cambanis, Simons, and Stout (1976) and Tchen (1980).

on beliefs was assumed to be a total order, Meyer and Strulovici's results apply to the setting of this chapter. Thus, if ν in Corollary 4 is supermodular, then the pure persuasion design problem can be simplified by first solving

$$\begin{aligned}
\sup_{\tau_1,\tau_2 \in \Delta(\Delta(\Theta))} & \sum_{\mu_1,\mu_2} \tau(\mu_1,\mu_2)\nu(\mu_1,\mu_2) \\
\text{s.t. } & \sum_{\mu_1} \tau_1(\mu_1)\mu_1 = \pi, \\
& \sum_{\mu_2} \tau_2(\mu_2)\mu_2 = \pi, \\
& \text{and } T = \overline{T},
\end{aligned}$$

and then verifying whether the resulting τ is induced by an information structures. Symmetrically, if ν is submodular the last constraint would be replaced by $T = \underline{T}$. In either case, the problem is simplified because the choice set contains only marginal distributions.

Kamenica and Gentzkow (2011) show that the value of the information-design problem with one receiver is equal to the concavification of the underlying utility function of the designer. This turns out to be a convenient way of solving the design problems for specific environments. Sometimes, the concavification approach is useful even for the case with two receivers, as considered in this chapter. Here, applying the concavification[65] to the

[65] See, for example, Rockafellar (1970, Corollary 17.1.5)

designer utility ν gives

$$\begin{aligned} \operatorname{cav} \nu(\pi_1, \pi_2) = \sup_{\tau \in \Delta(\Delta(\Theta)^2)} & \sum_{\mu_1, \mu_2} \tau(\mu_1, \mu_2) \nu(\mu_1, \mu_2) \\ \text{s.t.} & \sum_{\mu_1, \mu_2} \tau(\mu_1, \mu_2) \mu_1 = \pi_1, \\ & \sum_{\mu_1, \mu_2} \tau(\mu_1, \mu_2) \mu_2 = \pi_2. \end{aligned}$$

Thus, the concavification is just an (even more) relaxed version of the actual designer's problem. It suggests solving the concavification approach and then checking whether the resulting distribution actually satisfies the belief bounds. This might be useful for applications: as demonstrated later in Subsection 3.4.2, this approach simplifies the search for the optimal information structure in the CRO example, given that the CRO has supermodular preferences. This observation is formally recorded next.

Corollary 5. *Fix a design environment $\mathcal{D}$. The concavification of ν is an upper bound for the relaxed problem of Corollary 4 (*a fortiori *for the actual design problem), i.e.*

$$\operatorname{cav} \nu(\pi, \pi) \geq \overline{V}(\pi) \geq \sup_{I \in \mathcal{I}} V(I).$$

As explored above, Corollary 4 allows further simplifications of the maximization problem if the designer's utility function defined on the belief space takes particular forms. However, this utility function ν is an object derived from the primitive objects stated in a design environment $\mathcal{D}$. Next, I discuss a broad class of environments which provides easy verifiable sufficient conditions on primitives to ensure that the derived object ν satisfies sub- or supermodularity whenever the primitive function v satisfies these properties. In addition, a subclass of these environments allows me to provide an upper bound on the cardinality of the signal space (see Example 7).

Definition 17. *An economic environment* $\mathcal{E} = \langle \Theta, (A_i, u_i)_{i \in N} \rangle$ *is* monotone *if*

1. *the states of nature* Θ *are endowed with an total order,*

2. *for each player* $i \in N$*, the set of actions* A_i *is endowed with an total order, and*

3. *for each player* $i \in N$*, the utility function has increasing differences in* (a_i, θ)*, i.e. for all* $(a_i, \theta), (a_i', \theta') \in A_i \times \Theta$ *and all* $a_{-i} \in A_{-i}$*,*

$$a_i' \geq a_i \text{ and } \theta' \geq \theta \implies u_i(a_i', a_{-i}, \theta') + u_i(a_i, a_{-i}, \theta) \geq u_i(a_i', a_{-i}, \theta) + u_i(a_i, a_{-i}, \theta').$$

A design environment $\mathcal{D} = \langle \mathcal{E}, \pi, v \rangle$ *is* monotone *if*

1. *the economic environment* $\mathcal{E}$ *is monotone, and*

2. *the designer's utility function* $v : A \to \mathbb{R}$ *is increasing*[66] *with respect to the product order induced by the orders on the set of actions* A_i*, i.e. for all* $(a_1, a_2) \in A$*,*

$$a_i' \geq a_i, \text{ for all } i = 1, 2 \implies v(a_1', a_2') \geq v(a_1, a_2).$$

[66]Only monotonicity of v is needed for all of the following analysis. The definition uses increasingness to simplify the notation.

Supermodular games usually have an underlying economic environment that is monotone. However, the class of monotone environments is more general since it does not specify increasing differences in (a_i, a_{-i}), which is assumed to transform an economic environment to supermodular game. Thus, the class of environments here is quite general, but specific enough to translate the preference for complementarities from action space to belief space as formally stated in the next proposition. This proposition, therefore, provides a simple way to check the primitives to ensure that the derived Bernoulli utility in in belief space is either sub- or supermodular.

Proposition 7. *Consider a monotone design environment $\mathcal{D}$. Suppose the designer's utility $v : A \to \mathbb{R}$ is supermodular then the derived utility $\nu : \Delta(\Theta) \times \Delta(\Theta) \to \mathbb{R}$ on belief space (endowed with the first-order stochastic dominance order) is supermodular, where*

$$\nu(\mu_1, \mu_2) := \min_{\left(a_i \in R_i(\mu_i)\right)_{i \in N}} v(a_1, a_2).$$

Similarly, if v is submodular, then ν is submodular as well.

Proof. I will only prove the case of supermodularity. Consider $\mu_i \in \Delta(\Theta)$ and $\eta : \Theta \to \Delta(A_{-i})$ such that $\operatorname{supp} \nu(\cdot|\theta) \subseteq BFR_{-i}$ for all $\theta \in \Theta$. Since supermodularity is preserved under summation (i.e. expectation) , the best-reply is increasing (in the strong set-order) in first-order beliefs μ_i (holding η fixed), see van Zandt and Vives (2007). Thus the robust prediction correspondence is increasing (in the strong-set order). Now, let $b_i(\mu_i) = \min \{a_i \in R_i(\mu_i)\}$, which is increasing in μ_i. Because v is increasing,

$\nu(\mu_1, \mu_2) = v(b_1(\mu_1), b_2(\mu_2))$. If μ_1 first-order stochastic dominates μ_1', then $b_1(\mu_1) \geq b_1(\mu_1')$. Thus,

$$\nu(\mu_1, \mu_2) - \nu(\mu_1', \mu_2) = v(b_1(\mu_1), a_2(\mu_2)) - v(b_1(\mu_1'), a_2(\mu_2)),$$

is increasing in μ_2 because $b_2(\cdot)$ is and v is supermodular. ■

In the general problem, Example 7 illustrates that using recommendations similar to the usual revelation principle does not work. For monotone design environments with a restriction on information structures, action recommendations provide a rich enough signal space. Action recommendations turn out to be useful even when working in belief space as will be illustrated in Subsection 3.4.2.[67] For this, say that an information structure I is *direct* if for every $i \in N$, $S_i \subseteq A_i$ and for every signal $a = (a_1, a_2)$, it holds that $\min_{\left(a_i' \in R_i(s_i | I_i, \pi)\right)_{i \in N}} v(a_1', a_2') = v(a)$. Then, the following proposition is akin to a standard revelation principle.

Proposition 8 (Revelation Principle). *Suppose the design environment $\mathcal{D}$ is monotone. Restrict the choice of information structures to information structures that give rise to posteriors that are totally ordered by first-order stochastic dominance for each player.*[68] *Then, there exists an information structure I with value $V(I)$ if and only if there exists a direct information structure $\hat{I}$ such that $v(I) = v(\hat{I})$.*

[67]Whether this revelation principle argument is useful for working directly in signal space is an open question.

[68]For example, if the state space is binary, then this assumption is without loss of generality.

Proof. One direction is obvious. For the other fix an information structure I. Then, define[69] $S_i^1 = \left\{ s_i \in S_i : a_i^1 \in R_i(s_i) \right\}$ and for $1 < k \leq J_i$

$$S_i^k = \left\{ s_i \in S_i : a_i^k \in R_i(s_i) \text{ and } a_i^l \notin R_i(s_i) \text{ for all } l < k \right\}.$$

Now, let $\hat{S}_i = \left\{ a_i^j \in A_i : S_i^j \neq \emptyset \right\} \subseteq A_i$ and set the signal distribution to

$$\hat{\Psi}(a_1^{j_1}, a_2^{j_2} | \theta) = \sum_i \sum_{s_i \in S_i^{j_i}} \Psi(s_1, s_2 | \theta).$$

Now, for a given $a_i^j \in \hat{S}_i$, the induced first-order belief (call it μ) will be a convex combination of beliefs (i.e μ_{s_i} for $s_i \in S_i^j$). Since these beliefs are totally ordered, one of these beliefs is the lowest according to first-order stochastic dominance; call it $\underline{\mu}$. Thus, the convex combination (i.e. μ) is also greater than $\underline{\mu}$. As shown in Proposition 3.4.1, the robust-prediction correspondence is increasing. Thus, $R_i(\underline{\mu}) \leq R_i(\mu)$ in the strong set order.

By construction, I have $a_i^m \notin R_i(\underline{\mu})$ for $m < j$ implying that $a_i^m \notin R_i(\mu)$. Furthermore, a_i^j is conceivable for each μ_{s_i} for $s_i \in S_i^j$. That is, for each such $s_i \in S_i^j$ there exists $\eta_{s_i}(\cdot|\cdot) : \Theta \to \Delta(A_{-i})$ such that $a_i^j \in BR_i(\mu_{s_i} \circ \eta_{s_i})$. Consider[70] $\tilde{\mu} = \sum_{s_i \in \hat{S}_i^{\mathcal{A}_i}} \lambda_{s_i} \mu_{s_i} \circ \eta_{s_i}$, which has marginal μ by construction. And since a_i^j is a best-reply to each belief separately, it's also a best-reply to the convex combination. Proving $a_i^j \in R_i(\mu)$.

[69]The superscripts refer to the indexing set of the actions, i.e. $A_i = \{a_i^1, \ldots, a_i^{J_i}\}$.
[70]Let λ_{s_i} denote the coefficients of the convex combination.

So I established

$$a_i^j \in R_i(a_i^j) \text{ and } a_i^m \notin R_i(a_i^j), \quad \text{for all } m < j.$$

Thus, by Definition 17 for any $(a_1, a_2) \in \hat{S}_1 \times \hat{S}_2$

$$\min_{a_i' \in R_i(a_i)} v(a_1', a_2') = v(a_1, a_2).$$

Proving that the information structure is direct. That the values are the same follows trivially from the construction. ■

This result is interpreted slighlty different from the usual interpretation of the revelation principle as in Myerson (1982) or Kamenica and Gentzkow (2011). Here, the designer sends action recommendations to the receivers like in the usual version, but the receivers do not have to be obedient and follow the recommendation. Instead, whatever action the receiver chooses, for the designer the action will be at least as good as if the receiver had followed the recommendation.

3.4.2. The Problem of a CRO solved

Now, the problem of the CRO introduced in Subsection 3.1.2 can be solved. Recall that the economic environment $\mathcal{E}$ can be summarized by the two game tables in Subsection 3.1.2. This economic environment is actually a monotone one. Furthermore, the prior of both pharmaceutical companies was specified as $\pi = 1/3$, thus it remains to specify the preferences for the designer (i.e. the CRO) to get a design environment. For now, assume that preferences are such that the CRO prefers further research over dropping the project for both companies, i.e.

$$v(R,\cdot) > v(D,\cdot) \quad \text{and} \quad v(\cdot,R) > v(\cdot,D),$$

which makes the design environment monotone as well. Using Figure 3.1, it is easy to obtain the CRO utility function defined on belief space, as shown in Figure 3.4.

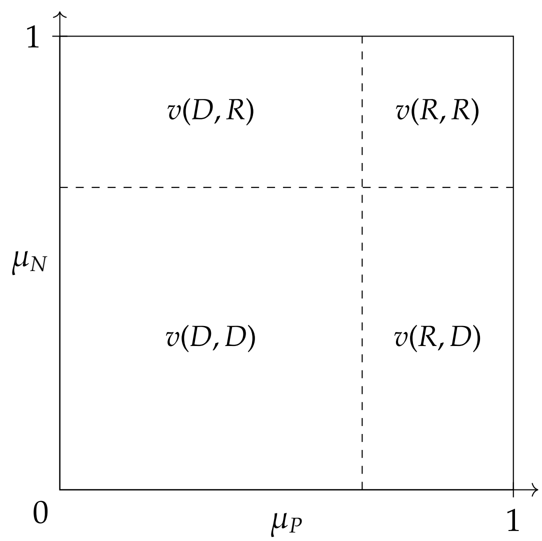

Figure 3.4: CRO utility function ν defined on belief space.

Given this derived utility function ν, the optimal information and the corresponding value can be obtained by applying Theorem 9. The problem is analyzed separately for two possible cases of sub- and supermodular preferences of the CRO. For the remainder, I also assume that the preferences are symmetric.[71]

[71] This is not without loss of generality!

Supermodular case: Suppose that the utility of the CRO is supermodular, i.e. $v(R,R) + v(D,D) \geq v(R,D) + v(D,R)$, then by Proposition 7 the induced belief utility function ν will be supermodular as well. In this case, the design problem can be easily solved by considering the relaxed version obtained by removing the belief-dependence bounds from the problem as stated in Theorem 9. Thus, the problem becomes equivalent to the concavification approach of Kamenica and Gentzkow (2011). Figure 3.5 plots the utility function ν in the left panel. The right panel superimposes the concavification $\operatorname{cav}\nu$. The optimal value corresponds to $\operatorname{cav}\nu(\pi,\pi)$ as indicated with an asterisk in the figure. Due to the supermodularity the CRO wants to make receivers' choices as positive dependent as possible, and the resulting belief distribution[72] (shown in Table 3.11) reflects this. It remains to verify that the belief-dependence bounds are satisfied by the solution resulting from the concavification approach. For this, recall that for symmetric marginal belief distributions the upper belief-dependence bound (which is attained due to supermodularity) is just the upper Fréchet-Hoeffding bound. Thus, the distribution in Table 3.11 is indeed a valid belief distribution. An information structure inducing this belief distribution is also shown in Table 3.11.

Submodular case: In the remaining case, the CRO is assumed to have submodular preferences. That is, $v(R,R)+v(D,D) \leq v(R,D)+v(D,R)$, which implies that ν is submodular similar to before. Here, the concavification approach is not useful since it would

[72]Since tie-breaking does not favor the designer, ν is not upper semicontinous and an optimal information structure does not exist. To simplify this illustration, the reported information structure ignores this issue. An ε-optimal information structure would ensure that the induced belief is strictly greater than $2/3$.

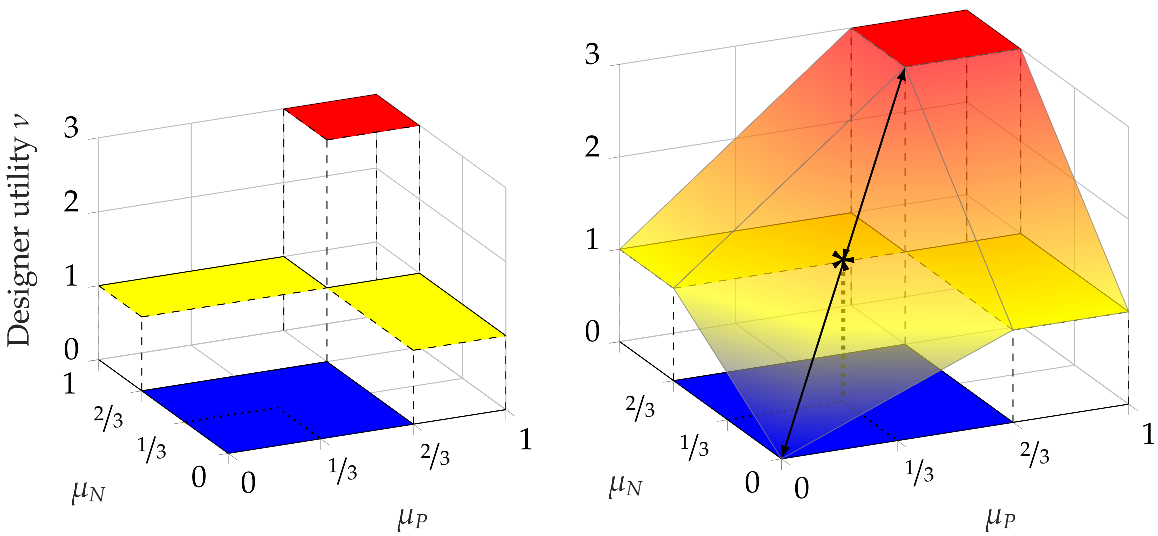

Figure 3.5: CRO utility and concavification superimposed (right panel).

Table 3.11: Optimal information for the CRO with supermodular preferences.

		Novarty							
		Signals						Belief	
		$\theta = 1$		$\theta = 0$					
		b	g	b	g			0	2/3
Signals for Pfizr	b	0	0	3/4	0	Belief of Pfizr	0	1/2	0
	g	0	1	0	1/4		2/3	0	1/2

yield a belief distribution (see Table 3.12) which cannot be induced by any information structure. This can be verified by checking that this distribution violates the lower belief-dependence bound. Thus, a different approach is needed for this case. By Proposition 8, it is sufficient to consider marginal belief distributions with binary

Table 3.12: Result from concavification approach for submodular preferences.

		Belief of Novarty	
		0	2/3
Belief of Pfizr	0	0	1/2
	2/3	1/2	0

support only: one supported belief leads to actions D in the worst-case and the other leads to action R in the worst-case. Therefore, for each receiver I need to consider beliefs $(\mu_i^D, \mu_i^R) \in [0, 2/3) \times [2/3, 1]$ only.[73] Moreover, it is easy to see that distributions leading to both actions with positive probability are better than just sticking to the prior (on each dimension). Thus, $(\mu_i^D, \mu_i^R) \in [0, 1/2) \times [2/3, 1]$ by Bayes plausibility. Using Theorem 9 the solution is readily available computationally. However, in this case the problem can be solved directly using Corollary 4. As binary signals suffice and the problem has only states of nature, the bounds of Proposition 6 coincide with the characterization of Arieli et al. (2020).[74] First, the lower belief-dependence bound[75] has to be binding due to submodularity. In the current situation this means that only $\underline{T}_1$ has to be considered. The reason is that with two states the beliefs are naturally ordered by first-order stochastic dominance and therefore the following lemma applies.

Lemma 8. *Fix two univariate belief-distributions $\tau_1, \tau_2 \in \Delta(\Theta)$ and a full-support prior $\pi \in \Delta(\Theta)$. Suppose that (i) $\mathbb{E}_{\tau_i}[\mu_i] = \pi$, and (ii) $\operatorname{supp}_i \tau_i$ is totally ordered by first-order stochastic dominance, then for every $L = 0, \ldots, K$*

$$T_i(\mu_i) - M_i(\mu_i, L) \leq T_i(\mu_i)\left[1 - \Pi(L)\right]. \tag{3.15}$$

[73] As before, I change the tie-breaking assumption here, which simplifies the notation, but does not change the essence of the argument.

[74] Arieli et al. (2020) discuss details in their Appendix B.

[75] For the binary state case first-order stochastic dominance is a total order. By Lemma 8, only $\underline{T}_1$ has to be considered for the lower bound.

Furthermore, $\underline{T}_2(\mu_1, \mu_2; L) \leq \max_L \underline{T}_1(\mu_1, \mu_2; L)$.

Proof. I use the same notation as in the proof of Proposition 6. Using the total order and (*i*), for every L and every $\mu_i \in \operatorname{supp} \tau_i$

$$\mathbb{E}_{\tau_i}[\sum_{k \leq L} \mu_i'(\theta_k) | \mu_i' \leq \mu_i] \mathbb{P}(\mu_i' \leq \mu_i) + \mathbb{E}_{\tau_i}[\sum_{k \leq L} \mu_i'(\theta_k) | \mu_i' > \mu_i] \mathbb{P}(\mu_i' > \mu_i) = \sum_{k \leq L} \mathbb{E}_{\tau_i}[\mu_i'(\theta_k)] = \Pi(L),$$

and by first-order stochastic dominance

$$\mathbb{E}_{\tau_i}[\sum_{k \leq L} \mu_i'(\theta_k) | \mu_i' \leq \mu_i] \geq \sum_{k \leq L} \mu_i(\theta_k) \geq \mathbb{E}_{\tau_i}[\sum_{k \leq L} \mu_i'(\theta_k) | \mu_i' > \mu_i].$$

Thus, $\Pi(L) \leq \mathbb{E}_{\tau_i}[\sum_{k \leq L} \mu_i'(\theta_k) | \mu_i' \leq \mu_i] = {}^{M_i(\mu_i, L)}/_{T_i(\mu_i)}$, which implies the first inequality in Equation 3.15.

For the second part, the inequality Equation 3.15 gives

$$\begin{aligned}
\underline{T}_2(\mu_1, \mu_2; L) &= T_1(\mu_1) - M_1(\mu_1, L) + T_2(\mu_1) - M_2(\mu_1, L) - [1 - \Pi(L)] \\
&\leq T_1(\mu_1)\,[1 - \Pi(L)] + T_2(\mu_2)\,[1 - \Pi(L)] - [1 - \Pi(L)] \\
&\leq T_1(\mu_1) + T_2(\mu_2) - 1 \leq \max_L \underline{T}_1(\mu_1, \mu_2; L).
\end{aligned}$$

∎

Furthermore, the lower belief-dependence bound has to be strictly tighter at some point than the usual Fréchet-Hoeffding lower bound, otherwise Table 3.12 would be the solution. Given the binary signals per receiver and the possible values for these, the only point where the bound is binding is at (μ_1^D, μ_2^D). For the other cases the Fréchet-Hoeffding bound is the same as the belief-dependence bound. Thus, letting τ_i denote the marginal distributions,

$$\tau(\mu_1^D, \mu_2^D) = \tau_i(\mu_1^D)\left(1 - \mu_1^D\right) + \tau_2(\mu_2^D)\left(1 - \mu_2^D\right) - (1 - \pi)\,,$$

has to hold for any possible joint distribution. This allows me to simplify the program as stated in Theorem 9 by making the problem separable between the two agents.[76] The reformulated program becomes

$$\begin{aligned}\sup_{\tau_1, \tau_2 \in \Delta(\Delta(\Theta))} \sum_{\mu_1} \tau_1(\mu_1) f(\mu_1) &+ \sum_{\mu_2} \tau_2(\mu_2) f(\mu_2) \\ \text{s.t.} \sum_{\mu_1} \tau_1(\mu_1)\mu_1 &= \pi, \\ \sum_{\mu_2} \tau_2(\mu_2)\mu_2 &= \pi,\end{aligned}$$

where $f(\mu) := \mathbf{1}\left[\mu < 2/3\right](2v + \mu - 1) + \mathbf{1}\left[\mu \geq 2/3\right](1 - \mu)$ using a normalization on the payoffs for the CRO.[77] The solution to this program determines the optimal marginal distributions, which are then combined to a joint distribution via the lower belief-dependence bound. Due to the established separability, the reformulation can

[76]Derivations are shown in detail below.

[77]In particular, $v(D, D) = -1$, $v(R, R) = 0$, and $v(R, D) = v(D, R) =: v \in [-1/2, 0]$. This is without loss of generality.

be solved with the concavification technique from Kamenica and Gentzkow (2011) yielding $\mu_i^{D,*} = 0$ and $\mu_i^{R,*} = 2/3$. By Bayes plausibility this gives the same marginal distribution as in Table 3.12, but these marginals must be put together with the lower belief-dependence bounds. This yields the optimal information structure as foreshadowed in the introduction and stated in Table 3.2.

Detailed calculations: To simplify notation let $\tau_{DD} := \tau(\mu_1^D, \mu_2^D)$ and similar for the other three cases and let $\tau_i := \tau_i(\mu_i^D)$. With this notation,

$$\tau_{DD} = \tau_1(1 - \mu_1^D) + \tau_2(1 - \mu_2^D) - (1 - \pi).$$

Since marginal distribution average out to the prior: $(1-\tau_i)(1-\mu_i^R)+\tau_i(1-\mu_i^D) = 1-\pi$. Hence,

$$\begin{aligned}\tau_{DD} &= \tau_1(1 - \mu_1^D) + \tau_2(1 - \mu_2^D) - (1 - \pi) - \sum_i \frac{(1 - \tau_i)(1 - \mu_i^R) + \tau_i(1 - \mu_i^D)}{2} \\ &= \frac{1}{2}\left[\tau_1(1 - \mu_1^D) - (1 - \tau_1)(1 - \mu_1^R) + \tau_2(1 - \mu_2^D) - (1 - \tau_2)(1 - \mu_2^R)\right].\end{aligned}$$

Given the normalization on the utility of the designer, the objective becomes $-\tau_{DD} + v(\tau_{DR} + \tau_{RD})$. Furthermore, the following equalities hold:

$$\begin{aligned}\tau_{DR} = \tau_1 - \tau_{DD} &= \frac{1}{2}\left[\left(\tau_1\mu_1^D + 1 - \mu_1^R + \tau_1\mu_1^R\right) - \left(\tau_2(1 - \mu_2^D) - (1 - \tau_2)(1 - \mu_2^R)\right)\right] \\ \tau_{RD} = \tau_2 - \tau_{DD} &= \frac{1}{2}\left[\left(\tau_2\mu_2^D + 1 - \mu_2^R + \tau_2\mu_2^R\right) - \left(\tau_1(1 - \mu_1^D) - (1 - \tau_1)(1 - \mu_1^R)\right)\right].\end{aligned}$$

Plugging into the objective (ignoring the $1/2$ scaling):

$$
\begin{aligned}
& v\left[\left(\tau_1\mu_1^D + 1 - \mu_1^R + \tau_1\mu_1^R\right) - \left(\tau_1(1-\mu_1^D) - (1-\tau_1)(1-\mu_1^R)\right)\right] - \left(\tau_1(1-\mu_1^D) - (1-\tau_1)(1-\mu_1^R)\right) \\
+ & v\left[\left(\tau_2\mu_2^D + 1 - \mu_2^R + \tau_2\mu_2^R\right) - \left(\tau_2(1-\mu_2^D) - (1-\tau_2)(1-\mu_2^R)\right)\right] - \left(\tau_2(1-\mu_2^D) - (1-\tau_2)(1-\mu_2^R)\right) \\
= & \left[2v\tau_1 - \tau_1(1-\mu_1^D) + (1-\tau_1)(1-\mu_1^R)\right] + \left[2v\tau_2 - \tau_2(1-\mu_2^D) + (1-\tau_2)(1-\mu_2^R)\right] \\
= & \tau_1\left(2v - (1-\mu_1^D)\right) + (1-\tau_1)(1-\mu_1^R) + \tau_2\left(2v - (1-\mu_2^D)\right) + (1-\tau_2)(1-\mu_2^R),
\end{aligned}
$$

so that the objective is separable. The analysis above implies $\mu_i^D < 2/3$ and $\mu_i^R \geq 2/3$. Thus, I can rewrite the objective as claimed before[78] with

$$
f(\mu) := \mathbf{1}\left[\mu < 2/3\right](2v + \mu - 1) + \mathbf{1}\left[\mu \geq 2/3\right](1-\mu)\,.
$$

Figure 3.6 plots this function and its concavification. Since $v \in [-1/2, 0]$ shifts f only vertically, it will not change the maximizer resulting from the concavification.

[78] To be precise, the values of f for $\mu \in [1/2, 2/3)$ can be set arbitrary as long as they are strictly below the resulting concavification. This can be done because from the previous analysis it is known that μ in this range cannot be optimal.

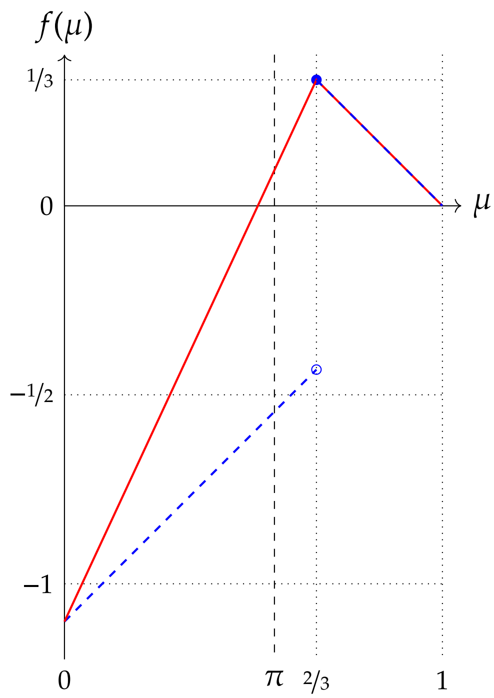

Figure 3.6: $f(\mu)$ in dashed blue (with $v = -0.05$) and concavification thereof in red.

3.5. Discussion

In this section, I discuss some extensions of the model and highlight some conceptual aspects.

3.5.1. Extension to Multiple Receivers

In this chapter, I have focused only on two players only. This simplifies the notation significantly. The solution concept introduced in Section 3.2 readily extends to any finite number of players if the definitions of belief-free rationalizability (Equation BFR^n) and rational-extended beliefs (Definition 5) are adapted to allow for general correlated beliefs about the opponents' actions.[79] Moreover, the general design problem discussed in Section 3.3 can be adjusted accordingly to multiple receivers. However, the bounds in Section 3.4 do not extend to multiple players without adaption. Of course, the functional form of the belief bounds is specific to two receivers, but a similar approach as in the proof of Proposition 6 can be adapted. I sketch this in accordance with the proof steps of Proposition 6.

Step 1 [Characterization of state-dependent distributions over beliefs]: This step generalizes as mentioned above.

[79] REF prev chapter, TODO

Step 2 [From τ to marginals of λ]: This generalizes as well, but now one has $|N| + 1$ marginals: the marginal τ on $\Delta(\Delta(\Theta)^N)$, and $|N|$ bivariate marginals $\lambda_{i,\theta}$ on $\Delta(\Delta(\Theta) \times \Theta)$.

Step 3 [Higher-order Fréchet-Hoeffding bounds]: This step is crucial for a generalization. Here an extension of Joe (1997, Theorem 3.11) is needed to obtain bounds for $|N| + 1$ dimensional distribution for marginals like the ones in Step 2, i.e. a version of

$$\underline{\Gamma}\left(\tau, (\lambda_{i,\theta})_{i\in N}\right) \leq \overline{\Gamma}\left(\tau, (\lambda_{i,\theta})_{i\in N}\right),$$

where $\underline{\Gamma}$ and $\overline{\Gamma}$ would be functionals mapping to CDFs of $|N| + 1$ dimensional distributions.

Deriving these bounds and studying their properties is left for future research.

3.5.2. The Economic Enviornment is Common Knowledge, so is Rationality

Throughout this chapter, I operated from the assumption that the economic environment is common knowledge among the players. In the examples this did not matter too much, but it this knowledge is crucial for the solution concept, which also requires common knowledge of rationality. A slight adaption of Battigalli et al. (2011, Section 3.1–3.2, see also Section 4.2) shows that the individual robust prediction corresponds to the behavioral implications of common belief of the economic environment *and* rationality, as well as *knowledge* of the marginal information structure. For certain economic environments, this has important consequences for the design of information structures. To see this, consider the following economic environment:

$\theta = 1$

Pfizr \ Novarty	R	D
R	1, -1	2, 0
D	0, -1	0, 0

$\theta = 2$

Pfizr \ Novarty	R	D
R	-2, -1	-1, 0
D	0, -1	0, 0

The payoffs for Pfizr are the same as in the CRO example, however Novarty now has an (ex-post) dominated action: R is always worse than D. For the same prior as before ($\pi = 5/9$) the robust prediction without any information would be $\{R\}$ for Pfizr (and, of course, D for Novarty). Thus, without providing any information the designer gets the best possible outcome. Suppose now, that the designer does not assume common knowledge of rationality among the receivers but still assumes rationality and knowledge of the marginal information structure for each receiver. The corresponding (even more) robust predication can be obtained by dropping part (2) in the definition of rCPS (Definition 5). In this example, this version of robust prediction (interpreted as a function of first-order beliefs) for Pfizr yields the same as in the running example in the main analysis of this chapter. Therefore, if the designer is concerned about robustness under these less restrictive assumptions, she will engage in Bayesian Persuasion á la Kamenica and Gentzkow (2011) with Pfizr. This means that the designer will optimally reveal the state of the drug being ineffective sometimes, which implies that Pfizr will drop the project occasionally. This is in contrast to the behavior under the assumption of common knowledge of rationality, where Pfizr will conduct further research with certainty. What is the *right* optimal information structure for the designer? This depends on the assumptions the designer wants to make. In this chapter, the designer imposes common knowledge of rationality.

3.5.3. Robust Information Design

The key aspect of robust mechanism design as initiated by Bergemann and Morris (2005), and the Wilson (1987)-doctrine more generally, is relaxing the implicit common knowledge assumption to obtain more realistic models. Given the discussion in the previous subsection, the model presented here can be interpreted likewise, but in the realm of information design. In robust mechanism design, the implicit assumptions are relaxed by considering a sufficiently rich Harsanyi type structure. In contrast, in information design the Harsanyi type structure is the actual designed information structure. Mathevet et al. (2020) provide a method to study this design problem. My model can be interpreted as relaxing the common knowledge assumption about the designed information structure. But to remain in the realm of information design, the players still know their designed marginal information structure. The solution concept proposed in this chapter captures these assumptions exactly as explained in the previous section. In addition, the adversarial selection assumption reflects the robustness aspect.

3.6. Conclusions

One of the primary tasks of modern economies is the provision of information. In this chapter, I provide a method to study the question of how to optimally provide information when agreements are made bilaterally between the sender and the receiver. In the case of multiple receivers, which is quite common in the pharmaceutical industry, for example, receivers might engage in a strategic game to compete in their market. These strategic considerations should be taken into account by the information provider. Since the previous literature assumed that the information provider can fully commit to a grand information structure that becomes common knowledge among the receivers, I cannot directly apply these existing methods. The full commitment assumption is in direct contrast to the bilateral-contracting assumption.

This chapter has several contributions, which provide a general, yet tractable, method to study bilateral information design. First, I propose a new solution concept that captures all actions that can be rationally chosen for a player with a given information about the fundamental of the economy. Second, I contribute to information design by characterizing the set of possible distributions over beliefs that can arise from any information structure. In doing so, I develop novel extremal distributions that capture how dependent these beliefs can be. Finally, I combine each of these insights to develop a representation theorem that provides a simple method to study bilateral information design, assuming the designer is concerned about robustness to strategic uncertainty arising from the bilateral arrangement. I illustrate the main theorem by solving for the optimal information structure in a stylized problem faced by contract research organizations.

Bibliography

Ahn, D. S. (2007): "Hierarchies of ambiguous beliefs," *Journal of Economic Theory*, 136, 286–301.

Aliprantis, C. and K. Border (2006): *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer.

Alonso, R. and O. Câmara (2016): "Bayesian Persuasion with Heterogeneous Priors," *Journal of Economic Theory*, 165, 672–706.

ans Daisuke Oyama, S. M. and S. Takahashi (2020): "Implementation via Information Design in Binary-Action Supermodular Games," Mimeo.

Anscombe, F. J. and R. J. Aumann (1963): "A definition of subjective probability," *The Annals of Mathematical Statistics*, 34, 199–205.

Aradillas-Lopez, A. and E. Tamer (2008): "The identification power of equilibrium in simple games," *Journal of Business & Economic Statistics*, 26, 261–310.

Arieli, I., Y. Babichenko, F. Sandomirskiy, and O. Tamuz (2020): "Feasible Joint Posterior Beliefs," Mimeo.

Armbruster, W. and W. Böge (1979): "Bayesian game theory," in *Game theory and related topics*, North Holland, Amsterdam.

Arrow, K. J. (1951): "Alternative Approaches to the Theory of Choice in Risk-Taking Situations," *Econometrica*, 19, 404–437.

Artemov, G., T. Kunimoto, and R. Serrano (2013): "Robust Virtual Implementation: Toward a Reinterpretation of the Wilson Doctrine," *Journal of Economic Theory*, 148, 424–447.

Asheim, G. and M. Dufwenberg (2003): "Admissibility and common belief," *Games and Economic Behavior*, 42, 208–234.

Aumann, R. J. (1962): "Utility theory without the completeness axiom," *Econometrica*, 30, 445–462.

——— (1987): "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1–18.

Aumann, R. J. and A. Brandenburger (1995): "Epistemic conditions for Nash equilibrium," *Econometrica*, 63, 1161–1180.

Aumann, R. J. and J. H. Dreze (2008): "Rational Expectations in Games," *American Economic Review*, 98, 72–86.

Barelli, P. and S. Galanis (2013): "Admissibility and event rationality," *Games and Economic Behaviour*, 77, 21–40.

Battigalli, P. (1999): "Rationalizability in Incomplete Information Games," EUI Working Paper ECO No. 99/17.

——— (2003): "Rationalizability in infinite, dynamic games with incomplete information," *Research in Economics*, 57, 1–38.

Battigalli, P., A. Di Tillio, E. Grillo, and A. Penta (2011): "Interactive Epistemology and Solution Concepts for Games with Asymmetric Information," *The B.E. Journal of Theoretical Economics*, 11.

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

——— (2002): "Strong belief and forward induction reasoning," *Journal of Economic Theory*, 106, 356–391.

——— (2003): "Rationalization and Incomplete Information," *Advances in Theoretical Economics*, 3.

——— (2007): "Interactive Epistemology in Games with Payoff Uncertainty," *Research in Economics*, 61, 165–184.

Beauchêne, D., J. Li, and M. Li (2019): "Ambiguous Persuasion," *Journal of Economic Theory*, 179, 312–365.

Beissner, P. and F. Riedel (2019): "Equilibria Under Knightian Price Uncertainty," *Econometrica*, 87, 37–64.

Bergemann, D. and S. Morris (2005): "Robust Mechanism Design," *Econometrica*, 73, 1771–1813.

——— (2009): "Robust Implementation in Direct Mechanisms," *The Review of Economic Studies*, 76, 1175–1204.

——— (2011): "Robust implementation in General Mechanisms," *Games and Economic Behavior*, 71, 261 – 281.

——— (2013): "Robust Predictions in Games With Incomplete Information," *Econometrica*, 81, 1251–1308.

——— (2016): "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," *Theoretical Economics*, 11, 487–522.

——— (2017): "Belief-free Rationalizability and Informational Robustness," *Games and Economic Behavior*, 104, 744–759.

——— (2019): "Information Design: A Unified Perspective," *Journal of Economic Literature*, 57, 44–95.

Bergemann, D., S. Morris, and O. Tercieux (2011): "Rationalizable implementation," *Journal of Economic Theory*, 146, 1253–1274.

Bergemann, D. and M. Pesendorfer (2007): "Information Structures in Optimal Auctions," *Journal of Economic Theory*, 137, 580–609.

Bernheim, B. D. (1984): "Rationalizable strategic behaviour," *Econometrica*, 52, 1007–1028.

Bewley, T. F. (2002): "Knightian Decision Theory. Part I," *Decisions in Economics and Finance*, 25, 79–110.

Blackwell, D. (1951): "Comparison of Experiments," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, Calif.: University of California Press, 93–102.

——— (1953): "Equivalent Comparisons of Experiments," *Ann. Math. Statist.*, 24, 265–272.

Blume, L., A. Brandenburger, and E. Dekel (1991): "Lexicographic probabilities and choice under uncertainty," *Econometrica*, 59, 61–79.

Böge, W. and T. Eisele (1979): "On solutions of Bayesian games," *The International Journal of Game Theory*, 8, 193–215.

Börgers, T. (1994): "Weak dominance and approximate common knowledge," *Jornal of Economic Theory*, 64, 265–276.

Börgers, T. and J. Li (2019): "Strategically Simple Mechanisms," *Econometrica*, 87, 2003–2035.

Brandenburger, A. (1992): *Economic Analysis of Markets and Games*, MIT Press, chap. Lexicographic Probabilities and Iterated Admissibility, 282–290, editors: Partha Dasgupta, Douglas Gale, Oliver Hart, and Eric Maskin.

Brandenburger, A. and E. Dekel (1987): "Rationalizability and Correlated Equilibria," *Econometrica*, 55, 1391–1402.

——— (1993): "Hierarchies of beliefs and common knowledge," *Journal of Economic Theory*, 59, 189–198.

Brandenburger, A. and A. Friedenberg (2010): "Self-admissible sets," *Journal of Economic Theory*, 145, 785–811.

Brandenburger, A., A. Friedenberg, and H. J. Keisler (2008): "Admissibiltiy in Games," *Econometrica*, 76, 307–352.

Bresnahan, T. and J. Levin (2012): "Vertical Integration and Market Structure," in *The Handbook of Organizational Economics*, ed. by R. Gibbons and J. Roberts, Princeton University Press, 853–890.

Brocas, I. and J. D. Carrillo (2007): "Influence Through Ignorance," *The RAND Journal of Economics*, 38, 931–947.

Calzolari, G. and A. Pavan (2006): "On the Optimality of Privacy in Sequential Contracting," *Journal of Economic Theory*, 130, 168–204.

Cambanis, S., G. Simons, and W. Stout (1976): "Inequalities for $\mathbb{E}k(X,Y)$ when the Marginals are Fixed," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 36, 285–294.

Carroll, G. (2016): "Informationally Robust Trade and Limits to Contagion," *Journal of Economic Theory*, 166, 334–361.

——— (2019): "Robustness in Mechanism Design and Contracting," *Annual Review of Economics*, 11, 139–166.

Catonini, E. and N. De Vito (2018): "Cautious Belief and Iterated Admissibility," Mimeo.

——— (2020): "Weak belief and permissibility," *Games and Economic Behavior*, 120, 154–179.

Cettolin, E. and A. Riedl (2019): "Revealed preferences under uncertainty: Incomplete preferences and preferences for randomization," *Journal of Economic Theory*, 181, 547–585.

Chambers, R. and T. Melkonyan (2020): "Incomplete Preferences and Equilibrium in Contingent Markets," *Economica*, 87, 108–131.

Chambers, R. G. (2014): "Uncertain equilibria and incomplete preferences," *Journal of Mathematical Economics*, 55, 48–54.

Chen, Y.-C., X. Luo, and C. Qu (2016): "Rationalizability in general situations," *Economic Theory*, 61, 147–167.

Chiesa, A., S. Micali, and Z. A. Zhu (2015): "Knightian Analysis of the Vickrey Mechanism," *Econometrica*, 83, 1727–1754.

Cramton, P. and S. Kerr (2002): "Tradeable carbon permit auctions: How and why to auction not grandfather," *Energy and Policy*, 30, 333–345.

Dana, R.-A. and F. Riedel (2013): "Intertemporal equilibria with Knightian uncertainty," *Journal of Economic Theory*, 148, 1582–1605.

Dekel, E., A. Friedenberg, and M. Siniscalchi (2016): "Lexicographic beliefs and assumption," *Journal of Economic Theory*, 163, 955–985.

Dekel, E. and D. Fudenberg (1990): "Rational behavior with payoff uncertainty," *Journal of Economic Theory*, 52, 243–267.

Dekel, E., D. Fudenberg, and S. Morris (2007): "Interim Correlated Rationalizability," *Theoretical Economics*, 2, 15–40.

Dekel, E. and M. Siniscalchi (2015): "Chapter 12 - Epistemic Game Theory," in *Handbook of Game Theory with Economic Applications*, ed. by H. P. Young and S. Zamir, Elsevier, vol. 4, 619–702.

Dequiedt, V. and D. Martimort (2015): "Vertical Contracting with Informational Opportunism," *American Economic Review*, 105, 2141–2182.

DI TILLIO, A. (2008): "Subjective expected utility in games," *Theoretical Economics*, 3, 387–323.

DIJKSTRA, E. W. (1975): "How do we tell truths that might hurt?" Memo EWD 498.

DOMINIAK, A. AND B. C. SCHIPPER (2020): "Common Belief in Choquet Rationality with an 'Attitude'," Mimeo.

DWORCZAK, P. AND A. PAVAN (2020): "Robust (Bayesian) Persuasion," Mimeo.

EASLEY, D. AND M. O'HARA (2010): "Liquidity and valuation in an uncertain world," *Journal of Financial Economics*, 97, 1–11.

ELY, J. C. (2017): "Beeps," *American Economic Review*, 107, 31–53.

EPSTEIN, L. G. (1997): "Preference, Rationalizability and Equilibrium," *Journal of Economic Theory*, 73, 1–29.

EPSTEIN, L. G. AND T. WANG (1996): ""Beliefs about beliefs" without probabilities," *Econometrica*, 64, 1343–1373.

ESŐ, P. AND B. SZENTES (2007): "Optimal Information Disclosure in Auctions and the Handicap Auction," *The Review of Economic Studies*, 74, 705–731.

FEDDERSEN, T. AND W. PESENDORFER (1997): "Voting behavior and information aggregation in elections with private information," *Econometrica*, 65, 1029–1058.

FISHBURN, P. C. (1970): *Utility theory for decision making*, John Wiley & Sons.

Forges, F. (1993): "Five Legitimate Definitions of Correlated Equilibrium in Games with Incomplete Information," *Theory and Decision*, 35, 277–310.

——— (2006): "Correlated Equilibrium in Games with Incomplete Information Revisited," *Theory and Decision*, 61, 329–344.

Fréchet, M. (1951): "Sur les tableaux de corrélation dont les marges sont données," *Annales de l'Université de Lyon. Section A: Sciences mathématiques et astronomie*, 9, 53–77.

Friedenberg, A. (2010): "When do type structures contain all hierarchies of beliefs?" *Games and Economic Behavior*, 68, 108–129.

Galperti, S. (2019): "Persuasion: The Art of Changing Worldviews," *American Economic Review*, 109, 996–1031.

Germano, F. and P. Zuazo-Garin (2017): "Bounded Rationality and Correlated Equilibria," *International Journal of Game Theory*, 46, 595–629.

Gilboa, I., F. Maccheroni, M. Marinacci, and D. Schmeidler (2010): "Objective and subjective rationality in a multiple prior model," *Econometrica*, 78, 755–770.

Goldstein, I. and C. Huang (2016): "Bayesian Persuasion in Coordination Games," *American Economic Review*, 106, 592–596.

Grand View Research (2019): "Healthcare Contract Research Organization Market Report, 2019-2025," Report 978-1-68038-688-2, Grand View Research Inc., San Francisco, USA.

Harsanyi, J. C. (1962): "Rationality Postulates for Bargaining Solutions in Cooperative and in Non-Cooperative Games," *Management Science*, 9, 141–153.

——— (1967): "Games with Incomplete Information Played by 'Bayesian' Players, I-III. Part I. The Basic Model," *Management Science*, 14, 159–182.

Heifetz, A., M. Meier, and B. C. Schipper (2019): "Comprehensive Rationalizability," *Games and Economic Behavior*, 116, 185–202.

Hoeffding, W. (1940): "Masstabinvariante Korrelationstheorie," *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, 8, 179–233.

Hoshino, T. (2019): "Multi-Agent Persuasion: Leveraging Strategic Uncertainty," Mimeo.

Inostroza, N. and A. Pavan (2018): "Persuasion in Global Games with Application to Stress Testing," Mimeo.

Joe, H. (1997): *Multivariate Models and Multivariate Dependence Concepts*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis.

Kajii, A. and T. Ui (2009): "Interim efficient allocations under uncertainty," *Journal of Economic Theory*, 144, 337–353.

Kamenica, E. (2019): "Bayesian Persuasion and Information Design," *Annual Review of Economics*, 11, 249–272.

Kamenica, E. and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615.

Keisler, H. J. and B. Lee (2015): "Common assumption of rationality," Mimeo.

Keynes, J. M. (1936): *The general theory of employment, interest and money*, Harcourt Brace and Co.

Knight, F. H. (1921): *Risk, uncertainty and profit*, Houghton Mifflin.

Kohlberg, E. and J.-F. Mertens (1986): "On the strategic stability of equilibria," *Econometrica*, 54, 1003–1037.

Kokkala, J., K. Berg, K. Virtanen, and J. Poropudas (2019): "Rationalizable strategies in games with incomplete preferences," *Theory and Decision*, 86, 185–204.

Kosterina, S. (2019): "Persuasion with Unknown Beliefs," Mimeo.

Lee, B. S. (2016): "Admissibility and assumption," *Journal of Economic Theory*, 163, 42–72.

Lehrer, E. and R. Tepper (2011): "Justifiable prerefences," *Journal of Economic Theory*, 146, 762–774.

Liu, Q. (2015): "Correlation and Common Priors in Games with Incomplete Information," *Journal of Economic Theory*, 157, 49–75.

Lopomo, G., L. Rigotti, and C. Shannon (2011): "Knightian uncertainty and moral hazard," *Journal of Economic Theory*, 146, 1148–1172.

——— (2017): "Uncertainty in Mechanism Design," Mimeo.

Luce, R. and H. Raiffa (1957): *Games and Decisions: Introduction and Critical Survey*, Dover books on advanced mathematics, Dover Publications.

Mathevet, L., J. Perego, and I. Taneva (2020): "On Information Design in Games," *Journal of Political Economy*, 128, 1370–1404.

McAfee, R. P. and M. Schwartz (1994): "Opportunism in Multilateral Vertical Contracting: Nondiscrimination, Exclusivity, and Uniformity," *American Economic Review*, 84, 210–230.

Mertens, J.-F. and S. Zamir (1985): "Formulation of Bayesian analysis for games with incomplete information," *International Journal of Game Theory*, 14, 1–29.

Meyer, M. and B. Strulovici (2015): "Beyond Correlation: Measuring Interdependence Through Complementarities," Mimeo.

Milgrom, P. (1998): "Game theory and the spectrum auctions," *European Economic Review*, 42, 771–778.

Monderer, D. and D. Samet (1987): "Approximating common knowledge with common beliefs," *Games and Economic Behavior*, 1, 170–190.

Moulin, H. (1979): "Dominance solvable voting schemes," *Econometrica*, 47, 1337–1351.

Myerson, R. B. (1982): "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems," *Journal of Mathematical Economics*, 10, 67–81.

NELSEN, R. B. (2006): *An Introduction to Copulas*, Springer Series in Statistics, New York, USA: Springer-Verlag.

OLLÁR, M. AND A. PENTA (2017): "Full Implementation and Belief Restrictions," *American Economic Review*, 107, 2243–77.

PAHLKE, M. (2019): "A Note on Dynamic Consistency in Ambiguous Persuasion," Center for Mathematical Economics Working Papers 611, Center for Mathematical Economics, Bielefeld University.

PALFREY, T. R. AND S. SRIVASTAVA (1991): "Nash implementation using undominated strategies," *Econometrica*, 59, 479–501.

PEARCE, D. G. (1984): "Rationalizable strategic behavior and the problem of perfection," *Econometrica*, 52, 1029–1050.

PEREA, A. (2012): *Epistemic Game Theory: Reasoning and Choice*, Cambridge University Press.

PIERMONT, E. AND P. ZUAZO-GARIN (2020): "Heterogeneously Perceived Incentives in Dynamic Environments: Rationalization, Robustness and Unique Selections," Mimeo.

RÉNYI, A. (1955): "On a New Axiomatic Theory of Probability," *Acta Mathematica Academiae Scientiarum Hungarica*, 6, 285–335.

RIGOTTI, L. AND C. SHANNON (2005): "Uncertainty and Risk in Financial Markets," *Econometrica*, 73, 203–243.

Rockafellar, R. T. (1970): *Convex Analysis*, Princeton Landmarks in Mathematics and Physics, Princeton University Press.

Samuelson, L. (1992): "Dominated strategies and common knowledge," *Games and Economic Behavior*, 4, 284–313.

Savage, L. (1954): *The Foundations of Statistics*, Dover Books on Mathematics, Dover Publications.

Segal, I. and M. D. Whinston (2012): "Property Rights," in *The Handbook of Organizational Economics*, ed. by R. Gibbons and J. Roberts, Princeton University Press, 100–158.

Shaked, M. and J. Shanthikumar (2007): *Stochastic Orders*, Springer Series in Statistics, Springer New York.

Sobel, J. (2017): "A note on pre-play communication," *Games and Economic Behavior*, 102, 477–486.

——— (2019): "Iterated weak dominance and interval-dominance supermodular games," *Theoretical Economics*, 14, 71–102.

Stahl, D. O. (1995): "Lexicographic rationalizability and iterated admissibility," *Economics Letters*, 47, 155–159.

Stauber, R. (2011): "Knightian games and robustness to ambiguity," *Journal of Economic Theory*, 146, 248–274.

——— (2014): "A framework for robustness to ambiguity of higher-order beliefs," *International Journal of Game Theory*, 43, 525–550.

Tan, T. and S. da Costa Werlang (1988): "The Bayesian foundations of solution concepts of games," *Journal of Economic Theory*, 45, 370–391.

Taneva, I. (2019): "Information Design," *American Economic Journal: Microeconomics*, 11, 151–185.

Tang, Q. (2015): "Hierarchies of Beliefs and the Belief-invariant Bayesian Solution," *Journal of Mathematical Economics*, 59, 111–116.

Tchen, A. H. (1980): "Inequalities for Distributions with Given Marginals," *Annals of Probability*, 8, 814–827.

van Zandt, T. and X. Vives (2007): "Monotone Equilibria in Bayesian Games of Strategic Complementarities," *Journal of Economic Theory*, 134, 339–360.

Varian, H. R. (2007): "Position auctions," *International Journal of Industrial Organizattion*, 25, 1163–1178.

von Goethe, J. W. (1841): *Faust: Eine Tragödie. Erster Theil*, Hermann Passarge.

Wald, A. (1939): "Contributions to the Theory of Statistical Estimation and Testing Hypotheses," *The Annals of Mathematical Statistics*, 10, 299–326.

——— (1949): "Statistical Decision Functions," *Ann. Math. Statist.*, 20, 165–205.

WILSON, R. (1987): "Game-Theoretic Analyses of Trading Processes," in *Advances in Economic Theory: Fifth World Congress*, ed. by T. Bewley, Cambridge University Press, Econometric Society Monographs, 33–70.

YANG, C.-C. (2015): "Weak assumption and iterated admissibilty," *Journal of Economic Theory*, 158, 87–101.

ZIEGLER, G. AND P. ZUAZO-GARIN (2020): "Strategic cautiousness as an expression of robustness to ambiguity," *Games and Economic Behavior*, 119, 197–215.

Vita

Gabriel Ziegler was born on November 19, 1987 in Styria, Austria. He is an Austrian citizen and grew up in Fernitz, Austria. In 2007 he graduated with distinction from HTBLA Kaindorf with a specialization in organization and information technology. After his mandatory service in the Austrian Armed Forces, he enrolled as a student at the University of Graz, Austria. First, he received a Bachelor of Science in Business Administration in 2010 and then continued to receive a Bachelor of Arts in Economics in 2012 with distinction. While studying in Graz, he spent a year at the University of Minnesota, USA as an exchange student. After his undergraduate studies, he moved to Vienna, Austria to earn a Master of Science in Economics from a join program of the Institute for Advanced Studies and Vienna University of Technology. Subsequently, he moved to Evanston, Illinois to enroll as a PhD Student of Economics at the Northwestern University. He received a Master of Arts in Economics from Northwestern University in 2016.

He is a member of the Econometric Society, the European Economic Association, the Game Theory Society, and the Society for the Advancement of Economic Theory. His research focuses on foundations of game theory and most recently on information design. One paper on the former was published in *Games and Economic Behavior* and is also part of this dissertation. During his studies he gained teaching experience by serving as a teaching assistant in classes on various topics in microeconomics at the graduate and undergraduate level.