Epistemic game theory*

Eddie Dekel and Marciano Siniscalchi
Tel Aviv and Northwestern University, and Northwestern University

January 14, 2014

Contents

*We thank Drew Fudenberg for his detailed comments, and Robert Molony and Luciano Pomatto for excellent research assistantship. We also thank Pierpaolo Battigalli, Adam Brandenburger, Yi-Chun Chen, Amanda Friedenberg, Joe Halpern, Qingmin Liu, Andres Perea, two anonymous referees, and the editors, Peyton Young and Shmuel Zamir, for helpful feedback. Eddie Dekel gratefully acknowledges financial support from NSF grant SES-1227434.

1 Introduction and Motivation

Epistemic game theory formalizes assumptions about rationality and mutual beliefs in a formal language, then studies their behavioral implications in games. Specifically, it asks: what do different notions of rationality and different assumptions about what players believe about ... what others believe about the rationality of players imply regarding play in a game? A well-known example is the equivalence between common belief in rationality and iterated deletion of dominated strategies.

The reason why it is important to be formal and explicit is the standard one in economics. Solution concepts are often motivated intuitively in terms of players' beliefs and their rationality.

However, the epistemic analysis may show limitations in these intuitions, reveal what additional assumptions are hidden in the informal arguments, clarify the concepts or show how the intuitions can be generalized. We now consider a number of examples.

Backwards induction was long thought to be obviously implied by "common knowledge of rationality." The epistemic analysis showed flaws in this intuition and it is now understood that the characterization is much more subtle (Sections 7.4.3 and 7.5).

Next, consider the solution concept that deletes one round of weakly dominated strategies and then iteratively deletes strictly dominated strategies. This concept was first proposed because it is robust to payoff perturbations, which were interpreted as a way to perturb players' rationality. Subsequent epistemic analysis showed this concept is exactly equivalent to "almost common belief" of rationality and of full-support conjectures – an explicit robustness check of common belief in rationality (see Section 5). Thus the epistemic analysis generalizes and formalizes the connection of this concept to robustness.

The common-prior assumption (Section 4.3) is used to characterize Nash equilibrium with $n > 2$ players, but not needed for two-player games (compare Theorems 5 and 7). This result highlights the difference between the concept in these environments. Furthermore, the common-prior is known to be equivalent to no betting when uncertainty is exogenous. We argue that the interpretation of the common-prior assumption and its connection to no-betting results must be modified when uncertainty is endogenous, e.g, about players' strategies (see Example 4).

Finally, recent work has shown how forward induction and iterated deletion of weakly dominated strategies can be characterized. These results turn out to identify important, non-obvious, assumptions and require new notions of "beliefs." Moreover, they clarify the connection between these concepts. (See Section 7.4.4.)

Epistemic game theory may also help provide a rationale, or 'justification,' for or against specific solution concepts. For instance, in Section 6 we identify those cases where interim independent rationalizability is and is not a "suitable" solution concept for games of incomplete information.

We view non-epistemic justifications for solution concepts as complementary to the epistemic approach. For some solution concepts, such as forward induction, we think the epistemic analysis is more insightful. For others, such as Nash equilibrium, learning theory may provide the more compelling justification. Indeed, we do not find the epistemic analysis of objective equilibrium notions (Section 4) entirely satisfactory. This is because the epistemic assumptions needed are often very strong and hard to view as a justification of a solution concept. Moreover, except for special cases (e.g. pure-strategy Nash equilibrium), it is not really possible to provide necessary and sufficient epistemic conditions for equilibrium *behavior* (unless we take the view that mixed strategies are actually available to the players). Rather, the analysis constitutes a fleshing-out of the textbook interpretation of equilibrium as 'rationality plus correct beliefs.' To us this suggests that equilibrium behavior cannot arise out of strategic reasoning alone. Thus, as discussed above, this epistemic analysis serves the role of identifying where alternative approaches are required to justify standard concepts.

While most of the results we present are known from the literature, we sometimes present them differently, to emphasize how they fit within our particular view. We have tried to present a wide swath of the epistemic literature, analyzing simultaneous-move games as well as dynamic games, considering complete and incomplete information games, and exploring both equilibrium and non-equilibrium approaches. That said, our choice of specific topics and results is still quite selective and we admit that our selection is driven by the desire to demonstrate our approach (discussed next), as well as our interests and tastes. Several insightful and important papers could not be included because they did not fit within our narrative. More generally, we have ignored several literatures. The connection with the robustness literature mentioned above (see Kajii and Morris

(1997b) for a survey) is not developed. Nor do we study self-confirming based solution concepts (Fudenberg and Levine, 1993; Battigalli, 1987; Rubinstein and Wolinsky, 1994).[1] Moreover, we do not discuss epistemics and k-level thinking (Crawford, Costa-Gomes, and Iriberri, 2012; Kets, 2012) or unawareness (see Schipper, 2013, for a comprehensive bibliography). We find all this work interesting, but needed to narrow the scope of this paper.

1.1 Philosophy/Methodology

The basic premise of this chapter is that the primitives of the model should be observable, at least in principle. The primitives of epistemic game theory are players' beliefs about the play of the game, their beliefs about players' beliefs about play, etc.; these are called *hierarchies of beliefs*. Obviously these cannot be observed directly, but we can ask that they be *elicitable* from observable choices, e.g., their betting behavior, as is standard in decision theory (De Finetti, 1992; Savage, 1972).

However, there are obvious difficulties with eliciting a players' beliefs about his own behavior and beliefs. Our basic premise then requires that we consider hierarchies of beliefs over *other* players' beliefs and rationality, thereby ruling out "introspective" beliefs (see also Section 2.6.3). With this stipulation it *is* possible to elicit such hierarchies of belief; see Section 2.6.2.

By contrast much of the literature, following Aumann's seminal developments, allows for introspective beliefs (Aumann, 1987). This modeling difference does have implications, in particular in characterization results that involve the common prior assumption (Theorems 4 and 8).

Rather than working with belief hierarchies directly, we use a convenient modeling device due to Harsanyi (1967), namely *type structures*. In the simple case of strategic-form games, these specify a set of "types" for each player, and for each type a belief over the opponents' strategies and types. Every type generates a hierarchy of beliefs over strategies, and conversely every hierarchy can be generated in some type structure; details are provided in Sections 2.3 and 2.4.

We emphasize that we use type structures solely as a modeling device. Types are not real-world objects; they simply represent hierarchies, which are. Therefore, although we will formally state epistemic assumptions on types, we will consider only those assumptions that can also be stated as restrictions on belief hierarchies, and we will interpret them as such. In particular, our assumptions cannot differentiate between two types that generate the same belief hierarchy. One concrete implication of this can be seen in the analysis of solution concepts for incomplete-information games (Section 6.1).

To clarify this point further, note that type structures can be used in a different way. In particular, they can be used to represent an information structure: in this case, a type represents the hard information a player can receive—for example, a possible outcome of some study indicting the value of an object being auctioned. Here, it makes perfect sense to distinguish between two types with different hard information, even if the two pieces of information lead to the same value for the object, and indeed the same belief hierarchy over the value of the object. However, in this chapter, types will *only* be used to represent hierarchies of belief, without any hard information.[2]

Finally, it is important to understand how to interpret epistemic results. One interpretation would go as follows. Assume we have elicited a player's hierarchy of beliefs. The theorems identify testable assumptions that determine whether that player's behavior is consistent with a particular solution concept. We do not find this interpretation very interesting: once we have elicited a player's

[1] The concept of RPCE (Fudenberg and Kamada, 2011) is a recent example where epistemics seem to us useful. Its definition is quite involved and, while examples illustrate the role of various assumptions, the epistemic analysis confirms the equivalence of the solution concept to the assumptions used in its intuitive description.

[2] We can add hard information to our framework, at the cost of notational complexity: see footnote 47.

hierarchy we know her best replies, so it is pointless to invest effort to identify what assumptions are satisfied. Instead our preferred interpretation of the results are as statements about play that follow without knowing the exact hierarchy. That is, the theorems we present answer the following question: if all we knew about the hierarchies of beliefs was that they satisfied certain assumptions, what would we be able to say about play? Naturally we cannot identify necessary conditions: a player might play a Nash equilibrium strategy "just because" he wanted to. (There is, however, a sense in which the results we present provide necessary conditions as well: see the discussion in Section 3.2.)

2 Main ingredients

In this section we introduce the basic elements of our analysis. We begin with notation and a formal definition of strategic-form games, continue with hierarchies of beliefs and type structures, and conclude with rationality and beliefs.

2.1 Notation

For any finite set Y, let $\Delta(Y)$ denote the set of probability distributions over Y and any subset E of Y is an event. For $Y' \subset Y$, $\Delta(Y')$ denotes the set of probabilities on Y that assign probability 1 to Y'. The support of a probability distribution $p \in \Delta(Y)$ is denoted by supp p. Finally, we adopt the usual conventions for product sets: given sets X_i, with $i \in I$, we let $X_{-i} = \prod_{j \neq i} X_j$ and $X = \prod_{i \in I} X_i$.

All our characterization theorems include results for which infinite sets are not required. However, infinite sets are needed to formally present hierarchies of beliefs, their relationship to type structures and for part of the characterization results. To minimize technical complications infinite sets are assumed to be compact metric spaces endowed with the Borel sigma algebra. We denote by $\Delta(Y)$ the set of Borel probability measures on Y and endow $\Delta(Y)$ with the weak convergence topology.[3] Cartesian product sets are endowed with the product topology and the product sigma-algebra. Events are a measurable subsets of Y.

2.2 Strategic-form games

We define finite strategic-form games and best replies.

Definition 1 *A (finite)* ***strategic-form game*** *is a tuple $G = (I, (S_i, u_i)_{i \in I})$, where I is finite and, for every $i \in I$, S_i is finite and $u_i : S_i \times S_{-i} \to \mathbb{R}$.*

As is customary, we denote expected utility from a mixed strategy of i, $\sigma_i \in \Delta(S_i)$, and a belief over strategies of opponents, $\sigma_{-i} \in \Delta(S_{-i})$, by $u_i(\sigma_i, \sigma_{-i})$. We take the view that players always choose *pure* strategies. On the other hand, certain standard solution concepts are defined in terms of mixed strategies. In the epistemic analysis, mixed strategies of i are replaced by strategic uncertainty of i's opponents, that is, their beliefs about i's choice of a pure strategy. We allow for mixed strategies as actual choices only when there is an explicit mixing device appended to the game.

Definition 2 *Fix a game $(I, (S_i, u_i)_{i \in I})$. A strategy $s_i \in S_i$ is a* ***best reply*** *to a belief $\sigma_{-i} \in \Delta(S_{-i})$ if, for all $s'_i \in S_i$, $u_i(s_i, \sigma_{-i}) \geq u_i(s'_i, \sigma_{-i})$; the belief σ_{-i} is said to* ***rationalize*** *strategy s_i.*

[3] For detailed definitions see, e.g., Billingsley (2008).

2.3 Belief hierarchies

The essential element of epistemic analysis is the notion of hierarchies of belief. These are used to define rationality and common belief in rationality, which are then used to characterize solution concepts. A belief hierarchy specifies a player's belief over the basic space of uncertainty (e.g., opponents' strategies), her beliefs over opponents' beliefs, etc..

To formally describe belief hierarchies, we first specify the basic space of uncertainty X_{-i} for each player i. In the epistemic analysis of a strategic-form game, the basic uncertainty is over the opponents' strategies, so $X_{-i} = S_{-i}$. More generally, we will allow for *exogenous* uncertainty as well, which is familiar from the textbook analysis of incomplete-information games. For instance, in a common-value auction, each player i is uncertain about the value of the object, so X_{-i} includes the set of possible values.

Once the sets X_{-i} have been fixed, each player i's hierarchy of beliefs is a sequence of probability measures $(p_i^1, p_i^2, \ldots)$. It is simpler to discuss these beliefs in the case of two players. Player i's *first-order belief* p_i^1 is a measure over the basic domain X_{-i}: $p_i^1 \in \Delta(X_{-i})$. Player i's *second-order belief* p_i^2 is a measure over the Cartesian product of X_{-i} and the set of all possible first-order beliefs for player $-i$: that is, $p_i^2 \in \Delta(X_{-i} \times \Delta(X_i))$, where X_i is the domain of $-i$'s first-order beliefs. The general form of this construction is as follows. First, let $X_{-i}^0 = X_{-i}$ for each player $i = 1, 2$; then, inductively, for each $k = 1, 2, \ldots$, let

$$X_{-i}^k = X_{-i}^{k-1} \times \Delta(X_i^{k-1}). \tag{1}$$

Then, for each $k = 1, 2, \ldots$, the domain of player i's k-th order beliefs is X_{-i}^{k-1}. Consequently, the set of all belief hierarchies for player i is $H_i^0 = \prod_{k \geq 0} \Delta(X_{-i}^k)$; the reason for the superscript "0" will be clear momentarily.

Note that, for $k \geq 2$, the domain of i's k-th order beliefs includes the domain of her $(k-1)$-th order beliefs. For instance, $p_i^2 \in \Delta(X_{-i} \times \Delta(X_i))$, so the marginal of p_i^2 also specifies a belief for i over X_{-i}, just like p_i^1. For an arbitrary hierarchy $(p_i^1, p_i^2, \ldots)$, these beliefs may differ. The reader may then wonder why we did not define i's second-order beliefs just over her opponent's first-order beliefs, i.e., as measures over $\Delta(X_i)$ rather than $X_{-i} \times \Delta(X_i)$.

Intuitively, the reason is that we need to allow for correlation in i's beliefs over X_{-i} and $-i$'s beliefs over X_i. Specifically, consider a simple 2×2 coordination game, with strategy sets $S_i = \{H, T\}$ for $i = 1, 2$. Suppose that the analyst is told that player 1 (i) assigns equal probability to player 2 choosing H or T, and (ii) also assigns equal probability to the events "player 2 believes that 1 chooses H" and "player 2 believes that 1 chooses T" (where by "believes that" we mean "assigns probability one to the event that"). Can the analyst decide whether or not player 1 believes that player 2 is rational? The answer is negative. Given the information provided, it may be the case that player 1 assigns equal probability to the events "player 2 plays H and believes that 1 plays T" and "player 2 plays T and believes that 1 plays H."

To sum up, i's second-order belief p_i^2 must be an element of $\Delta(X_{-i} \times \Delta(X_i))$. Hence, we need to make sure that its marginal on X_{-i} coincides with i's first-order belief p_i^1. More generally, we restrict attention to *coherent* belief hierarchies, i.e. sequences $(p_i^1, p_i^2, \ldots) \in H_i^0$ such that, for all $k \geq 2$,

$$\mathrm{marg}_{X_{-i}^{k-2}} p_i^k = p_i^{k-1}. \tag{2}$$

Let H_i^1 denote the subset of H_i^0 consisting of coherent belief hierarchies.

Brandenburger and Dekel (1993) use Kolmogorov's theorem (see Dellacherie and Meyer (1978), p. 68, or Aliprantis and Border (2007) Section 15.6) to show that there exists a homeomorphism

$$f_i : H_i^1 \to \Delta\left(X_{-i} \times H_{-i}^0\right) \tag{3}$$

that "preserves beliefs" in the sense that for $h_i = \left(p_i^k\right)_{k=1}^{\infty}$, $\mathrm{marg}_{X_{-i}^k} f_i\left(h_i\right) = p_i^{k+1}$. To understand this, first note that f_i maps a coherent hierarchy h_i into a belief over i's basic space of uncertainty, X_{-i}, and $-i$'s hierarchies, H_{-i}^0. Therefore, we want this mapping to preserve i's first-order beliefs. In particular, h_i's first-order beliefs should equal the marginal of $f_i(h_i)$ on X_{-i}. Now consider second-order beliefs. Recall that $H_{-i}^0 = \prod_{\ell \geq 0} \Delta(X_i^\ell) = \Delta(X_i^0) \times \prod_{\ell \geq 1} \Delta(X_i^\ell)$. Therefore, $X_{-i} \times H_{-i}^0 = X_{-i} \times \Delta(X_i) \times \prod_{\ell \geq 1} \Delta(X_i^\ell) = X_{-i}^1 \times \prod_{\ell \geq 1} \Delta(X_i^\ell)$. Hence we can view $f_i(h_i)$ as a measure on $X_{-i}^1 \times \prod_{\ell \geq 1} \Delta(X_i^\ell)$, so we can consider its marginal on X_{-i}^1. Preserving beliefs means that this marginal is the same as i's second-order belief p_i^2 in the hierarchy h_i. Higher-order beliefs are similarly preserved.

The function f_i in Equation 3 maps coherent hierarchies of player i to beliefs about the basic uncertainty X_{-i} and the hierarchies of the other player, H_{-i}^0. Thus, in a sense, f_i determines a "first-order belief" over the expanded space of uncertainty $X_{-i} \times H_{-i}^0$. However, since f_i is onto, some coherent hierarchies of i assign positive probability to *in*coherent hierarchies of $-i$. These hierarchies of $-i$ do *not* correspond to beliefs over $X_i \times H_i^0$. Therefore, there are coherent hierarchies of i for which "second-order beliefs" over the expanded space $X_{-i} \times H_{-i}^0$ are not defined. To address this, we impose the restriction that coherency is "common belief"; that is, we restrict attention to

$$H_i = \cap_{k=0}^{\infty} H_i^k, \tag{4}$$

where for $k > 0$ $H_i^k = \left\{h_i \in H_i^{k-1} : f_i\left(h_i\right)\left(X_{-i} \times H_{-i}^{k-1}\right) = 1\right\}$. It can then be shown that the function f_i in Eq. (3) restricted to H_i is one-to-one and onto $\Delta(X_{-i} \times H_{-i})$.[4] In the next subsection we will interpret the elements of H_i as "types." With this interpretation, that f_i is one-to-one means that distinct types have distinct beliefs over X_{-i} and the opponent's types. That f_i is onto means that any belief about X_{-i} and the opponent's types is held by some type of i.

It is important to note that belief hierarchies are elicitable via bets. We elaborate on this point in Section 2.6.

2.4 Type structures

As Harsanyi noted, type structures provide an alternative way to generate hierarchies of beliefs. A type structure specifies for each player i the space X_{-i} over which i has uncertainty, the set T_i of types of i, and each type t_i's hierarchy of beliefs, $\beta_i\left(t_i\right)$.[5]

Definition 3 *For every player $i \in I$, fix a compact metric space X_{-i}. An $(X_{-i})_{i \in I}$-based* ***type structure*** *is a tuple $\mathcal{T} = (I, (X_{-i}, T_i, \beta_i)_{i \in I})$ such that each T_i is a compact metric space and each $\beta_i : T_i \to \Delta(X_{-i} \times T_{-i})$ is continuous.*[6] *A type structure is* ***complete*** *if the maps β_i are onto.*

We discuss the notion of completeness immediately before Definition 7.

An *epistemic type structure* for a strategic-form game of complete information models players' strategic uncertainty: hierarchies are defined over opponents' strategies. This is just a special case

[4] For further details on the construction of belief hierarchies, see Armbruster and Böge (1979), Böge and Eisele (1979) Mertens and Zamir (1985), Brandenburger and Dekel (1993), Heifetz (1993) and Heifetz and Samet (1998), among others.

[5] As we discussed in the Introduction, in this definition players do not have *introspective* beliefs—that is, beliefs about their own strategies and beliefs: see Sec. 2.6 for a discussion of this modeling choice.

[6] The topological assumptions we adopt are for convenience; we do not seek generality. For instance, compactness of the type spaces and continuity of the belief maps β_i provides an easy way to show that sets corresponding to assumptions such as, "Player i is rational," or "Player i believes that Player j is rational," are closed, and hence measurable.

of Def. 3. However, since epistemic type structures play a central role in this chapter, we provide an explicit definition for future reference. Also, when it is clear from the context, we will omit the qualifier 'epistemic.'

Definition 4 *An* ***epistemic type structure*** *for the complete-information game* $G = (I, (S_i, u_i)_{i\in I})$ *is a type structure* $\mathcal{T} = (I, (X_{-i}, T_i, \beta_i)_{i\in I})$ *such that* $X_{-i} = S_{-i}$ *for all* $i \in I$.

Given an (epistemic) type structure $\mathcal{T}$ we can assess the belief hierarchy of each type t_i. As discussed earlier, type t_i's first-order belief is what she believes about S_{-i}; her second-order belief is what she believes about S_{-i} and about other player j's beliefs about S_{-j}, and so on. Also recall (Eq. 4) that the set of all hierarchies of beliefs over strategies for a player i is denoted by H_i.

Definition 5 *Given a type structure* $\mathcal{T}$*, the function mapping types into hierarchies is denoted by* $\varphi_i(\mathcal{T}) : T_i \to H$. *The type structure* $\mathcal{T}$ *is redundant if there are two types* $t_i, t_i' \in T_i$ *with the same hierarchy, i.e., such that* $\varphi_i(\mathcal{T})(t_i) = \varphi_i(\mathcal{T})(t_i')$*; such types are also called redundant.*

When the type structure $\mathcal{T}$ is clear from the context, we will write $\varphi_i(\cdot)$ instead of $\varphi_i(\mathcal{T})(\cdot)$.

Because a type's first-order beliefs – those over S_{-i} – play a particularly important role, it is convenient to introduce specific notation for them:

Definition 6 *The* ***first-order beliefs map*** $f_i : T_i \to \Delta(S_{-i})$ *is defined by* $f_i(t_i) = \mathrm{marg}_{S_{-i}}\beta_i(t_i)$ *for all* $t_i \in T_i$.

Example 1 We illustrate these notions using a finite type structure.

	L	C	R
T	2,1	3,1	0,0
M	4,3	0,2	4,0
B	3,0	1,2	2,5

	(L,t_2^1)	(L,t_2^2)	(C,t_2^1)	(C,t_2^2)	(R,t_2^1)	(R,t_2^2)
$\beta_1(t_1^1)$	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0	0
$\beta_1(t_1^2)$	1	0	0	0	0	0

	(T,t_1^1)	(T,t_1^2)	(M,t_1^1)	(M,t_1^2)	(B,t_1^1)	(B,t_1^2)
$\beta_2(t_2^1)$	1	0	0	0	0	0
$\beta_2(t_2^2)$	0	0	$\frac{1}{2}$	0	0	$\frac{1}{2}$

Figure 1: A strategic-form game and an epistemic type structure

In the type structure on the right-hand side of Fig. 1, type t_1^1 of player 1 (the row player) assigns equal probability to player 2 choosing L and C: these are type t_1^1's first-order beliefs. Similarly, the first-order beliefs of type t_1^2 of player 1 assign probability one to player 2 choosing L. The second-order beliefs of player 1's types are straightforward, because both t_1^1 and t_1^2 assign probability one to t_2^1, and hence to the event that player 2 is certain that (i.e., assigns probability one to the event that) 1 chooses T. Thus, for example, the second-order beliefs of type t_1^1 are that, with probability $\frac{1}{2}$, player 2 chooses L and believes that 1 chooses T, and with probability $\frac{1}{2}$, player 2 chooses C and believes that 1 chooses T.

Now consider type t_2^2 of player 2, who assigns equal probability to the pairs (M, t_1^1) and (B, t_1^2). This type's first-order beliefs are thus that player 1 is equally likely to play M or B; his second-order beliefs are that, with equal probability, either (i) player 1 plays M and expects player 2 to choose L and C with equal probability, or (ii) player 1 plays B and is certain that 2 plays L. We can easily describe type t_2^2's third-order beliefs as well: this type believes that, with equal probability, either (i) player 1 plays M, expects 2 to choose L and C with equal probability, and is certain that 2 is certain that 1 plays T; or (ii) player 1 plays B, is certain that 2 chooses L, and is certain that 2 is certain that 1 plays T. ■

A number of questions arise in connection with type structures. Is there a type structure that generates *all* hierarchies of beliefs? Is there a type structure into which any other type structure can be embedded?[7] Is a given type structure complete, as in Definition 3, i.e., such that every belief over Player i's opponents' strategies and types is generated by some type of Player i? These are all versions of the same basic question: is there a rich enough type structure that allows for "all possible beliefs?" We ask this question because we take beliefs as primitive objects; hence, we want to make sure that using type structures as a modeling device does not rule out any beliefs.

Under our assumptions on the sets X_{-i}, the answer to these questions is affirmative. Indeed, we can consider H_i (defined in Eq. 4) as a set of type profiles and define $\mathcal{T} = (I, (X_{-i}, T_i, \beta_i)_{i\in I})$ where $T_i = H_i$ and $\beta_i = f_i$ (where f_i was defined in Eq. 3). This is the "largest" non-redundant type structure, that generates all hierarchies, embeds all other type structures, and is complete.[8]

Once again, type structures are devices and belief hierarchies are the primitive objects of interest. Therefore, asking whether a player's hierarchy of beliefs "resides" in one type structure or another is meaningless. In particular, we cannot ask whether it "resides" in a rich type structure. We state results regarding the implications of epistemic assumptions in both rich and arbitrary type structures. The interest in rich type structures is twofold. First, one of convenience: while rich type structures are uncountable and complex mathematical objects, the fact that they are complete simplifies the statements of our characterization results. The second appeal of rich type structure is methodological: because they generate *all* hierarchies, they impose no implicit assumption on beliefs. Any smaller type structure does implicitly restrict beliefs; we explain this point in Section 7.4.4, because it is particularly relevant there. On the other hand, small (in particular, finite) type structures are convenient to discuss examples of epistemic conditions and characterization results.

2.5 Rationality and belief

We can now define rationality (expected payoff maximization) and belief, by which we mean "belief with probability one."

Definition 7 *Fix a type structure $(I, (S_{-i}, T_i, \beta_i)_{i\in I})$ for a strategic game $(I, (S_i, u_i)_{i\in I})$. For every player $i \in I$:*

1. *strategy $s_i \in S_i$ is **rational** for type $t_i \in T_i$ if it is a best reply to $f_i(t_i)$; let*
$$R_i = \{(s_i, t_i) \in S_i \times T_i \,:\, s_i \text{ is rational for } t_i\}.$$

2. *type $t_i \in T_i$ **believes** event $E_{-i} \subset S_{-i} \times T_{-i}$ if $\beta_i(t_i)(E_{-i}) = 1$; let*
$$B_i(E_{-i}) = \{(s_i, t_i) \in S_i \times T_i \,:\, t_i \text{ believes } E_{-i}\}.$$

Note that R_i, the set of strategy-type pairs of i that are rational for i, is defined as a subset of $S_i \times T_i$, rather than a subset of $S \times T$. This is notationally convenient, and also emphasizes that R_i is an assumption about Player i alone.

For any event $E_{-i} \subset S_{-i} \times T_{-i}$, $B_i(E_{-i})$ represents the types of i that believe E_{-i} obtains. It is convenient to define it as an event in $S_i \times T_i$, but it is clear from the definition that no restriction

[7] We are not going to formally define the relevant notion of embedding. Roughly speaking, it is that each type in one type space can be mapped to a type in the other in such a way as to preserve hierarchies of beliefs.

[8] Because we restrict attention to compact type spaces and continuous belief maps, these notions of "richness" are all equivalent. See the references in Footnote 4 for details, as well as Friedenberg (2010). In particular, it is sufficient that X_{-i} and T_i are compact metrizable, and that the sets T_i are non-redundant.

is imposed on i's strategies.[9] We abuse terminology and for sets $E_i \subset S_i \times T_i$ write "t_i in E_i" if there exists s_i such that $(s_i, t_i) \in E_i$.

The map associating with each event $E_{-i} \subset S_{-i} \times T_{-i}$ the subset $B_i(E_{-i})$ is sometimes called Player i's *belief operator*. While we do not develop a formal syntactic analysis, we do emphasize two important related properties satisfied by probability-one belief, *Monotonicity* and *Conjunction*:[10] for all events $E_{-i}, F_{-i} \subset S_{-i} \times T_{-i}$,

$$E_{-i} \subset F_{-i} \quad \Rightarrow \quad B_i(E_{-i}) \subset B_i(F_{-i}) \qquad \text{and} \qquad B_i(E_{-i} \cap F_{-i}) = B_i(E_{-i}) \cap B_i(F_{-i}). \quad (5)$$

Finally, we define mutual and common belief. Consider events $E_i \subset S_i \times T_i$ (with $E = \prod_i E_i$ and $E_{-i} = \prod_{j \neq i} E_i$ as usual). Then the events "E is *mutually believed*," "*k-th order believed*," and "*commonly believed*" are

$$B^1(E) = B(E) = \prod_{i \in I} B_i(E_{-i}), \quad B^k(E) = B\left(B^{k-1}(E)\right) \text{ for } k > 1, \quad CB(E) = \bigcap_{k \geq 1} B^k(E). \quad (6)$$

Note that each of these events is a Cartesian product of subsets of $S_i \times T_i$ for $i \in I$, so we can write $B_i^k(E)$ and $CB_i(E)$ for the i-th projection of these events.[11]

2.6 Discussion

The above definitions of a game, type structure, rationality, and belief all incorporate the assumption that players have state-independent expected-utility preferences. This modeling assumption raises three issues, discussed next: relaxing state independence, relaxing expected utility, and eliciting beliefs. Another modeling assumption discussed subsequently is that the type structure in principle allows any strategy to be played by any type. We conclude this discussion section by commenting on our use of semantic models rather than the alternative syntactic approach.

2.6.1 State dependence and non-expected utility

A more general definition of a game would specify a consequence for each strategy profile, and a preference relation over acts that map opponents' strategies into consequences.[12] Maintaining the expected-utility assumption one could allow for state dependence: the ranking of consequences may depend on the opponents' strategies (as in Morris and Takahashi, 2011). One could also allow for a richer model where preferences may be defined over opponents' beliefs (as in Geanakoplos, Pearce, and Stacchetti, 1989) or preferences (as in Gul and Pesendorfer, 2010), as well as material consequences. All these interesting directions lie beyond the scope of this chapter.

Moreover, type structures can also be defined without making the expected-utility assumption. Some generalizations of expected utility are motivated by refinements: in particular, lexicographic beliefs (Blume, Brandenburger, and Dekel, 1991) and conditional probability systems (Myerson,

[9] That is, if $(s_i, t_i) \in B_i(E_{-i})$ for some $s_i \in S_i$, then $(s'_i, t_i) \in B_i(E_{-i})$ for all $s'_i \in S_i$.

[10] We can split the "=" in the Conjunction property into two parts, "$\subset$" and "$\supset$." It is easy to see that the "$\subset$" part is equivalent to Monotonicity for any operator, no matter how it is defined.

The "p-belief," "strong belief" and "assumption"operators we consider in Sections 5, 7 and 8 respectively do not satisfy Monotonicity and hence the "$\subset$" part of Conjunction—a fact that has consequences for the epistemic analysis conducted therein.

[11] Thus $CB(E) = \bigcap_{k \geq 1} B^k(E) = \prod_i B_i(E_{-i}) \cap \bigcap_{k \geq 2} \prod_i B_i(B^{k-1}_{-i}(E)) = \prod_i \left(B_i(E_{-i}) \cap \bigcap_{k \geq 2} B_i(B^{k-1}_{-i}(E))\right) \equiv \prod_i CB_i(E)$. Hence, $CB_i(E) = B^1_i(E) \cap \bigcap_{k \geq 2} B_i(B^{k-1}_{-i}(E)) = \bigcap_{k \geq 1} B^k_i(E)$, and also $CB_i(E) = B_i(E_{-i} \cap \bigcap_{k \geq 1} B^k_{-i}(E)) = B_i(E_{-i} \cap CB_{-i}(E))$.

[12] As in Anscombe and Aumann (1963), consequences could be lotteries over prizes.

1997; Siniscalchi, 2011).[13] We discuss these, and the type structures they induce, in Sections 7 and 8. Other generalizations of expected utility are motivated by the Allais and Ellsberg paradoxes; Epstein and Wang (1996) show how type spaces can be constructed for a wide class of non-expected utility preferences.[14]

2.6.2 Elicitation

Our analysis puts great emphasis on players' beliefs; thus, as discussed in the Introduction, it is crucial that such beliefs can in fact be elicited from preferences. Indeed one would expect that Player 1's beliefs about 2's strategies can be elicited by asking 1 to bet on which strategy 2 will in fact play, as in Savage (1972) and Anscombe and Aumann (1963). Given this, one can then elicit 2's beliefs about 1's strategies and beliefs by having 2 bet on 1's strategies and bets, etc.[15] (Similarly, we could elicit utilities over consequences.) However, adding these bets changes the game, because the strategy space and payoffs now must include these bets. Potentially, this may change the players' beliefs about the opponents' behavior in the original game. Hence, some delicacy is required in adding such bets to elicit beliefs.[16]

2.6.3 Introspective beliefs

Our definition of a type structure assumes that each type t_i has beliefs over $S_{-i} \times T_{-i}$. This has two implications. First, players do not hold introspective beliefs, as we noted in the Introduction. Second, by specifying a type space, the analyst restricts the hierarchies the player may hold, but does not restrict play. An alternative popular model (Aumann, 1999a,b) associates with each type t_i a belief on opponents' types *and* a strategy, $\sigma_i(t_i)$. Such a model restricts the strategies that a player with a given hierarchy may choose; moreover, such restrictions are common belief among the players. We can incorporate such assumptions as well, but, in keeping with our view of the goals of the epistemic literature, we make them explicit: see for example Sec. 4.5.

2.6.4 Semantic/syntactic models

Finally we note that our modeling approach is what is called *semantic*: it starts from a type structure, and defines the belief operator, B_i, using the elements of the type structure; its properties, such as conjunction and monotonicity, follow from the way it is defined. An alternative approach, called *syntactic*, is to start with a formal language in which a belief operator is taken as a primitive; properties such as the analogs of conjunction and monotonicity are then explicitly imposed as axioms. There is a rich literature on the relation between the semantic and syntactic approaches; see for example Fagin, Halpern, Moses, and Vardi (1995), Aumann (1999a,b) and Heifetz and Mongin (2001). Due to its familiarity to economists we adopt the semantic approach here.

---

[13] See also Asheim and Perea (2005). Morris (1997) considers alternative, preference-based definitions of belief.

[14] See also Ahn (2007), Di Tillio (2008) and Chen (2010).

[15] See Morris (2002) and Dekel, Fudenberg, and Morris (2006). More generally, one can in principle elicit Player 1's preferences over acts mapping 2's strategies to consequences, then elicit 2's preferences over acts mapping 1's (strategies and) preferences to consequences, etc.; this underlies the aforementioned construction of Epstein and Wang (1996).

[16] Aumann and Dreze (2009) raise this concern, and propose a partial resolution, though they do not elicit unique beliefs, and only study first-order beliefs. (A related concern was raised by Mariotti (1995) and addressed by Battigalli (1996b).) Aumann and Dreze (2009) also note that, by assuming common belief in rationality—as we will through most of this paper—beliefs can also be elicited by adding to the game bets with payoffs that are suitably small. Aumann and Dreze (2004) Sec. 6.5 also note that elicitation of preferences may suffer from an additional problem: to elicit the ranking of two acts by direct comparison requires restricting the choice set, and hence, again, changing the game. Siniscalchi (2011) adds bets differently, avoiding all these concerns.

3 Strategic games of complete information

In this section we study common belief in rationality, as this is a natural starting point. Like the assumptions of perfect competition or rational expectations, common belief in rationality is not meant to be descriptively accurate. However, like those notions, it is a useful benchmark. We present the equivalence of the joint assumptions of rationality and common belief in rationality with iterated deletion of dominated strategies, i.e. (correlated) rationalizability, and best-reply sets. We also discuss a refinement of rationalizability that allows for additional restrictions on beliefs.

3.1 Rationality and Common Belief in Rationality

As noted, we focus on the joint assumptions of rationality and common belief in rationality. There is more than one way of stating this assumption. The immediate definition is

$$RCBR = R \cap B(R) \cap B^2(R) \cap \ldots \cap B^m(R) \cap \ldots = R \cap CB(R).\text{[17]} \tag{7}$$

In words, RCBR is the event that everybody is rational, everybody believes that everone else is rational, everybody believes that everyone else believes that others are rational, etc.. However there is an alternative definition. For all $i \in I$ let:

$$R_i^1 = R_i; \tag{8}$$

and for any $m \geq 1$,

$$R_i^{m+1} = R_i^m \cap B_i(R_{-i}^m). \tag{9}$$

Finally, we let

$$RCBR_i = \bigcap_{m \geq 1} R_i^m \quad \text{and} \quad RCBR = \prod_{i \in I} RCBR_i. \tag{10}$$

To see how 7 and 10 relate consider the case $m = 3$. We have

$$R_1^3 = R_1 \cap B_1(R_2) \cap B_1(R_2 \cap B_2(R_1))$$

whereas

$$R_1 \cap B_1(R_2) \cap B_1^2(R_1) = R_1 \cap B_1(R_2) \cap B_1(B_2(R_1)).$$

Inspecting the last term, the definition of R_1^3 is seemingly more demanding. However, thanks to monotonicity and conjunction (see Eq. 5), the two are equivalent. Inductively, it is easy to see that the two definitions of $RCBR_i$ in Eqs. (7) and (10) are also equivalent. However, when we consider non-monotonic belief operators—as we will have to for studying refinements—this equivalence will fail.

Having defined the epistemic assumptions of interest in this section, we now turn to the relevant solution concepts. In general different (but obviously related) solution concepts characterize the behavioral implications of epistemic assumptions such as RCBR in complete type structures, where nothing is assumed beyond RCBR, and smaller type structures, in which players' beliefs satisfy additional (commonly believed) assumptions. Here, the relevant concepts are rationalizability and best-reply sets.

[17] A typographical note: we write "RCBR" in the text as the acronym for "rationality and common belief in rationality," and "$RCBR$" in equations to denote the event that corresponds to it.

Definition 8 (Rationalizability) *Fix a game $(I, (S_i, u_i)_{i \in I})$. Let $S_i^0 = S_i$ for all $i \in I$. Inductively, for $m \geq 0$, let S_i^{m+1} be the set of strategies that are best replies to conjectures $\sigma_{-i} \in \Delta(S_{-i}^m)$. The set $S_i^\infty = \bigcap_{m \geq 0} S_i^m$ is the set of* ***(correlated) rationalizable*** *strategies of Player i.*

Bernheim (1984) and Pearce (1984) propose the solution concept of rationalizability, which selects strategies that are best replies to beliefs over strategies that are themselves best replies, and so on. Intuitively, one expects this to coincide with the iterative deletion procedure in Def. 8. Indeed these authors prove this, except that they focus on beliefs that are product measures, i.e. stochastically independent across different opponents' strategies.

A strategy $s_i \in S_i$ is **(strictly) dominated** if there exists a distribution $\sigma_i \in \Delta(S_i)$ such that, for all $s_{-i} \in S_{-i}$, $u_i(\sigma_i, s_{-i}) > u_i(s_i, s_{-i})$. It is well-known (Pearce, 1984; Gale and Sherman, 1950; Van Damme, 1983) that a strategy is strictly dominated if and only if it is not a best reply to any belief about the opponents' play.[18] Therefore, S_i^∞ is also the set of strategies of i that survives **iterated strict dominance**, i.e., the solution concept that selects the iteratively undominated strategies for each player. In the game of Fig. 1, it is easy to verify that $S^1 = \{T, M\} \times \{L, C, R\}$ and $S^2 = S^\infty = \{T, M\} \times \{L, C\}$.

A best-reply set is a collection of strategy profiles with the property that every strategy of every player is rationalized by (i.e., is a best response to) a belief restricted to opponents' strategy profiles in the set. A best-reply set is full if, in addition, *all* best replies to each such rationalizing belief also belong to the set.[19]

Definition 9 *Fix a game $(I, (S_i, u_i)_{i \in I})$. A set $B = \prod_{i \in I} B_i \subset S$ is a* ***best-reply set*** *(or BRS) if, for every player $i \in I$, every $s_i \in B_i$ is a best reply to a belief $\sigma_{-i} \in \Delta(B_{-i})$.*

B is a ***full BRS*** *if, for every $s_i \in B_i$, there is a belief $\sigma_{-i} \in \Delta(B_{-i})$ that rationalizes s_i and such that all best replies to σ_{-i} are also in B_i.*

Notice that the player-by-player union of (full) BRSs is again a (full) BRS.[20] Thus, there exists a unique, maximal BRS, which is itself a full BRS, and is equal to S^∞.

To clarify the notion of full BRS, refer to the game in Fig. 1. The profile (T, C) is a BRS, but not a full BRS, because, if player 1 plays T, then L yields the same payoff to player 2 as C. On the other hand, $\{T\} \times \{L, C\}$ is a full BRS, because T is the unique best reply for player 1 to a belief that assigns equal probability to L and C, and L and C are the only best replies to a belief concentrated on T.

We can now state the epistemic characterization result.

Theorem 1 (Brandenburger and Dekel (1987b), Tan and da Costa Werlang (1988)) [21] *Fix a game $G = (I, (S_i, u_i)_{i \in I})$.*

1. *In any type structure $(I, (S_{-i}, T_i, \beta_i)_{i \in I})$ for G, $\mathrm{proj}_S RCBR$ is a full BRS.*
2. *In any complete type structure $(I, (S_{-i}, T_i, \beta_i)_{i \in I})$ for G, $\mathrm{proj}_S RCBR = S^\infty$.*
3. *For every full BRS B, there exists a finite type structure $(I, (S_{-i}, T_i, \beta_i)_{i \in I})$ for G such that $\mathrm{proj}_S RCBR = B$.*

[18] This equivalence holds for games with compact strategy sets and continuous payoff functions (in particular, the finite games we consider here). See also Dufwenberg and Stegeman (2002) and Chen, Long, and Luo (2007).

[19] For related notions, see Basu and Weibull (1991).

[20] That is: if $B = \prod_i B_i$ and $C = \prod_i C_i$ are (full) BRSs, then so is $\prod_i (B_i \cup C_i)$.

[21] See also Armbruster and Böge (1979) and Böge and Eisele (1979).

We do not provide proofs in this chapter; they can be found in the cited papers, or can be adapted from arguments therein. For some results, we provide the details in an online appendix, Dekel, Pomatto, and Siniscalchi (2013b). We discuss this result in Section 3.2. For an example of Theorem 1, consider the type structure of Fig. 1. Then $\text{proj}_S RCBR = \{T, M\} \times \{L, C\}$, which is a full BRS and indeed equals S^∞. Next, consider the smaller type structure $\mathcal{T}'$ containing only type t_1^1 for player 1 and type t_2^1 for player 2. Now $\text{proj}_S RCBR = \{T\} \times \{L, C\}$, which, as noted above, is indeed a full BRS.

3.2 Discussion

Theorem 1 characterizes the implications of RCBR. One could also study the weaker assumption of common belief in rationality (CBR). The latter is strictly weaker because belief in an event does not imply it is true. Hence, CBR only has implications for players' beliefs; we focus on RCBR because it also restricts behavior. Epistemic models that allow for introspective beliefs have the feature that, if a player has correct beliefs about her own strategy and beliefs, then, if she believes that she is rational she is indeed rational. Hence, in such models, CBR is equivalent to RCBR.

The interpretation of (1) in Theorem 1 is that, if the analyst assumes that RCBR holds, but allows for the possibility that the players' beliefs may be further restricted (i.e., something in addition to rationality is commonly believed), then the analyst can only predict that play will be consistent with *some* full BRS.[22] This implies that, unless the analyst knows what further restrictions on players' beliefs hold, he must allow for the player-by-player union of all full BRSs. As we noted, this is equal to S^∞.

Part (2) in Theorem 1 is an epistemic counterpart to this. A complete type structure embeds all other type structures; it is therefore "natural" to expect that the predictions of RCBR in a complete structure should also be S^∞. Theorem 1 shows that this is the case. This convenient equivalence fails when we consider refinements.

Part (3) confirms that the result in (1) is tight: *every* full BRS represents the behavioral implications of RCBR in some type structure. If this was not the case, then RCBR would have more restrictive behavioral implications than are captured by the notion of full BRS. Furthermore, the result in (3) indicates a sense in which RCBR is "necessary" for behavior to be consistent with a full BRS. While, as noted in the Introduction, players may choose strategies in a given full BRS B by accident, or following thought processes altogether different from the logic of RCBR, the latter is always a *possible* reason why individuals may play strategy profiles in B.

3.3 Δ-rationalizability

As we discussed, a type structure encodes assumptions about players' hierarchies of beliefs. This may provide a convenient way to incorporate specific assumptions of interest. For example, one may wish to study the assumption that players' beliefs over opponents' play are independent, or that for some reason a particular strategy – even if it is rationalizable – will not be played, and so on. An alternative approach (Battigalli and Siniscalchi, 2003) is to make them explicit. In this subsection we outline one way to do so.

For every player $i \in I$, fix a subset $\Delta_i \subset \Delta(S_{-i})$. Given a type structure, the event that Player i's beliefs lie in the set Δ_i is

$$[\Delta_i] = \{(s_i, t_i) : f_i(t_i) \in \Delta_i\}.$$

We wish to characterize RCBR combined with common belief in the restrictions Δ_i.[23]

[22] It must be a full BRS because we do not restrict play: see the discussion at the end of Sec. 2.

[23] For related solution concepts (albeit without a full epistemic characterization) see Rabin (1994) and Gul (1996).

Definition 10 *Fix a game $(I, (S_i, u_i)_{i\in I})$ and a collection of restrictions $\Delta = (\Delta_i)_{i\in I}$. A set $B = \prod_{i\in I} B_i \subset S$ is a **Δ-best-reply set** (or Δ-BRS) if, for every player $i \in I$, every $s_i \in B_i$ is a best reply to a belief $\sigma_{-i} \in \Delta(B_{-i}) \cap \Delta_i$; it is a **full Δ-BRS** if, for every $s_i \in B_i$, there is a belief $\sigma_{-i} \in \Delta(B_{-i}) \cap \Delta_i$ that rationalizes s_i and such that all best replies to σ_{-i} are also in B_i.*

Definition 11 *Fix a game $(I, (S_i, u_i)_{i\in I})$. For each $i \in I$, let $S_i^{\Delta,0} = S_i$. Inductively, for $m \geq 0$, let $S_i^{\Delta,m+1}$ be the set of strategies that are best replies to conjectures $\sigma_{-i} \in \Delta_i$ such that $\sigma_{-i}(S_{-i}^{\Delta,m}) = 1$. The set $S_i^{\Delta,\infty} = \bigcap_{m\geq 0} S_i^{\Delta,m}$ is the set of **Δ-rationalizable** strategies of Player i.*

Obviously the set of Δ-rationalizable strategies may be empty for certain restrictions Δ. Also, note that $S_i^{\Delta,\infty}$ is a full Δ-BRS.

Theorem 2 *Fix a game $G = (I, (S_i, \beta_i)_{i\in I}$ and a collection of restrictions $\Delta = (\Delta_i)_{i\in I}$.*

1. *In any type structure $(I, (S_{-i}, T_i, \beta_i)_{i\in I})$ for G, $\mathrm{proj}_S\,(RCBR \cap CB([\Delta]))$ is a full Δ-BRS.*
2. *In any complete type structure $(I, (S_{-i}, T_i, \beta_i)_{i\in I})$ for G, $\mathrm{proj}_S\,(RCBR \cap CB([\Delta])) = S^{\Delta,\infty}$.*
3. *For every full Δ-BRS B, there exists a finite type structure $(I, (S_{-i}, T_i, u_i)_{i\in I})$ for G such that $\mathrm{proj}_S\,(RCBR \cap CB([\Delta])) = B$.*[24]

Results (1)–(3) in Theorem 2 correspond to results (1)–(3) in Theorem 1.

The notion of Δ-rationalizability extends easily to games with incomplete information, and is especially useful in that context. We provide an example in Section 6.5; more applied examples can be found, e.g., in Battigalli and Siniscalchi (2003). In the context of complete-information games, Bernheim's and Pearce's original definition of rationalizability required that beliefs over opponents' strategies be independent. This can also be formulated using Δ-rationalizability: for every player i, let Δ_i be the set of product measures over S_{-i}. Restrictions on first-order beliefs may also arise in a learning setting, where players observe only certain aspects of play in each stage. We return to this point in Section 7.6, where we discuss self-confirming equilibrium in extensive games.

4 Equilibrium Concepts

4.1 Introduction

A natural question is what epistemic characterizations can be provided for equilibrium concepts. By this we mean solution concepts where players best-reply to opponents' *actual* strategies. This is in contrast with solution concepts like rationalizability, where players best-reply to conjectures about opponents' strategies that may be incorrect.

Before turning to Nash equilibrium, we consider two weaker solution concepts, objective and subjective correlated equilibrium. Somewhat surprisingly, it turns out that the latter equilibrium concept is equivalent to correlated rationalizability. Hence, RCBR does provide a characterization of an equilibrium concept as well. Subsection 4.2 develops this point. The main idea is that any incorrect beliefs about opponents' strategies can be "shifted" to incorrect beliefs about a correlating

[24] In fact, one can define a type structure in which the restrictions Δ hold for *every* type. In such a structure, $\mathrm{proj}_S\,(RCBR \cap CB([\Delta])) = \mathrm{proj}_S RCBR$. It is also possible to construct a type structure that is not complete, but is infinite and contains all belief hierarchies that are consistent with it being common belief that players' first-order beliefs satisfy the restrictions Δ. Unlike the type structures constructed to obtain part (3), this contains no other assumptions on beliefs beyond common belief in Δ. Consequently, RCBR characterizes $S^{\Delta,\infty}$ in this structure.

device, thereby maintaining the assumption that players have correct beliefs about the mapping from correlating signals to strategies.

Subsection 4.3 and 4.4 characterize objective correlated and Nash equilibrium. In contrast to other results in this paper, in which epistemic conditions fully characterize *play*, in Subsections 4.3 and 4.4 epistemic conditions only imply that *beliefs* correspond to equilibrium. For example, we do not show that under certain conditions players play Nash-equilibrium strategies, only that the profile of their first-order beliefs is an equilibrium profile. Indeed this is one of the insights that emerges from the epistemic analysis: Nash equilibrium is usefully interpreted as a property of (first-order) beliefs, not play. This point was made by Harsanyi (1973) and Aumann (1987), among others.

A critical assumption in Subsections 4.3 and 4.4 is the existence of a "common prior" that generates beliefs. While we mostly follow Aumann (1987) and Aumann and Brandenburger (1995), in contrast to their approach we do not allow players to have beliefs over their own strategies.[25] Due to this difference, a direct adaptation of the common prior assumption to our setting turns out to be weaker than in those papers (indeed betting becomes possible). Hence, we formulate an additional assumption that is needed for a full characterization of these equilibrium concepts. The final subsection presents an alternative sufficient – but not necessary – assumption to obtain these concepts.

4.2 Subjective Correlated Equilibrium

Aumann (1987) defined (subjective) correlated equilibria; these are equivalent to Nash equilibria of a game in which players observe signals from a correlating device prior to choosing their actions. A correlating device consists of a finite set Ω of realizations, and, for each player, a partition Π_i of this finite set, and a conditional probability distribution $\mu_i(\cdot|\pi_i)$ for each cell π_i in the partition.[26] In a correlated equilibrium of a strategic-form game $G = (I, (S_i, u_i)_{i\in I})$, players choose strategies in S_i as a function of their signal $\pi_i \in \Pi_i$, so as to maximize their conditional expected payoff, taking as given the equilibrium behavior of their opponents.

Definition 12 *Fix a game $G = (I, (S_i, u_i)_{i\in I})$.*

*A **correlating device** for the game G is a tuple $\mathcal{C} = (\Omega, (\Pi_i, \mu_i)_{i\in I})$, where Ω is a finite set, for every $i \in I$, Π_i is a partition of Ω with typical element π_i, and μ_i is a conditional belief, i.e., $\mu_i : 2^\Omega \times \Pi_i \to [0, 1]$ satisfies $\mu_i(\cdot|\pi_i) \in \Delta(\Omega)$ and $\mu_i(\pi_i|\pi_i) = 1$ for all $\pi_i \in \Pi_i$. If there exists $\mu \in \Delta(\Omega)$ such that, for every $i \in I$, and $\pi_i \in \Pi_i$, $\mu_i(\cdot|\pi_i) = \mu(\cdot|\pi_i)$, then it is an **objective correlating device**.*

*A **subjective correlated equilibrium** is a correlating device and a tuple $(\mathbf{s}_i)_{i\in I}$ where, for each $i \in I$, $\mathbf{s}_i : \Omega \to S_i$ is measurable with respect to Π_i and, for every $\pi_i \in \Pi_i$,*

$$\sum_{\omega\in\pi_i} \mu_i(\{\omega\}|\pi_i)u_i(\mathbf{s}_i(\omega), \mathbf{s}_{-i}(\omega)) \geq \sum_{\omega\in\pi_i} \mu_i(\{\omega\}|\pi_i)u_i(s_i, \mathbf{s}_{-i}(\omega)) \qquad \forall s_i \in S_i. \tag{11}$$

*An **objective correlated equilibrium** is a subjective correlated equilibrium where the correlating device is objective.*

*Given an objective correlated equilibrium $\mathcal{C} = (\Omega, \Pi_i, \mu)$, the **objective correlated equilibrium distribution** induced by $\mathcal{C}$ is the probability distribution $\sigma \in \Delta(S)$ defined by $\sigma(s) = \mu(\{\omega : \mathbf{s}(\omega) = s\})$ for all $s \in S$.*

[25] For other approaches, see Tan and da Costa Werlang (1988), Brandenburger and Dekel (1987b) and Perea (2007).

[26] Aumann (1987) defines correlating devices slightly differently; for details and to see how this affects the results herein, see his paper. (See also Brandenburger and Dekel, 1987a)

There are obvious formal similarities between type structures and correlating devices—although their interpretation is very different. As noted repeatedly, a type structure is merely a mathematical construct used to represent belief hierarchies. On the other hand, a correlating device is meant to represent a real signal structure—something the players and (potentially) the analyst observe. The formal similarities yield the following result.

Theorem 3 (Brandenburger and Dekel (1987b)) *Fix a game $G = (I, (S_i, u_i)_{i \in I})$.*

1. *For any type structure $(I, (S_{-i}, T_i, \beta_i)_{i \in I})$ for G, there exists a subjective correlated equilibrium $\big((\Omega, \Pi_i), \mathbf{s}\big)$ of G such that, for all $i \in I$, $\mathrm{proj}_{S_i} RCBR_i = \mathbf{s}_i(\Omega)$.*
2. *Given any subjective correlated equilibrium $\big((\Omega, \Pi_i), \mathbf{s}\big)$ of G there exists a type structure $(I, (S_{-i}, T_i, \beta_i)_{i \in I})$ for G such that $\mathrm{proj}_{S_i} RCBR_i \supseteq \mathbf{s}_i(\Omega)$ for all $i \in I$.*

The essence of this theorem is that, under RCBR, the strategic uncertainty in an epistemic type structure (or, equivalently, in rationalizability) is interchangeable with the exogenous uncertainty of a (subjective) correlating device in a correlated equilibrium. The proof of part 1 sets $\Omega = RCBR \subseteq S \times T$, defines the cells in each player's partition Π_i to be of the form $\{(s_i, t_i)\} \times RCBR_{-i}$, and chooses the belief μ_i given the cell $\{(s_i, t_i)\} \times RCBR_{-i}$ so that its marginal on $RCBR_{-i}$ equals $\beta_i(t_i)$. For part 2, let $T_i = \Pi_i$ and, for any $\pi \in \Pi_i$, let $\beta_i(\pi)(s_{-i}, \pi_{-i}) = \mu_i(\{\omega : \forall j, \mathbf{s}_j(\omega) = s_j \text{ and } \Pi_j(\omega) = \pi_j\} | \pi_i)$, where $\Pi_j(\omega)$ denotes the element of Π_j that contains ω. Note that, in part 2, there may be strategies in $RCBR_i$ that are not played in the correlated equilibrium, but the set of (interim) payoffs under RCBR and in the equilibrium are the same.

4.3 Objective correlated equilibrium

To characterize objective correlated equilibrium and Nash equilibrium, we want to define the event that Player i's beliefs are "consistent with a common prior." By this we mean that her belief hierarchy can be generated in some type structure where the beliefs held by each type t_i can be obtained from some probability measure μ (the common prior) over the profiles of strategies and types $S \times T$, by conditioning on the event that i's type is indeed t_i. Note that we state the common-prior assumption as a property of belief hierarchies, rather than type structures; in this, we deviate from the received literature, but are consistent with our premise that the primitives of our analysis should be elicitable.

To make this formal, we proceed in two steps. First, a type structure $\mathcal{T} = (I, (S_{-i}, T_i, \beta_i)_{i \in I})$ admits a common prior μ on $S \times T$ if the belief maps β_i are obtained from μ by conditioning on types. This differs from the standard definition (e.g., Aumann, 1987) because it conditions only on types, not on strategies; we discuss this important point after Example 3. However, it is as "close" as possible to the standard definition, given our premise that players do not hold beliefs about their own strategies.

Definition 13 *A finite type structure $\mathcal{T} = (I, (S_{-i}, T_i, \beta_i)_{i \in I})$* ***admits*** *(or* ***is generated by****)* ***a common prior*** *$\mu \in \Delta(S \times T)$ if, for all $t_i \in T_i$, $\mu(S_i \times \{t_i\} \times S_{-i} \times T_{-i}) > 0$ and $\beta_i(t_i) = \mathrm{marg}_{S_{-i} \times T_{-i}} \mu(\cdot | S_i \times \{t_i\} \times S_{-i} \times T_{-i})$.*

We now translate Definition 13 into one that is stated in terms of hierarchies, rather than types. Given any type structure $\mathcal{T}$, we deem type $t_i \in T_i$ consistent with a common prior μ if its induced hierarchy is the same as the one which would arise in the ancillary type structure $\mathcal{T}^\mu$ which admits

μ as a common prior, where the type spaces of $\mathcal{T}^\mu$ are subsets of those in $\mathcal{T}$.[27] The following definition makes this precise.

Definition 14 *Fix a finite type structure $\mathcal{T} = (I, (S_{-i}, T_i, \beta_i)_{i\in I})$, a player $i \in I$, and a probability $\mu \in \Delta(S \times T)$. Consider the type structure $\mathcal{T}^\mu = (I, (S_{-i}, T_i^\mu, \beta_i^\mu)_{i\in I})$ that admits μ as a common prior and such that, for every $i \in I$, $T_i^\mu \subseteq T_i$. The event "Player i's beliefs are consistent with a **common prior** μ" is*

$$CP_i(\mu) = \Big\{(s_i, t_i) \,:\, \mu(S \times \{t_i\} \times T_{-i}) > 0 \text{ and } \varphi_i(\mathcal{T})(t_i) = \varphi_i(\mathcal{T}^\mu)(t_i)\Big\}. \tag{12}$$

*The prior μ is **minimal** for t_i if t_i is in $CP_i(\mu)$ and, for all $\nu \in \Delta(S \times T)$ with t_i in $CP_i(\nu)$, $\text{supp}\nu \not\subset \text{supp}\mu$.*

The following examples illustrate Definition 14.

Example 2 Definition 14 is stated in terms of hierarchies, which are elicitable. This example shows a further benefit of this formulation. Let $T_1 = \{t_1\}, T_2 = \{t_2\}, T_3 = \{t_3', t_3''\}$, $S_i = \{s_i\}$ for all i, $\beta_1(t_1)(t_2, t_3', s_2, s_3) = 1$, $\beta_2(t_2)(t_1, t_3'', s_1, s_3) = 1$, $\beta_3(t_3)(t_1, t_2, s_1, s_2) = 1$ for $t_3 \in T_3$. Since all players have a single strategy, all types commonly believe the profile (s_1, s_2, s_3), so the hierarchies of beliefs over strategies *should* be deemed consistent with a common prior; indeed, this is the case according to our Definition 14. Yet this type space does not have a common prior in the standard sense of Definition 13 because of the redundancy. Specifically, t_3' and t_3'' induce the same hierarchies, even though they are distinct types: Player 1 is sure 3's type is t_3' while 2 is sure 3's type is t_3''. Note that, as an alternative to the definition above, another way around this difficulty is to rule out redundant type spaces. ■

Example 3 We illustrate two aspects of Definition 14: first, the role of minimality, and second, why we allow the type spaces T_i^μ in the ancillary structure $\mathcal{T}^\mu$ to be a strict subset of the type spaces T_i in the original structure.

Let $\mathcal{T}$ be the type structure with $T_i = \{t_i^a, t_i^b\}, S_i = \{s_i^a, s_i^b\}$, $\beta_i(t_i^k)(s_{-i}^k, t_{-i}^k) = 1$. This type structure is really the combination of two separate structures, $\mathcal{T}^a$ and $\mathcal{T}^b$: in each structure $\mathcal{T}^k$, for $k = a, b$, the type spaces are $T_i^k = \{t_i^k\}$ for $i = 1, 2$ and the profile s^k is commonly believed. The structure $\mathcal{T}$ is consistent with any common prior μ that assigns probability $\mu_k > 0$ to $(s_i^k, t_i^k)_{i\in I}$ with $\mu_a + \mu_b = 1$. However, focusing on minimal common priors treats the two components $\mathcal{T}^a$ and $\mathcal{T}^b$ distinctly. In particular, the minimal common prior for both types t_i^a assigns probability one to $(s_i^a, t_i^a)_{i\in I}$; it generates the beliefs in the ancillary type structure $\mathcal{T}^a$.

Treating $\mathcal{T}^a$ as distinct from $\mathcal{T}^b$ is important to characterize correlated equilibrium. Assume that s_i^a is strictly dominant for $i = 1, 2$. Consider type structure $\mathcal{T}$. Then at $(s_i^a, t_i^a)_{i=1,2}$ RCBR holds and by construction beliefs are consistent with the common prior μ. However, $\text{marg}_S\mu$ assigns positive probability to the strictly dominated strategies s_i^b, $i = 1, 2$, and hence it is not a correlated equilibrium distribution. On the other hand, $\text{marg}_S\mu^a$ is a correlated equilibrium distribution. ■

Aumann proved the important result that a common prior together with common belief in rationality implies that the distribution over actions is an objective correlated equilibrium distribution. Aumann's framework is different from ours: in his model, player i's "type" incorporates i's strategy, and hence corresponds to a pair (s_i, t_i) in our framework. As a result, the existence

[27] To clarify, to obtain $\mathcal{T}^\mu$ from $\mathcal{T}$ we may eliminate some types, and replace the belief maps β_i with maps β_i^μ derived from μ by conditioning.

of a common prior in Aumann's framework requires that t_i's beliefs be obtained by conditioning on (s_i, t_i), rather than just t_i as in Def. 13. This implies that a common prior in the sense of Def. 13 need not be a common prior in Aumann's framework. The implications of this distinction are apparent in the following example (due to Brandenburger and Dekel, 1986).

Example 4 Consider the game in Table 1 (due to Bernheim, 1984) and the type structure $\mathcal{T}$ generated by the common prior μ in Table 2 as per Definition 13 (adapted from Brandenburger and Dekel, 1986). A fortiori, each type's belief hierarchy is consistent with the common prior μ in the sense of Def. 14; furthermore, RCBR holds for every strategy-type profile, i.e., $RCBR = S \times T$. For instance, (T, t_1) and (M, t_1) are both rational because T and M are best replies to type t_1's first-order belief that 2 plays L and C with equal probability. Yet, the distribution over strategies induced by μ is *not* an objective correlated equilibrium distribution. (The only objective correlated equilibrium places probability one on the profile (M, C).) Thus, a common prior in the sense of Definition 13 and RCBR do not characterize objective correlated equilibrium.

	L	C	R
T	$7,0$	$0,5$	$0,7$
M	$5,0$	$2,2$	$5,0$
B	$0,7$	$0,5$	$7,0$

Table 1: A two-player game

	(L,t_2)	(C,t_2)	(C,t_2')	(R,t_2')
(T,t_1)	0	0	$\frac{1}{4}$	0
(M,t_1)	$\frac{1}{4}$	0	0	0
(M,t_1')	0	0	0	$\frac{1}{4}$
(B,t_1')	0	$\frac{1}{4}$	0	0

Table 2: The common prior

	(L,t_2)	(C,t_2)	(C,t_2')	(R,t_2')
$\beta_1(t_1)$	$\frac{1}{2}$	0	$\frac{1}{2}$	0
$\beta_1(t_1')$	0	$\frac{1}{2}$	0	$\frac{1}{2}$

Table 3: Player 1's beliefs over $S_2 \times T_2$

	(T,t_1)	(M,t_1)	(M,t_1')	(B,t_1')
$\beta_2(t_2)$	0	$\frac{1}{2}$	0	$\frac{1}{2}$
$\beta_2(t_2')$	$\frac{1}{2}$	0	$\frac{1}{2}$	0

Table 4: Player 2's beliefs over $S_1 \times T_1$

Moreover, the beliefs in the example permit a form of betting between individual players and an outside observer (or dummy player) whose beliefs are given by the prior μ. Consider the bet described in Table 5, where the numbers specify the payments from the outside observer to Player 1.

	L	C	R
T	1	$-\frac{1}{2}$	0
M	$-\frac{1}{2}$	1	$-\frac{1}{2}$
B	0	$-\frac{1}{2}$	1

Table 5: A bet between Player 1 and an outside observer

	L	C	R
T	0	$\frac{1}{4}$	0
M	$\frac{1}{4}$	0	$\frac{1}{4}$
B	0	$\frac{1}{4}$	0

Table 6: The outside observer's beliefs over S

The outside observer computes the value of this bet using the prior μ from Table 2; the marginal on S is given in Table 6. The observer expects to receive $\frac{1}{2}$ with probability one. The unusual feature of this example (relative to the literature) is that Player 1 is betting on his own actions as well as those of Player 2. Suppose his type is t_1; in this case he is indifferent between T and M in terms of his payoffs in the game. Moreover, by playing either T or M, he expects to get 1 or $-\frac{1}{2}$ with equal probability from the bet. As type t_1 is indifferent between T and M, and strictly prefers both to B in the game and in the bet, the outside observer has no concern that the bet will affect type t_1's incentives. The same analysis applies to type t_1'. Therefore, the observer and both types of Player 1 expect strictly positive payoffs from the bet. ■

The preceding example may seem puzzling in light of the so-called "no-trade theorems." These results state that, in a setting in which type structures are used to model hierarchical beliefs about exogenous events, rather than strategic uncertainty, the existence of a common prior is equivalent to the absence of mutually agreeable bets.[28] Example 4 instead shows that, in an environment in which the events of interest are endogenous—each player chooses his strategy—certain bets are not ruled out by the existence of a common prior in the sense of Def. 14.[29] These bets can be ruled out if we impose a further assumption on the common prior—one that is automatically satisfied in Aumann's model, due to the way "types" are defined. This assumption, condition AI of Definition 15, states that conditioning the prior on a player's strategy does not imply more information than conditioning only on his type. Clearly, the prior in Table 2 violates this: $\mu(\cdot|(t_1, T)) \neq \mu(\cdot|(t_1, M))$. We conjecture that, in the present setting where uncertainty is strategic, a suitable definition of "no betting" that takes into account the fact that players can choose their own strategies, do not have beliefs about them, but can bet on them as well as on opponents' play, can characterize this additional assumption on the common prior.[30]

Definition 15 *A prior $\mu \in \Delta(S \times T)$ satisfies **Condition AI** if, for every $i \in I$, event $E_{-i} \subset S_{-i} \times T_{-i}$, strategies $s_i, s_i' \in S_i$ and type $t_i \in T_i$ with $\mu(\{(s_i, t_i)\} \times S_{-i} \times T_{-i}) > 0$ and $\mu(\{(s_i', t_i)\} \times S_{-i} \times T_{-i}) > 0$,*

$$\mu(E_{-i} \times S_i \times T_i | \{(s_i, t_i)\} \times S_{-i} \times T_{-i}) = \mu(E_{-i} \times S_i \times T_i | \{(s_i', t_i)\} \times S_{-i} \times T_{-i}). \tag{13}$$

Roughly speaking, Condition AI requires that the conditional probability $\mu(E_{-i} \times S_i \times T_i | \{(s_i, t_i)\} \times S_{-i} \times T_{-i})$ be independent of s_i. We discuss this condition further in Subsection 4.6.1.

We then obtain a version of Aumann's celebrated result: correlated equilibrium is equivalent to RCBR and hierarchies consistent with a common prior that is minimal and satisfies Condition AI.[31]

Theorem 4 *Fix a game $G = (I, (S_i, u_i)_{i \in I})$.*

1. *For every type structure $(I, (S_{-i}, T_i, \beta_i)_{i \in I})$, if $(s_i, t_i) \in CP_i(\mu) \cap RCBR_i$ for some $\mu \in \Delta(S \times T)$, and μ is minimal for t_i and satisfies Condition AI, then* $\text{marg}_S \mu$ *is an objective correlated equilibrium distribution.*

2. *Conversely, for every objective correlated equilibrium distribution ν of G, there is a type structure $(I, (S_{-i}, T_i, \beta_i)_{i \in I})$, a prior $\mu \in \Delta(S \times T)$ that satisfies Condition AI, and such that* $\text{marg}_S \mu = \nu$ *and, for all states $(s, t) \in \text{supp}\ \mu$, $(s, t) \in CP(\mu) \cap RCBR$.*

[28] That a common prior implies no betting follows from Aumann (1976) and the subsequent no-trade literature (Milgrom and Stokey, 1982; Rubinstein and Wolinsky, 1990). The opposite direction requires a more involved statement; see Morris (1994), Bonanno and Nehring (1999), Feinberg (2000), Samet (1998a). For different characterizations (not in terms of betting), see Samet (1998b) and Heifetz (2006).

[29] As noted previously, another difference with the received literature is that here players do not have beliefs about their own strategies, whereas no-trade theorems consider environments in which every agent's beliefs are defined over the entire state space.

[30] For a different perspective on obtaining objective correlated equilibrium from no-betting conditions, see Nau and McCardle (1990).

[31] Barelli (2009) shows that a characterization of correlated equilibrium in the Aumann (1987) setting can be obtained using a weaker common-prior assumption that only restricts first-order beliefs.

4.4 Nash Equilibrium

We turn now to Nash equilibrium. We start with two players as the epistemic assumptions required are much weaker. In particular this is the only objective equilibrium assumption for which no cross-player consistency assumptions (such as a common prior or agreement) must be imposed.

For every $i \in I$, let $\phi_i \in \Delta(S_{-i})$ be a conjecture of Player i about her opponents' play. In any type structure $(I, (S_{-i}, T_i, \beta_i)_{i\in I})$, let $[\phi_i] = \{(s_i, t_i) : f_i(t_i) = \phi_i\}$ be the event that Player i's first-order beliefs are given by ϕ_i. Then the following theorem says that if 1's and 2's first-order beliefs are (ϕ_2, ϕ_1), their first-order beliefs are mutually believed to be (ϕ_2, ϕ_1) and rationality is also mutually believed, then (ϕ_1, ϕ_2) is a Nash equilibrium. As discussed in the introduction to this Section, this result, as well as the subsequent generalizations, provide conditions under which beliefs—not play—form a Nash equilibrium.

Theorem 5 *Assume that $I = 2$. If $[\phi] \cap B(R \cap [\phi]) \neq \emptyset$, then (ϕ_2, ϕ_1) is a Nash equilibrium.*[32]

We can obtain a straightforward extension of Theorem 5 to n-player games by explicitly adding the assumptions that players' beliefs over opponents' strategies are independent, and that any two players have the same beliefs over any common opponent's strategies. In particular, define the events "i has independent first-order beliefs" and "i's opponents agree":

$$\begin{aligned}
&Ind_i = \left\{(s_i, t_i) : f_i(t_i) = \Pi_{j\neq i}\mathrm{marg}_{S_j} f_i(t_i)\right\},\\
&Agree_{-i} = \left\{(s_{-i}, t_{-i}) : \forall j, k, \ell \in I \text{ s.t. } j \neq i, k \neq i, j \neq \ell, k \neq \ell, \mathrm{marg}_{S_\ell} f_j(t_j) = \mathrm{marg}_{S_\ell} f_k(t_k)\right\}.
\end{aligned}$$

It is worth emphasizing that $Agree_{-i}$, like the common prior, is a restriction that relates different players' beliefs, in contrast to all the other assumptions throughout this paper.

Theorem 6 *If $[\phi] \cap \bigcap_{i\in I} B_i(R_{-i} \cap [\phi]_{-i} \cap Ind_{-i} \cap Agree_{-i}) \neq \emptyset$, then there exist $\sigma_j \in \Delta(S_j)$ for all j such that σ is a Nash equilibrium and $\phi_j = \prod_{k\neq j} \sigma_j$ for all j.*

Aumann and Brandenburger (1995) show that these additional conditions can be derived from arguably more primitive assumptions: the common prior and common belief in the conjectures. For the reasons discussed in the preceding section, we need to add Condition AI.[33]

[32] Aumann and Brandenburger (1995) require only mutual belief in rationality and in the conjectures. This is because, in their framework, players have beliefs about their own strategies and hierarchies, and furthermore these beliefs are correct. Thus, mutual belief in the conjectures ϕ implies that i's conjecture *is* ϕ_i. As we do not model a player's introspective beliefs (here beliefs about her own beliefs), we need to explicitly assume that the conjectures are indeed ϕ.

Alternatively, in Theorems 5, 6, 7 and 9 we could drop the event $[\phi]$ and replace "B" with "B^2." We could also state these results using only assumptions on one player's beliefs, as in Theorem 4. For example, in Theorem 5 and for $i = 1$, the assumptions would be $B_1([\phi_2]) \cap B_1(B_2(R_1)) \cap B_1(B_2([\phi_1]))$ and $B_1(B_2([\phi_1])) \cap B_1(B_2(B_1(R_2))) \cap B_1(B_2(B_1([\phi_2])))$.

Finally, Aumann and Brandenburger allow for incomplete information in the sense of Sec. 6, but assume that there is common belief in the game being played. Liu (2010) shows that this assumption can be weakened to second-order mutual belief in the game, but not to mutual belief.

[33] To see why assumption AI is necessary, consider the three-player game in Fig. 5 of Aumann and Brandenburger (1995). Let $T_i = \{t_i\}$ for $i = 1, 2, 3$, and define $\mu \in \Delta(S \times T)$ by $\mu(H, t_1, h, t_2, W, t_3) = \mu(T, t_1, t, t_2, W, t_3) = 0.4$, $\mu(H, t_1, t, t_2, W, t_3) = \mu(T, t_1, h, t_2, W, t_3) = 0.1$. Define β_1 and β_2 via μ, as in Example 4 The first-order beliefs of types t_1 and t_2 place equal probability on H, T and h, t respectively; therefore, $R_i = S_i \times T_i = S_i \times \{t_i\}$ for $i = 1, 2$. Furthermore, player 3 assigns a high probability to players 1 and 2 playing either (H, h) or (T, t), so that $R_3 = \{(W, t_3)\}$. Thus, there is common belief in rationality and the first-order beliefs, as well as a common prior in the sense of Def. 14. However, player 3 has a correlated first-order belief, so we do not get a Nash equilibrium. Furthermore, players 1 and 3 could bet on the correlation between 1's and 2's strategies, so once again the common prior does not preclude bets.

Theorem 7 *If there is a probability $\mu \in \Delta(S \times T)$ that satisfies Condition AI, and a tuple $(t_1, \ldots, t_I)$ in $CP(\mu) \cap [\phi] \cap CB([\phi]) \cap B(R)$ for which μ is minimal, then there exist $\sigma_j \in \Delta(S_j)$ for all j such that σ is a Nash equilibrium and $\phi_j = \prod_{k \neq j} \sigma_j$ for all j.*

4.5 The book-of-play assumption

We now consider a related approach to dealing with the issues pointed out in Example 4 and footnote 33. We introduce a "book of play": a commonly believed function from hierarchies into strategies (Brandenburger and Dekel, 1986). The interpretation is that, once we condition on a player's hierarchical beliefs, there is no residual uncertainty about her play. The existence of such a function reflects a (perhaps naive) determinism perspective—a player's hierarchical beliefs uniquely determine his strategy—and hence may be of interest in its own right.

It turns out that common belief in a "book of play" implies that Condition AI in Definition 15 holds. Therefore, we can obtain sufficient epistemic conditions for objective correlated and Nash equilibrium by replacing Condition AI with common belief in a "book of play." We do so in Theorems 8 and 9. The advantage relative to Theorems 4 and 7 is that common belief in a "book of play" is a more easily interpretable assumption. However, we will see in Example 5 below that, in the absence of any exogenous uncertainty, this assumption is restrictive: essentially, it rules out certain forms of randomization.[34]

Consider a game $G = (I, (S_i, u_i)_{i \in I})$, a type structure $\mathcal{T} = (I, (S_{-i}, T_i, \beta_i)_{i \in I})$, and a function

$$n_i : \varphi_i(\mathcal{T})(T_i) \to S_i; \tag{14}$$

this specifies, for each type of Player i, the strategy she is " expected" to play, where any given hierarchy is associated with a unique (pure) strategy.[35] We then define the event that "i's play adheres to the book n_i":

$$[n_i] = \{(s_i, t_i) \, : \, s_i = n_i(\varphi_i(\mathcal{T})(t_i))\} \, . \tag{15}$$

Theorem 8 *Fix a game $G = (I, (S_i, u_i)_{i \in I})$. For every type structure $(I, (S_{-i}, T_i, \beta_i)_{i \in I})$ and book of play n, if $(s_i, t_i) \in CP_i(\mu) \cap RCBR_i \cap CB([n])_i$ for some $\mu \in \Delta(S \times T)$, and μ is minimal for t_i, then $\mathrm{marg}_S \mu$ is an objective correlated equilibrium distribution.*

Theorem 9 *Fix a type structure $(I, (S_{-i}, T_i, \beta_i)_{i \in I})$. If there is a book of play n and a probability $\mu \in \Delta(S \times T)$, and it is the case that $[\phi] \cap B(R) \cap CP(\mu) \cap CB([\phi]) \cap CB([n]) \neq \emptyset$, then there exist $\sigma_j \in \Delta(S_j)$ for all j such that σ is a Nash equilibrium and $\phi_j = \prod_{k \neq j} \sigma_j$ for all j.*

Notice that we did not state a converse to the preceding theorems. In fact, as the following example due to Du (2011) shows, the converses are false.

Example 5 Consider Matching Pennies. The unique correlated and Nash equilibrium is of course $\sigma_1 = \sigma_2 =$"$\frac{1}{2}$ Heads, $\frac{1}{2}$ Tails."

First, consider the converse to Theorem 9. Fix a type t_i. If there is common belief that the conjectures are (σ_2, σ_1), then every type t_j to which t_i assigns positive probability must have the

[34] One can also explore the implications of this assumption in non-equilibrium contexts. Under RCBR, Brandenburger and Friedenberg (2008) consider weaker conditions that enable them to study the notion of 'intrinsic' correlation in games with more than 2 players, which corresponds to it being common belief that there are no *exogenous* unmodeled correlating devices. Peysakhovich (2011) shows that objective correlated equilibrium outcomes are also consistent with RCBR and intrinsic correlation. The converse is false, as Example 4 shows.

[35] Here, as is the case throughout this chapter with the exception of Theorem 10, players do not have access to randomizing devices. Rather, randomizations reflect opponents' beliefs, as discussed in Section 2.2.

same belief hierarchy, and hence, by the book-of-play assumption, must play the same strategy, either Heads *or* Tails. But then type t_i's first-order belief cannot be σ_{-i}.

Next, consider the converse to Theorem 8. Fix a type structure, a type t_1, and a common prior μ minimal for t_1 and such that its marginal on S is the equilibrium distribution. Because common belief in the book of play holds (in the eyes of t_1), we can partition the types of each player i in the support of μ into those that play Heads and those that play Tails, say T_i^H and T_i^T. Consider a type $t_1' \in T_1^H$; assuming wlog that he wants to match, (common belief in) rationality requires that t_1' assigns probability at least $\frac{1}{2}$ to Heads. Repeating the argument for all types in T_1^H implies that the common prior must assign probability at least $\frac{1}{2}$ to Heads conditional on 1's type being in T_1^H. This is equivalent to saying, that, conditional on 1 playing Heads, 2 must play Heads with probability at least $\frac{1}{2}$. But by assumption, conditional on 1 playing Heads, 2 plays heads with probability exactly $\frac{1}{2}$. This implies that all types in T_1^H have the same first-order beliefs, i.e., σ_2. Repeating the argument shows that all types of i in the support of μ have the same first-order beliefs, for $i = 1, 2$. Hence, all types of each player i have the same hierarchy of beliefs, and by common belief in the book of play, they must actually be playing the same strategy. ■

The essence of the example is that the book-of-play assumption makes it impossible, in certain games, to attribute mixed-strategies to beliefs and not to actual mixing. Thus this important perspective, highlighted by Harsanyi and Aumann – indeed one of the benefits of the epistemic perspective – is not possible under the book-of play assumption. Indeed, one way to obtain a converse to Theorems 8 and 9 is to either explicitly allow for mixing, or to add extrinsic uncertainty (so as to "purify" the mixing). Once we allow for mixed strategies to be played, first-order beliefs are over mixed strategies; that is, for player j, they are measures $\phi_j \in \Delta(\prod_{k \neq j} \Delta(S_k))$. It is then useful to have a notation for the "expected belief" over pure strategies; given a first-order belief ϕ_j, let $\mathbb{E}\phi_j \in \Delta(S_{-j})$ be defined by $\mathbb{E}\phi_j(s_{-j}) = \int_{\prod_{k \neq j} \Delta(S_k)} \bar{\sigma}_{-j}(s_{-j})\, \phi_j(d\bar{\sigma}_{-j})$ for all s_{-j}. We can then state the following converse to Theorem9: given a Nash equilibrium, there is a type structure where hierarchies of beliefs are consistent with a common prior, and there is common belief in rationality, the book of play, and first-order beliefs whose expectations are the equilibrium strategy profile.

Theorem 10 *For any Nash equilibrium $(\sigma_i)_{i \in I}$, $\sigma_i \in \Delta(S_i)$, there is a first-order belief $\phi_j \in \Delta\left(\prod_{j \neq i} \Delta(S_j)\right)$ such that $\mathbb{E}\phi_j = \prod_{k \neq j} \sigma_k$ for every j, and a type structure $(I, (\prod_{j \neq i} \Delta(S_j), T_i, \beta_i)_{i \in I})$, such that, for all i, $T_i = CP_i(\mu) \cap CB(R)_i \cap CB([n]) \cap CB([\phi])$.*

4.6 Discussion

4.6.1 Condition AI

In Theorem 4, we assume that a player's hierarchy is consistent with a common prior μ which satisfies Condition AI. This is an elicitable assumption because it is about beliefs, but it is arguably somewhat opaque. As noted above, we conjecture that common priors satisfying Condition AI may be characterized via a suitable no-betting condition. This would provide a more transparent behavioral characterization.

We emphasize that we cannot interpret Condition AI, i.e., Eq. (13), directly as a restriction on Player i's beliefs: in our environment, players do not have beliefs about their own strategies. Instead, it is a restriction on the beliefs of the other players, and perhaps those of an outside observer whose beliefs are given by μ. In Example 4, it implies in particular that, conditional on t_1, an outside observer must believe that Player 1's and 2's strategies are stochastically independent.

Hence, the name of the condition: *A*umann *I*ndependence.[36]

We observed that Condition AI is implied by the common-prior assumption in Aumann's model, due to the fact that a "type" therein comprises both a belief about the other players and a strategy.[37] Our framework instead requires that we make this independence assumption explicit, and hence helps us highlight a key epistemic condition required to characterize objective correlated and Nash equilibrium.

Finally, at the risk of delving too far into philosophy, one can also relate Condition AI to the notion of free will. If there is "free will" then players cannot learn anything from their own choice of strategy. If instead players' choices are pre-determined, then it is possible they would learn something from their own choices, unless one explicitly assumes that there is enough independence to rule out any such learning. Condition AI requires that there be no such learning, either because of "free will" or because there is sufficient independence so that there is nothing to learn.

4.6.2 Comparison with Aumann (1987)

Our characterization of objective correlated equilibrium in Theorem 4 seems more complex than Aumman's original result. Translated to our setting, his result is as follows:

> Let $\mathcal{T}$ be a type structure that admits a common prior μ (in the sense of Def. 13) which satisfies Condition AI. If supp $\mu \subset R$, then $\text{marg}_S\mu$ is an objective correlated equilibrium distribution.

While elegant, this formulation involves hypotheses that are not directly verifiable. To begin with, we cannot verify whether "the type structure admits a common prior" because, as we have noted several times, we cannot elicit the type structure that generated the players' actual belief hierarchies. This is why we were led to Definition 14 rather than 13. Once we have elicited a player's hierarchy, we *can* verify whether that hierarchy is consistent with a common prior. Our theorem 4 translates Aumann's result above into the language of hierarchies. Moreover, Aumann's hypothesis of rationality for every strategy-type pair in the support of the prior implies, but does not explicitly state, that RCBR will also hold. Our focus on hierarchies forces us to make this latter assumption explicit.

4.6.3 Nash equilibrium

As discussed, the epistemic analysis leads to the interpretation of Nash equilibria in mixed strategies as descriptions of players' conjectures, rather than their actual behavior. The definition of Nash equilibrium then becomes a mutual consistency requirement: conjectures must be correct and consistent with rationality. In games with more than two players, they must also be independent and suitably consistent across players. Theorems 5–6 may be seen as essentially formalizing the Nash consistency requirement in the language of belief hierarchies.

A separate question is whether these results provide a "justification" for equilibrium concepts. In games with more than two players, the assumptions of independence and agreement (corresponding to the events Ind_i and $Agree_{-i}$ used in Theorem 6) appear strong; understandably, the literature has sought more basic conditions. Theorems 7–10 clarify the need for Condition AI or common

[36] Aumann deserves no blame for this definition; the label only indicates that it is inspired by his analysis.

[37] By the CPA, the beliefs of an Aumann type for player i about the other players' Aumann types is derived from a common prior μ by conditioning on i's Aumann type, and hence by definition on both that type's beliefs *and* strategy. Therefore, if we condition μ on two Aumann types that feature the same beliefs about the other players, but different strategies, the two resulting measures must obviously have the same marginal on the set of other players' Aumann types. This corresponds to Eq. (13).

belief in the "book of play." Ultimately, we interpret Theorems 6–10 as negative: they highlight the demanding epistemic assumptions needed to obtain equilibrium concepts. Naturally, there is room for other types of justification for equilibrium analysis, such as learning and evolution.

5 Strategic-form refinements

In this section we provide a first introduction to the epistemic analysis of refinements of rationalizability. These are important for two reasons. First, refinements yield tighter predictions, and hence can be useful in applications. Second, the epistemic conditions that yield them turn out to be of interest. In particular, this section introduces the notions of admissibility / weak dominance and common p-belief.

Weak dominance is a strengthening of Bayesian rationality (expected-utility maximization). A strategy $s_i \in S_i$ is **weakly dominated** if there exists a distribution $\sigma_i \in \Delta(S_i)$ such that, for all $s_{-i} \in S_{-i}$, $u_i(\sigma_i, s_{-i}) \geq u_i(s_i, s_{-i})$, and the inequality is strict for at least one $s^*_{-i} \in S_{-i}$. A strategy is **admissible** if it is not weakly dominated. Analogously to strict dominance, a strategy is weakly dominated if and only if it is not a best reply to any full-support belief about the opponents' play (that is, a belief that assigns strictly positive probability to every opponents' strategy profile).

A natural first step to refine the assumption of RCBR might be to consider "admissibility and common belief in admissibility" and try to obtain an analog of Theorem 1 above. However, there is a tension between the logic of admissibility and that of common belief. Loosely speaking, the former is motivated from the perspective that anything is possible, whereas the latter does restrict what is possible.[38] To address this tension, one can relax the definition of belief.

Monderer and Samet (1989) introduce the notion of p-belief, i.e., belief with probability at least p, to game theory.[39] For $p = 1$, this is the notion of belief we have considered so far. For p close to 1, p-belief has similar behavioral implications, but it enables us to resolve the tension just discussed. It makes it possible to formulate "almost" common belief in admissibility, which *is* consistent with full-support beliefs (anything is possible).

As we just noted, one motivation for our discussion of p-belief is the observation that, while admissibility is an interesting and common strengthening of rationality (expected-payoff maximization), common belief in admissibility leads to difficulties. Common p-belief in admissibility may be viewed as one one way to approximate these epistemic conditions of interest. However, there are two additional reasons. First, as we shall see momentarily, for p sufficiently high, common p-belief in admissibility characterizes a solution concept that was originally motivated by different considerations. Thus, Theorem 11 below provides a different perspective on the derived solution concept. Second, we can employ the notion of p-belief to carry out a robustness check for our analysis of RCBR. It can be shown that, for p sufficiently close to 1, rationality and common p-belief of rationality has the same behavioral implications as RCBR (in stark contrast with admissibility and common p-belief thereof): see Hu (2007).

Definition 16 *Fix a game $G = (I, (S_i, u_i)_{i \in I})$ and a type structure $\mathcal{T} = (I, (S_{-i}, T_i, \beta_i)_{i \in I})$ for G. The event that Player i* assigns probability at least $p \in [0,1]$ to $E_{-i} \subset S_{-i} \times T_{-i}$ *is*

$$B_i^p(E_{-i}) = \{(s_i, t_i) \,:\, \beta_i(t_i)(E_{-i}) \geq p\}\,. \tag{16}$$

[38] See e.g. Samuelson (1992).

[39] This notion originates in modal logic: see e.g. Fagin and Halpern (1994), Fagin, Halpern, and Megiddo (1990), and the references therein.

The event that Player i has full-support beliefs *is*

$$FS_i = \{(s_i, t_i) \, : \, \text{supp } f_i(t_i) = S_{-i}\} \,. \tag{17}$$

We can now define "admissibility" and "mutual and common p-belief of admissibility" as follows. We proceed analogously to the definition of the event $RCBR_i$ in Equations 8 and 9. We assume full-support belief in addition to rationality, and weaken belief to p-belief. (As noted, rationality and full support are equivalent to admissibility). For every $i \in I$, let

$$\begin{aligned}
&ACBA_i^{p,0} = R_i \cap FS_i, \\
&ACBA_i^{p,k} = ACBA_i^{p,k-1} \cap B_i^p(ACBA_{-i}^{p,k-1}) \text{ and} \\
&ACBA_i^p = \bigcap_{k \geq 0} ACBA_i^{p,k}.
\end{aligned} \tag{18}$$

Here it *does* matter whether we define mutual and common p-belief as above, or by iterating the p-belief operator as in Eq. (7).[40] The reason is that p-belief does not satisfy the Conjunction property in Eq. (5). On the other hand, given any *finite* type structure, there is a $\pi \in (0, 1)$ such that, for all $p \geq \pi$, Conjunction holds, i.e., player i p-believes events A_{-i} and B_{-i} if and only if she p-believes $A_{-i} \cap B_{-i}$. A similar statement holds for Monotonicity.

To capture the behavioral implications of ACBA, consider the following adaptation of the notion of best-reply sets (Def. 9).[41]

Definition 17 *Fix $p \in [0, 1]$. A set $B = \prod_{i \in I} B_i \subset S$ is a p-**best-reply set** (or p-BRS) if, for every player $i \in I$, every $s_i \in B_i$ is a best reply to a full-support belief $\sigma_{-i} \in \Delta(S_{-i})$ with $\sigma_{-i}(B_{-i}) \geq p$; it is a **full** p-**BRS** if, for every $s_i \in B_i$, there is a rationalizing full-support belief $\sigma_{-i} \in \Delta(B_{-i})$ with $\sigma_{-i}(B_{-i}) \geq p$ such that all best replies to σ_{-i} are also in B_i.*

A basic refinement of rationalizability is to carry out one round of elimination of weakly dominated strategies, followed by the iterated deletion of strictly dominated strategies. This procedure was introduced in Dekel and Fudenberg (1990), who—following Fudenberg, Kreps, and Levine (1988)—were motivated by robustness considerations.[42] Let $S^\infty W$ denote the set of strategy profiles that survive this procedure.

For every game, there exists $\pi \in (0, 1)$ such that every p-BRS with $p \geq \pi$ is contained in $S^\infty W$. Furthermore, $S^\infty W$ is itself a p-BRS, for $p \geq \pi$. This inclusion is a consequence of the fact that, as discussed above, p-belief satisfies Conjunction and Monotonicity for p large enough.

Theorem 11 *Fix a game $G = (I, (S_i, u_i)_{i \in I})$. Then there is $\pi \in (0, 1)$ such that, for $p \geq \pi$:*[43]

1. *in any type structure $(I, (S_{-i}, T_i, \beta_i)_{i \in I})$, $\text{proj}_S ACBA^p$ is a full p-BRS contained in $S^\infty W$;*

[40] Indeed, $B_1^p(R_2 \times S_3 \times T_3) \cap B_1^p(S_2 \times T_2 \times R_3) \neq B_1^p(R_2 \times R_3)$ in general, whereas equality does hold for $p = 1$.

[41] For related notions see Tercieux (2006) and Asheim, Voorneveld, W Weibull, et al. (2009).

[42] Like Bernheim's perfect rationalizability (Bernheim, 1984), this procedure is a non-equilibrium analog to trembling-hand perfection (Selten, 1975). Borgers (1994) (see also Hu, 2007) provided a characterization using common p-belief. The main difference with perfect rationalizability is that in $S^\infty W$ it is not assumed that players agree about the the trembles of other players, and trembles are not required to be independent. For refinements of rationalizability motivated by proper equilibrium (Myerson, 1978), see Pearce (1984) section 3, Schuhmacher (1999), and Asheim (2002).

[43] If $p < \pi$, then these results are modified as follows. First, (1) holds except for the claim that the p-BRS is contained in $S^\infty W$. Regarding (2), ACBA characterizes the largest p-BRS, which can be computed using the procedure in Borgers (1994). Finally, (3) holds for all p.

2. *in any complete type structure,* $(I, (S_{-i}, T_i, \beta_i)_{i \in I})$, $\mathrm{proj}_S ACBA^p = S^\infty W$*;*

3. *for every full p-BRS B, there exists a finite type structure* $(I, (S_{-i}, T_i, \beta_i)_{i \in I})$ *such that* $\mathrm{proj}_S ACBA^p = B$.

Note that π depends upon the game G. Consequently, the epistemic conditions that deliver $S^\infty W$ depend upon the game. We view this as unappealing, because to some extent the assumptions are tailored to the game. We will shortly mention an alternative approach that avoids this issue. However, this approach requires a different notion of type structure. The advantage of Theorem 11 is that it can be stated using the machinery developed so far.

An alternative way to characterize $S^\infty W$ builds on Schuhmacher (1999). Instead of weakening belief to p-belief, we weaken rationality to "ϵ-rationality." This is easiest to implement in a model where players can explicitly randomize, as in Theorem 10.[44] Consider the mixed extension $(I, (\Delta(S_i), u_i)_{i \in I}$ of the original game $(I, (S_i, u_i)_{i \in I})$, and a type structure $\left(I, \left(\prod_{j \neq i} \Delta(S_j), T_i, \beta_i\right)_{i \in I}\right)$. For given $\epsilon > 0$, a (mixed) strategy-type pair (σ_i, t_i) is **ϵ-rational** if σ_i assigns probability at most ϵ to any pure strategy that is *not* a best reply to type t_i's first-order beliefs;[45] we thus define the event

$$R_i^\epsilon = \left\{ (\sigma_i, t_i) \ : \ u_i(s_i, \mathbb{E} f_i(t_i)) < \max_{s_i' \in S_i} u_i(s_i', \mathbb{E} f_i(t_i)) \Rightarrow \sigma_i(s_i) \leq \epsilon. \right\},$$

where, as in Theorem 10, $\mathrm{E} f_i(t_i)$ is the reduction of the measure $f_i(t_i)$, which is

As in Theorem 11, we need a full-support assumption in order to obtain admissibility. Given that we consider the mixed extension of the game, as in the discussion preceding Theorem 10, player i's first-order beliefs are now a probability measure over profiles of *mixed* strategies of the opponents. The appropriate full-support assumption remains over *pure* strategy profiles. We formalize this using "expected" first-order beliefs: e.g., for type t_i, $\mathbb{E} f_i(t_i) \in \Delta(S_{-i})$ has full support. The event where this is the case is

$$\widehat{FS}_i = \{ (\sigma_i, t_i) \ : \ [\mathbb{E} f_i(t_i)](s_{-i}) > 0 \ \forall s_{-i} \in S_{-i} \}.$$

Theorem 12 *Fix a game* $G = (I, (S_i, u_i)_{i \in I})$. *Then there is* $\bar{\epsilon} \in (0, 1)$ *such that, for* $\epsilon \leq \bar{\epsilon}$,

1. *in any complete type structure* $(I, (\prod_{j \neq i} \Delta(S_j), T_i, \beta_i)_{i \in I})$, $s \in S^\infty W$ *if and only if there is* $(\sigma_i, t_i)_{i \in I} \in R^\epsilon \cap \widehat{FS} \cap CB(R^\epsilon \cap \widehat{FS})$ *such that* $\sigma_i(s_i) > \epsilon$ *for each* i.

2. *in any type structure* $(I, (\prod_{j \neq i} \Delta(S_j), T_i, \beta_i)_{i \in I})$, *if* $(\sigma_i, t_i)_{i \in I} \in R^\epsilon \cap \widehat{FS} \cap CB(R^\epsilon \cap \widehat{FS})$ *then* $\sigma_i(s_i) > \epsilon$ *for each* i *implies* $s \in S^\infty W$.

Schuhmacher (1999) uses this approach to define a counterpart to Myerson (1978)'s notion of proper equilibrium. He strengthens ϵ-rationality to "ϵ-properness": σ_i must be completely mixed and, if a strategy s_i is worse than another strategy s_i' given player i's first-order beliefs, then $\sigma_i(s_i) \leq \epsilon \sigma_i(s_i')$, where σ_i is i's mixed strategy.[46] A strategy σ_i is then deemed "ϵ-properly

[44] This characterization (Theorem 12 below) could also be stated without the mixed-strategy extension, but as a result concerning beliefs and not play (see our discussion in the third paragraph of Section 4.1).

[45] Note that this is not the same as saying that the strategy obtains within ϵ of the maximal payoff; this is also often called "ϵ-rationality." The definition in the text is in the spirit of Selten (1975) and Myerson (1978), as well as Schuhmacher (1999).

[46] Without the requirement that σ_i be completely mixed, ϵ-properness may lose its bite: if, given i's beliefs, strategy s_i is strictly better than s_i', which in turn is strictly better than s_i'', then a mixed strategy that assigns probability one to s_i would formally be ϵ-proper, but one could not say that s_i' is "much more likely" than s_i''.

rationalizable" if there is a type structure and a type t_i such that the pair (s_i, t_i) is consistent with ϵ-properness and common belief thereof. Notice that this definition is epistemic; Schuhmacher (1999) provides an algorithmic procedure that yields some, but not all properly rationalizable strategies; Perea (2011a) provides a full algorithmic characterization. Finally, Asheim (2002) provides an epistemic definition of proper rationalizability using lexicographic probability systems.

At this point it would be natural to investigate epistemic conditions leading to iterated admissibility. These would require strengthening common p-belief. It turns out that it is possible to do so by replacing probabilistic beliefs with lexicographic probability systems. This in turn necessitates modifying the notion of a type structure. (This is the different type structure alluded to above in which an alternative version of Theorem 11 can be given: see Brandenburger (1992).) It is convenient to present this material after studying extensive-form refinements: see sections 8.

6 Incomplete Information

6.1 Introduction

A game has *incomplete information* if the payoff to one or more players is not fully determined by the strategy profile; we therefore allow for a parameter $\theta \in \Theta$ that enters players' payoff functions.[47] In this section we provide an epistemic analysis of such games, focusing mainly on RCBR.

The key issue that arises is to what extent the model is meant to be "complete," that is, to describe all possible aspects of the world that might be relevant to the agents. The alternative is to adopt a "small worlds" perspective, where we understand that many aspects are not included in our specification. This is a general issue in modeling, that is particularly relevant in this paper, and that is especially critical in this section. The particular aspect of concern is whether there might be additional uncertainty and information beyond what the model describes. Such information could enable correlations that we might otherwise exclude.

One way to deal with this is to adopt the small-worlds approach and study solution concepts that are "robust" to adding such unmodeled uncertainty explicitly. The other is to insist on the model being complete. We consider both in this section.

To clarify this issue we consider two distinct solution concepts that embody different degrees of correlation: interim independent and correlated rationalizability (denoted IIR and ICR respectively). Consider the game of incomplete information in Tab. 7.

θ_1	L	R
U	1,1	0,0
D	$\frac{1}{4}$,0	$\frac{1}{4}$,0

θ_2	L	R
U	0,0	1,1
D	$\frac{1}{4}$,0	$\frac{1}{4}$,0

Table 7: An incomplete-information game

The players' hierarchy of beliefs over $\Theta = \{\theta_1, \theta_2\}$ corresponds to it being common belief that the two parameters are equally likely. (Neither player receives any hard information.) These hierarchies can be modeled using two distinct type structure based on Θ (i.e., $X_{-i} = \Theta$), denoted $\mathcal{T}^N = \{I, (\Theta, T_i^N, \beta_i^N)_{i\in I}\}$ and $\mathcal{T}^R = \{I, (\Theta, T_i^R, \beta_i^R)_{i\in I}\}$. For both structures, player 1 has a single type: $T_1^N = \{t_1^N\}$ and $T_1^R = \{t_1^R\}$. However, $T_2^N = \{t_2^N\}$, whereas $T_2^R = \{t_2^R, \bar{t}_2^R\}$. The belief maps β_i^N and β_i^R are described in Table 8. Notice that these type structures describe beliefs

[47] To economize on notation, this formulation does not allow for hard private information (signals) the players may receive. To accommodate private information, one can let $\Theta = \Theta_0 \times \prod_{i\in I} \Theta_i$, where Θ_i is the set of signals that i may receive and Θ_0 represents residual uncertainty. For example, see Battigalli, Di Tillio, Grillo, and Penta (2010).

about Θ alone, and not also about players' strategies (that is, $X_{-i} = \Theta$). Solution concepts for incomplete-information games typically use such Θ-based type structures. When we turn to the epistemic analysis, we will need to consider type structures that model beliefs about both Θ and players' strategies (i.e., $X_{-i} = \Theta \times S_{-i}$).

	(θ_1, t_2^N)	(θ_2, t_2^N)
$\beta_1^N(t_1^N)$	$\frac{1}{2}$	$\frac{1}{2}$

	(θ_1, t_1^N)	(θ_2, t_1^N)
$\beta_2^N(t_2^N)$	$\frac{1}{2}$	$\frac{1}{2}$

	(θ_1, t_2^R)	$(\theta_1, \bar{t}_2^R)$	(θ_2, t_2^R)	$(\theta_2, \bar{t}_2^R)$
$\beta_1^R(t_1^R)$	$\frac{1}{2}$	0	0	$\frac{1}{2}$

	(θ_1, t_1^R)	(θ_2, t_1^R)
$\beta_2^R(t_2^R)$	$\frac{1}{2}$	$\frac{1}{2}$
$\beta_2^R(\bar{t}_2^R)$	$\frac{1}{2}$	$\frac{1}{2}$

Table 8: A non-redundant (top) and a redundant (bottom) type structure for the game in Fig. 7

Structure $\mathcal{T}^R$ has redundant types (cf. Definition 5), since the hierarchies of beliefs of types t_2^R and $\bar{t}_2^R$ are the same. Furthermore, there is no hard private information in this example: types are used solely to model hierarchies of beliefs.[48] In this sense, t_2^R and $\bar{t}_2^R$ are indistinguishable.

We can (iteratively) delete strategies that are dominated (i.e., non-best replies) for a given type. We now present some intuitive arguments about two different deletion procedures; these are formally defined in sections 6.2 and 6.3. First consider structure $\mathcal{T}^R$. One might argue that D is not dominated for t_1^R: it is a best reply to the belief that t_2^R plays R and $\bar{t}_2^R$ plays L. This conclusion crucially depends on the fact that type t_1^R believes that 2's type is t_2^R when $\theta = \theta_1$ and $\bar{t}_2^R$ when $\theta = \theta_2$. This induces a correlation between the strategy that t_1^R expects 2 to play and the payoff parameter θ; this correlation is essential for D to be a best reply.

For structure $\mathcal{T}^N$ the analysis is more subtle. One could argue that D is dominated for player 1's sole type t_1^N since, for any belief over S_2 *independently* combined with the belief that θ_1 and θ_2 are equally likely, D is not a best reply. This is the perspective underlying the solution concept of IIR, which we analyze in subsection 6.3. Alternatively one could argue that D is not dominated: it is a best reply to the belief that, with probability one-half, player 2's sole type t_2^N plays R and the state is θ_1, and otherwise t_2^N plays L and the state is θ_2. This corresponds to ICR (subsection 6.2). Is this latter belief "reasonable"? Certainly yes if there is unmodeled uncertainty: player 1 can believe that player 2's actions are correlated with θ through some unmodeled payoff-irrelevant signal.[49] If there is no unmodeled uncertainty however, one might want to exclude such beliefs.

To do so, we introduce an explicit independence assumption into the epistemic model. Intuitively, in the absence of unmodeled hard information, we want to rule out the possibility of "excessive" correlation between the payoff-relevant parameter θ and 2's strategy. However, what is "excessive" needs to be defined with care. It certainly seems reasonable to allow player 1 to believe that 2 plays differently depending on 2's hierarchical beliefs about Θ. However, *conditional on 2's hierarchy of beliefs over* Θ, 1's beliefs about θ and 2's strategies should be independent. Thus, by definition – since types that have the same hierarchy must be treated the same – an epistemic analysis that adopts this independence assumption will not result in different solutions for the two type structures.[50]

[48] If there was private information, we would model it explicitly: see footnote 47.

[49] Indeed, one view of ICR is that it is the same as IIR when certain types of unmodeled correlation are explicitly added. For related ideas see Liu (2009), who discusses this idea in the context of Bayesian Nash equilibrium. See also Sadzik (2011) and Bergemann and Morris (2011).

[50] This formulation of independence is due to Battigalli, Di Tillio, Grillo, and Penta (2011a). These authors emphasize that epistemic analysis should be carried out solely in terms of *expressible* assumptions about the primitives of the model. For incomplete-information games, the primitives are the payoff states θ and the strategy sets; in our

In the next two subsections we develop this formally. First we show ICR characterizes RCBR. Then we show that IIR corresponds to RCBR plus common belief of a suitable independence assumption *if the type space is non-redundant* (and IIR is a coarser solution concept in general). We then briefly discuss Δ-rationalizability (section 3.3) for incomplete-information games. In the last subsection we briefly discuss equilibrium concepts. Little has been done here in terms of using the ideas of Section 4 above under incomplete information to characterize standard equilibrium concepts, and in particular the many different versions of correlated equilibrium concepts in the literature.

6.2 Interim Correlated Rationalizability

We begin by formally defining incomplete-information games. We then define ICR and conclude this subsection by relating this solution concept to RCBR.

Definition 18 *A (finite)* ***(strategic-form) incomplete-information game*** *is a tuple* $G = (I, \Theta, (S_i, u_i)_{i\in I})$*, where* I *and* Θ *are finite and, for every* $i \in I$*,* S_i *is finite and* $u_i : S_i \times S_{-i} \times \Theta \to \mathbb{R}$*.*

This description is partial because it does not specify the players' (hierarchies of) beliefs about Θ. Thus, we will append to the game the players' hierarchies of beliefs over Θ. We model these hierarchies using a Θ-based type structure, i.e., a type structure as in Def. 3 where we set $X_{-i} = \Theta$.

As discussed, ICR is the solution concept that iteratively eliminates strategies that are not best replies to beliefs over $\Theta \times S_{-i}$, where beliefs allow for correlation.[51]

Definition 19 [52] *Consider a game* $G = (I, \Theta, (S_i, u_i)_{i\in I})$ *and a* Θ*-based type structure* $\mathcal{T}^\Theta = (\Theta, (T_i^\Theta, \beta_i^\Theta)_{i\in I})$*. For every* $t_i^\Theta \in T_i^\Theta$*, let*

- $ICR_i^0(t_i^\Theta) = S_i$*;*
- *for* $k > 0$*,* $s_i \in ICR_i^k(t_i^\Theta)$ *if there exists a map* $\sigma_{-i} : \Theta \times T_{-i}^\Theta \to \Delta(S_{-i})$ *such that, for all* $\theta \in \Theta$ *and* $t_{-i}^\Theta \in T_{-i}^\Theta$*,* $\sigma_{-i}(\theta, t_{-i}^\Theta)(ICR_{-i}^{k-1}(t_{-i}^\Theta)) = 1$ *and*

$$\forall s_i' \in S_i, \sum_{\theta, t_{-i}^\Theta} \beta_i^\Theta(t_i^\Theta)(\theta, t_{-i}^\Theta) \sum_{s_{-i}} \sigma_{-i}(\theta, t_{-i}^\Theta)(s_{-i}) u_i(s_i, s_{-i}, \theta) \geq$$
$$\sum_{\theta, t_{-i}^\Theta} \beta_i^\Theta(t_i^\Theta)(\theta, t_{-i}^\Theta) \sum_{s_{-i}} \sigma_{-i}(\theta, t_{-i}^\Theta)(s_{-i}) u_i(s_i', s_{-i}, \theta).$$

The set $ICR_i^\infty(t_i^\Theta) = \bigcap_{k\geq 0} ICR_i^k(t_i^\Theta)$ *is the set of* ***interim correlated rationalizable*** *strategies for type* t_i^Θ*.*

---

probabilistic setting, the only expressible assumptions are those about each player i's hierarchies of beliefs on $\Theta \times S_{-i}$. We agree with these authors' emphasis on expressibility. In fact, as argued in the Introduction, we take the stronger stand that assumptions should be elicitable.

[51] With $I > 2$ players, there are two forms of correlation: that between the underlying uncertainty and opponents' strategies, and (as in correlated rationalizability—Def. 8—and correlated equilibrium—Def. 12) that among opponents' strategies, even conditioning on the underlying uncertainty. One could allow one but not the other, in principle. For simplicity we allow for both.

[52] The definition of ICR we adopt differs in inessential ways from the one originally proposed by Dekel, Fudenberg, and Morris (2007). See also Liu (2011) and Tang (2011).

To understand this definition, recall that, in a Θ-based structure $(\Theta, (T_i^\Theta, \beta_i^\Theta)_{i\in I})$, player i's type t_i^Θ represents her beliefs about $\Theta \times T_{-i}^\Theta$, but not S_{-i}. ICR then assumes that beliefs over S_{-i} are determined by a function $\sigma_{-i} : \Theta \times T_{-i}^\Theta \to \Delta(S_{-i})$. Specifically, the probability that type t_i^Θ assigns to opponents playing a given profile s_{-i} equals

$$\sum_{(\theta, t_{-i}^\Theta)} \beta_i^\Theta(t_i^\Theta)(\theta, t_{-i}^\Theta)) \cdot \sigma_{-i}(\theta, t_{-i}^\Theta)(s_{-i}). \tag{19}$$

The fact that $\sigma_{-i}(\cdot) \in \Delta(S_{-i})$ depends upon both θ and t_{-i}^Θ allows for the possibility of unmodeled correlating information received by i's opponents, as discussed. For example, in the game of Table 7 augmented with Θ-based type structure $\mathcal{T}^1$, the strategy D of player 1 is a best reply for type t_1 given the belief on $\Theta \times S_{-i}$ constructed by defining $\sigma_2(\theta_1, t_2')(R) = 1$ and $\sigma_2(\theta_2, t_2'')(L) = 1$. The strategy D is also a best reply for type t_1 of player 1 in the structure $\mathcal{T}^2$ when we define $\sigma_2(\theta_1, t_2)(R) = 1$ and $\sigma_2(\theta_2, t_2)(L) = 1$.

To study the relationship between RCBR and ICR, we start with an epistemic type structure where $X_{-i} = \Theta \times S_{-i}$. It is important to keep track of the difference between this and the Θ-based type structure appended to the game of incomplete information and used in defining ICR (Definition 19). Of course, in our epistemic analysis, we will need to relate the belief hierarchies generated by the Θ-based type structure to the $\Theta \times S_{-i}$-based hierarchies in the epistemic type structure.

Thus, consider an epistemic type structure $\mathcal{T} = (I, (\Theta \times S_{-i}, T_i, \beta_i)_{i\in I})$. As in Def. 6, continue to denote the first-order belief map for player i by $f_i : T_i \to \Delta(\Theta \times S_{-i})$, defined by $f_i(t_i) = \text{marg}_{\Theta\times S_{-i}} \beta_i(t_i)$. Naturally, first-order beliefs are now over $\Theta \times S_{-i}$. Analogously to Def. 7, a strategy s_i is *rational* for type $t_i \in T_i$, written $(s_i, t_i) \in R_i$, iff

$$\forall s_i' \in S_i, \qquad \sum_{\theta\in\Theta, s_{-i}\in S_{-i}} f_i(t_i)(\theta, s_{-i}) u_i(s_i, s_{-i}, \theta) \geq \sum_{\theta\in\Theta, s_{-i}\in S_{-i}} f_i(t_i)(\theta, s_{-i}) u_i(s_i', s_{-i}, \theta).$$

The event, "Player i believes event $E_{-i} \subset \Theta \times S_{-i} \times T_{-i}$" is defined as in Def. 7 part (2): $(s_i, t_i) \in B_i(E_{-i})$ if $\beta_i(t_i)(E_{-i}) = 1$. As before, both R_i and $B_i(\cdot)$ are subsets of $S_i \times T_i$. Recall that we defined mutual belief in a product event $E = \prod_i E_i$, where $E_i \subseteq S_i \times T_i$, as $B(E) = \prod_i B_i(E_{-i})$. Since now beliefs are also about Θ, in order simplify the definition of $B(B(E))$ and so on, it is convenient to instead have $B(E)$ be a subset of $\Theta \times S \times T$, as follows. For $F = Q \times \prod_i E_i$, where $Q \subset \Theta$ and $E_i \subseteq S_i \times T_i$, let

$$B(F) = \Theta \times \prod_{i\in I} B_i(Q \times E_{-i}).$$

This way, $B(F)$ has the same product structure as F, so common belief can be defined as before by iterating $B(\cdot)$. Correspondingly, we adapt the definition of RCBR:

$$RCBR = (\Theta \times R) \cap CB(\Theta \times R).$$

As noted, in order to provide an epistemic analysis of ICR, we must discuss the relationship between the epistemic type structure and the Θ-based type structure that we append to the game of incomplete information and use to define solution concepts. We start with a type t_i^Θ in a Θ-based type structure—such as t_i^N in $\mathcal{T}^N$, or t_i^R in $\mathcal{T}^R$—which induces a belief hierarchy over Θ. Denote this hierarchy by $\varphi_i^\Theta(t_i^\Theta)$. We then ask whether a type t_i in the epistemic type structure, which induces belief hierarchies over $\Theta \times S_{-i}$, has the same "marginal" hierarchy on Θ, denoted $\varphi_{i,\Theta}(t_i)$. To illustrate this, we relate the type structure $\mathcal{T}^N$ of Table 8 to the epistemic type structure defined in Table 9 below.

	(θ_1, L, t_2)	(θ_1, L, t_2')	(θ_1, R, t_2)	(θ_1, R, t_2')	(θ_2, L, t_2)	(θ_2, L, t_2')	(θ_2, R, t_2)	(θ_2, R, t_2')
$\beta_1(t_1)$	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0	0
$\beta_1(t_1')$	$\frac{1}{2}$	0	0	0	0	$\frac{1}{2}$	0	0

	(θ_1, U, t_1)	(θ_1, U, t_1')	(θ_1, D, t_1)	(θ_1, D, t_1')	(θ_2, U, t_1)	(θ_2, U, t_1')	(θ_2, D, t_1)	(θ_2, D, t_2')
$\beta_2(t_2)$	$\frac{1}{2}$	0	0	0	$\frac{1}{2}$	0	0	0
$\beta_2(t_2')$	0	$\frac{1}{3}$	0	0	0	$\frac{2}{3}$	0	0

Table 9: An epistemic type structure for the game in Tab. 7.

In this epistemic type structure, types t_1 and t_2 both believe that θ_1 and θ_2 are equally likely; furthermore, t_1 assigns probability one to 2's type being t_2, and conversely. So, if players' belief hierarchies over $\Theta \times S_{-i} \times T_{-i}$ are described by t_1 and t_2 respectively, there is common belief that each payoff parameter θ_i is equally likely. This is the same belief hierarchy over Θ held by types t_1^N and t_2^N in the Θ-based type structure $\mathcal{T}^N$. Formally, we have $\varphi_i^\Theta(t_i^N) = \varphi_{i,\Theta}(t_i)$.

To further illustrate how to construct $\varphi_{i,\Theta}(\cdot)$, consider type t_1'. This type also believes that θ_1 and θ_2 are equally likely; however, t_1' has more complex second-order beliefs. Specifically, t_1' assigns probability $\frac{1}{2}$ to the event that the payoff parameter is θ_1 and that 2 thinks θ_1 and θ_2 are equally likely, and probability $\frac{1}{2}$ to the event that the payoff state is θ_2 and 2 thinks that the probability of θ_1 is $\frac{1}{3}$. Iterating this procedure yields the hierarchical beliefs on Θ held by t_1', i.e., $\varphi_{i,\Theta}(t_1')$. Notice that no type in the Θ-based structure $\mathcal{T}^N$ (or $\mathcal{T}^R$) generates this hierarchy over Θ.

Thus, for an epistemic type structure $\mathcal{T}$ and Θ-based type structure $\mathcal{T}^\Theta$, we can define the event that each player i's hierarchy over Θ is the one generated by some type t_i^Θ in $\mathcal{T}^\Theta$:

$$[\varphi^\Theta(t^\Theta)] = \left\{(\theta, s, t) \ : \ \forall i, \ \varphi_{i,\Theta}(t_i) = \varphi_i^\Theta(t_i^\Theta)\right\}.$$

Then, we can ask what strategies are consistent with RCBR and the assumption that hierarchical beliefs on Θ are generated by a given type t_i^Θ in the Θ-based structure $\mathcal{T}^\Theta$. The following theorem states that these are precisely the strategies in $ICR_i^\infty(t_i^\Theta)$.[53]

Theorem 13 *Fix a Θ-based type structure and a complete epistemic type structure for a game $\left(I, (S_i)_{i\in I}, \Theta, (u_i)_{i\in I}\right)$. Then, for any Θ-type profile $t^\Theta \in T^\Theta$,*

$$ICR^\infty(t^\Theta) = \mathrm{proj}_S \left(RCBR \cap [\varphi^\Theta(t^\Theta)]\right).$$

One important implication of this characterization is that the set of ICR strategies for a Θ-based type t_i^Θ depends solely upon the hierarchical beliefs on Θ that it generates. In particular, if two such types $t_i^\Theta, \hat{t}_i^\Theta$ induce the same hierarchies, i.e., if they are redundant, they share the same set of ICR strategies (see Dekel et al., 2007).

6.3 Interim Independent Rationalizability

As noted above, a key feature of ICR is the fact that it allows a player to believe that her opponents' strategic choices are correlated with the uncertainty Θ. The definition of ICR does so by introducing maps $\sigma_{-i} : \Theta \times T_{-i}^\Theta \to \Delta(S_{-i})$, which explicitly allow player i's conjecture about her opponents' play to depend upon the realization of θ, in addition to their Θ-based types t_{-i}. Correspondingly,

[53] For brevity, we only state the analog to part (2) in Theorems 1, 2, and 11. We could also define a notion of best-response set and provide analogs to parts (1) and (3) as well. In particular, in any epistemic type structure, $ICR^\infty(t^\Theta) \supseteq \mathrm{proj}_S \left(RCBR \cap [\varphi^\Theta(t^\Theta)]\right)$.

in an epistemic type structure, any correlation between the Θ and S_{-i} components of the first-order beliefs is allowed.

As discussed, if there is no unmodeled uncertainty one might want to rule out such correlations and assume that, conditioning on opponents' hierarchies, opponents' strategies should be uncorrelated with Θ. *Interim Independent Rationalizability*, or IIR, reflects such considerations. Like ICR, this procedure applies to an incomplete-information game augmented with a Θ-based type structure $(I, (\Theta, T_i^\Theta, \beta_i^\Theta)_{i\in I})$, and iteratively eliminates strategies that are not best replies for each player type. The difference is in the way beliefs about S_{-i} are constructed: IIR employs maps $\sigma_{-i} : T_{-i}^\Theta \to \Delta(S_{-i})$ that associate with each profile $t_{-i} \in T_{-i}^\Theta$ a distribution over strategy profiles. This explicitly rules out the possibility that opponents' strategies may be directly correlated with the payoff parameter θ. The probability that a type t_i^Θ attaches to strategy profile s_{-i}, given the function σ_{-i}, is then

$$\sum_{t_{-i}^\Theta} \beta_i^\Theta(t_i^\Theta)(\Theta \times \{t_{-i}^\Theta\})) \cdot \sigma_{-i}(t_{-i}^\Theta)(s_{-i}), \tag{20}$$

in contrast with Eq. (19). Any correlation between the payoff parameter θ and opponents' play must thus come from correlation between θ and opponents' types, because direct correlation between θ and s_{-i} is ruled out by the definition of the maps $\sigma_{-i}(\cdot)$. This implies that redundant types can matter. Consider the game in Tab. 7 and the type structure $\mathcal{T}^R$ defined in Tab. 8. Strategy D is a best reply for type t_1^R, given the function $\sigma_2 : T_2^R \to \Delta(S_2)$ such that $\sigma_2(t_2^R)(R) = 1 = \sigma_2(\bar{t}_2^R)(L)$; notice how the belief over $\Theta \times S_2$ derived from $\beta_1^R(t_1^R) \in \Delta(\Theta \times T_2^R)$ and σ_2 induces correlation between θ and 2's strategy via correlation between θ and 2's type. If we instead consider the type structure $\mathcal{T}^N$, it is impossible to induce correlation between θ and 2's strategy in this indirect way, because 2 has only one type. As a result, D is not a best reply given this type structure. This indicates that IIR can deliver different predictions in type structures that generate the same hierarchical beliefs about Θ.

The IIR procedure (Ely and Peski, 2006) is formally defined as follows.

Definition 20 [54] *Consider a game $G = (I, \Theta, (S_i, u_i)_{i\in I})$ and a Θ-based type structure $\mathcal{T}^\Theta = (\Theta, (T_i^\Theta, \beta_i^\Theta)_{i\in I})$. For every $t_i^\Theta \in T_i^\Theta$, let*

- $IIR_i^0(t_i^\Theta) = S_i$;
- *for $k > 0$, $s_i \in IIR_i^k(t_i^\Theta)$ if there exists a map $\sigma_{-i} : T_{-i}^\Theta \to \Delta(S_{-i})$ such that, for all $t_{-i}^\Theta \in T_{-i}^\Theta$, $\sigma_{-i}(t_{-i}^\Theta)(IIR_{-i}^{k-1}(t_{-i}^\Theta)) = 1$ and*

$$\begin{aligned} \forall s_i' \in S_i, \sum_{\theta, t_{-i}^\Theta} \beta_i^\Theta(t_i^\Theta)(\theta, t_{-i}^\Theta) \sum_{s_{-i}} \sigma_{-i}(t_{-i}^\Theta)(s_{-i}) u_i(s_i, s_{-i}, \theta) \geq \\ \sum_{\theta, t_{-i}^\Theta} \beta_i^\Theta(t_i^\Theta)(\theta, t_{-i}^\Theta) \sum_{s_{-i}} \sigma_{-i}(t_{-i}^\Theta)(s_{-i}) u_i(s_i', s_{-i}, \theta). \end{aligned}$$

*The set $IIR_i^\infty(t_i^\Theta) = \bigcap_{k\geq 0} IIR_i^k(t_i^\Theta)$ is the set of **interim independent rationalizable** strategies for type t_i^Θ.*

We now turn to the epistemic characterization of IIR. The key is to formalize the assumption that player i's beliefs about $\Theta \times S_{-i}$ are independent, conditional upon any hierarchical beliefs

[54] IIR can also be described as "rationalizability in the agent strategic form": see Battigalli et al. (2011a).

about Θ that i thinks may be held by her opponents. Thus, fix a finite Θ-based type structure $\mathcal{T}^\Theta$ and an epistemic type structure $\mathcal{T}$. We denote by $T_{i,CI}$ the set of i's types t_i whose beliefs satisfy this assumption: formally, $t_i \in T_{i,CI}$ if

$$\mathrm{marg}_{\Theta\times S_{-i}}\beta_i(t_i)\left(\cdot\middle|[\varphi^\Theta_{-i}(t^\Theta_{-i})]\right)$$

is the product of its marginals on Θ and S_{-i}, whenever the above conditional probability is well-defined, i.e., for every type t^Θ_{-i} in the Θ-based structure $\mathcal{T}^\Theta$ such that $\beta_i(t_i)([\varphi^\Theta_{-i}(t^\Theta_{-i}]) > 0$.[55] Finally, let

$$CI = \{(\theta, s, t)\ :\ \forall i,\ t_i \in T_{i,CI}\}.$$

For example, consider the epistemic type structure in Table 9 and the Θ-based type structure $\mathcal{T}^R$. Recall that both types t^R_2 and $\bar{t}^R_2$ in $\mathcal{T}^R$ generate the same hierarchy of beliefs about Θ, namely that it is common belief that θ_1 and θ_2 are equally likely. Observe that epistemic type t_2 generates precisely this hierarchical belief about Θ, whereas epistemic type t'_2 generates a different hierarchy. Therefore, the events $[\varphi^\Theta_2(t^R_2)]$ and $[\varphi^\Theta_2(\bar{t}^R_2)]$ coincide and are equal to $\Theta\times S_2\times\{t_2\}$ in the epistemic structure of Table 9. Now consider player 1's type t_1. This type assigns probability one to 2's type t_2, and hence to the event $[\varphi^\Theta_2(t^R_2)]$. Conditional on this event, t_1's beliefs over $\Theta \times S_2$ assign equal probability to (θ_2, L) and (θ_1, R), which is not an independent product. Therefore, $t_1 \notin T_{1,CI}$. On the other hand, type t'_1 assigns positive probability to both types t_2 and t'_2 of player 2. Conditional on t_2, i.e., conditional on $[\varphi^\Theta_2(t^R_2)]$, t'_2 assigns probability one to (θ_1, L), which is trivially an independent product. Conditional independence does not impose any further restriction on t'_1, because the Θ-hierarchy generated by t'_2 differs from the hierarchy of any type in the Θ-based structure $\mathcal{T}^R$. Therefore, $t'_1 \in T_{1,CI}$.

We then have the following characterization (Battigalli et al., 2010).

Theorem 14 *Fix a finite Θ-based type structure $\mathcal{T}^\Theta$ and a* complete *epistemic type structure $\mathcal{T}$ for the game $\left(I, (S_i)_{i\in I}, \Theta, (u_i)_{i\in I}\right)$. Then, for any profile of Θ-types $t^\Theta \in T^\Theta$,*

$$IIR^\infty(t^\Theta) \supset \mathrm{proj}_S\left(RCBR \cap CI \cap CB(CI) \cap [\varphi^\Theta(t^\Theta)]\right);$$

if, furthermore, the Θ-based type structure is not redundant (see Definition 5), then the above inclusion is an equality.

Note that the inclusion in Theorem 14 may be strict when the Θ-based structure is redundant, even if the epistemic type structure is complete. This contrasts with the results in Theorems 1, 2, 11 and 13, where completeness implies equality. In those four theorems, inclusion may be strict only if some rationalizing beliefs are simply not present in a given incomplete epistemic type structure.

To see that the inclusion in Theorem 14 may be strict in a redundant Θ-based structure, consider the game in Tab. 7 augmented with the type structure $\mathcal{T}^R$. As was argued before Definition 20, $IIR^\infty(t^R_1) = \{U, D\}$. Now consider a strategy-type pair $(s_1, t_1) \in RCBR_1 \cap CI_1 \cap CB_1(CI_2) \cap [\varphi^\Theta_1(t^R_1)]$ in a complete epistemic type structure for this game. Since $(s_1, t_1) \in [\varphi^\Theta_1(t^R_1)$, epistemic type t_1 must satisfy common belief that θ_1, θ_2 are equally likely. Hence, t_1 must believe that 2 also commonly believes this. Therefore, $(s_1, t_1) \in CI_1$ implies that t_1's beliefs about Θ and S_2 must be independent conditional on 2 commonly believing this. But then, D cannot be a best reply: that is, $(s_1, t_1) \in RCBR_1 \cap CI_1 \cap CB_1(CI_2 \cap [\varphi^\Theta_1(t^R_1)]$ implies $s_1 = U$. Thus the epistemic assumptions result in $\{U\}$, a strict subset of the IIR prediction of $\{U, D\}$. On the other hand, repeating the analysis in the structure $\mathcal{T}^N$ leads to an equality: the only IIR strategy for type t^N_1 is U.

[55] For completeness, $[\varphi^\Theta_{-i}(t^\Theta_{-i})] = \left\{(\theta, s_{-i}, t_{-i}) : \forall j \neq i,\ \varphi_{j,\Theta}(t_j) = \varphi^\Theta_j(t^\Theta_j)\right\}$.

To summarize, the main point is quite simple: if the type space has redundant types, then the solution concept should treat them symmetrically, since they are decision-theoretically indistinguishable.[56]

6.4 Equilibrium concepts

The characterization of Nash equilibrium in games of incomplete information requires significantly stronger assumptions relative to the complete-information case (cf. Theorem 5). There are two differences. One is that the first-order belief of a type in an epistemic type structure is a belief about opponents' strategies, $s_{-i} \in S_{-i}$, whereas an equilibrium of an incomplete-information game specifies maps from Θ-types into strategies. Therefore, while Theorem 5 obtained Nash equilibrium by assuming that first-order beliefs are mutually believed, now this needs to be modified so that the maps from Θ-hierarchies into strategies are (at least) mutually believed. The other, and more interesting, difference is that the assumption of mutual belief of these maps and of rationality needs to be strengthened to common belief. The following example illustrates this.

Example 6 Consider a 2-person game with payoff irrelevant uncertainty $\Theta = \{\theta, \theta'\}$; player 2 has only one action, L, player 1 has two actions, U, D where U is strictly dominant, and the Θ-based type structure $\mathcal{T}^\Theta$ is generated by the common prior in Table 10:

θ	t_2^Θ	$\hat{t}_2^\Theta$
t_1^Θ	$\frac{1}{2}$	$\frac{1}{4}$
$\hat{t}_1^\Theta$	0	0

θ'	t_2^Θ	$\hat{t}_2^\Theta$
t_1^Θ	0	0
$\hat{t}_1^\Theta$	0	$\frac{1}{4}$

Table 10: A Θ-based type structure

Consider the following maps from types in $\mathcal{T}^\Theta$ to strategies: $\psi_1(t_1^\Theta) = U$, $\psi_1(\hat{t}_1^\Theta) = D$, $\psi_2(t_2^\Theta) = \psi_2(\hat{t}_2^\Theta) = L$. The pair $\psi = (\psi_1, \psi_2)$ is obviously not a Bayesian Nash equilibrium. However, consider the common-prior epistemic type structure $\mathcal{T}$ obtained from $\mathcal{T}^\Theta$ and ψ being common belief: see Table 11.

θ	L, t_2	$L, \hat{t}_2$
U, t_1	$\frac{1}{2}$	$\frac{1}{4}$
$D, \hat{t}_1$	0	0

θ'	L, t_2	$L, \hat{t}_2$
U, t_1	0	0
$D, \hat{t}_1$	0	$\frac{1}{4}$

Table 11: An epistemic type structure

The Θ-based hierarchies generated by the epistemic type structure $\mathcal{T}$ coincide with those generated by $\mathcal{T}^\Theta$: for example, $\varphi_{1,\Theta}(t_1) = \varphi_1^\Theta(t_1^\Theta)$. The type profile (t_1, t_2) satisfies rationality and mutual, but not common, belief in rationality. Finally, by construction, Player 2's beliefs about 1's strategies, conditional on 1's Θ-hierarchy, are as specified by ψ_1, and this is common belief (the same is trivially true for Player 1's beliefs about 2's sole strategy). This shows that mutual belief

[56]In an intriguing and surprising result, Ely and Peski (2006) show that one can associate with each type a hierarchy of beliefs about $\Delta(\Theta)$, rather than Θ, in such a way as to distinguish between types that are redundant in the usual sense. However, we do not understand how their notion of hierarchy can be elicited. For example, the first-order belief of player i in an Ely-Pesky hierarchy is an element of $\Delta(\Delta(\Theta))$, representing i's beliefs about her own beliefs about Θ (for details, see p. 28 in their paper). As we argued in Sec. 2.6, introspective beliefs cannot easily be interpreted behaviorally. Moreover, a probability measure in $\Delta(\Delta(\Theta))$ is meant to represent how i's beliefs about Θ would change *if i were informed of j's type*. This obviously depends on the type space chosen to represent beliefs; we do not see how one could elicit these beliefs.

in rationality, even with common belief in the maps ψ_i, is not enough to obtain Bayesian Nash equilibrium.[57] ■

We now show that strengthening the assumptions of Theorem 5 as indicated above yields a characterization of Bayesian Nash equilibrium in two-player games (see Pomatto (2011) and Sadzik (2011)). We believe (but have not verified) that a similar analog to Theorem 7 holds for games with more than two players. There exist several incomplete-information versions of correlated equilibrium (Forges, 1993, 2006; Liu, 2011; Bergemann and Morris, 2011). The epistemic characterizations for these concepts may be insightful, and have not yet been developed.

Fix a non-redundant Θ-based type structure T^{Θ} on Θ, and maps $\psi_i : T^{\Theta}_{-i} \to \Delta(S_{-i})$. For every i, we interpret this map as Player i's conjecture about the behavioral strategy of her opponents.[58]

Given an epistemic type structure $\mathcal{T}$, let $[s_{-i}] = \Theta \times \{s_{-i}\} \times T_{-i}$ and $[t^{\Theta}_{-i}] = \Theta \times S_{-i} \times \{t_{-i} : \varphi_{-i,\Theta}(t_{-i}) = \varphi^{\Theta}_{-i}(t^{\Theta}_{-i})\}$. These are the events that "the opponents play s_{-i}" and "the opponents' Θ-hierarchies are as specified by t^{Θ}_{-i}." Then, the event that "each player's first-order beliefs are consistent with ψ" is

$$[\psi] = \Big\{(\theta, s, t) \,:\, \forall i,\ \forall t^{\Theta}_{-i} \in T^{\Theta}_{-i} \text{ s.t. } \beta_i(t_i)([t^{\Theta}_{-i}]) > 0,\ \beta_i(t_i)([s_{-i}] | [t^{\Theta}_{-i}]) = \psi_i(t^{\Theta}_{-i})(s_{-i})\Big\}. \tag{21}$$

Define "Θ-hierarchies are consistent with a given Harsanyi type structure $\mathcal{T}^{\Theta}$":

$$[\mathcal{T}^{\Theta}] = \left\{(\theta, s, t) \,:\, \forall i,\ \varphi_{i,\Theta}(t_i) \in \varphi^{\Theta}_i(T^{\Theta}_i).\right\}.$$

We need one more definition. Θ-based structure $\mathcal{T}^{\Theta}$ is *minimal* if, for every pair of players and types $i, j \in I$, $t^{\Theta}_i \in T^{\Theta}_i$, $t^{\Theta}_j \in T^{\Theta}_j$, there is a finite sequence $t^1_{i(1)}, \ldots, t^N_{i(N)}$ such that $i(1) = 1$, $t^1_{i(1)} = t^{\Theta}_i$, $i(N) = j$, $t^N_{i(N)} = t^{\Theta}_j$, and for all $n = 2, \ldots, N$, $\beta^{\Theta}_{i(n-1)}(t^{n-1})([t^n_{i(n)}]) > 0$. That is, loosely speaking, it is not possible to partition $\mathcal{T}^{\Theta}$ into two components such that each component is a type structure in and of itself.

Theorem 15 *Assume that there are two players. Fix an incomplete-information game G and a non-redundant, minimal Θ-based type structure $\mathcal{T}^{\Theta}$ and maps ψ_1, ψ_2 as above. If there is an epistemic type structure $\mathcal{T}$ in which $CB([\mathcal{T}^{\Theta}]) \cap CB([\psi]) \cap CB(\Theta \times R) \neq \emptyset$, then (ψ_2, ψ_1) is a Bayesian Nash equilibrium of the Bayesian game $(G, \mathcal{T}^{\Theta})$.*

6.5 Δ-rationalizability

The role of Θ-based type structures in the definition of ICR and IIR is to represent assumptions about players' interactive beliefs concerning exogenous payoff uncertainty, Θ. An alternative approach is to adapt the notion of Δ-rationalizability discussed in Sec. 3.3. Doing so is straightforward: for every player i, let $\Delta_i \subset \Delta(\Theta \times S_{-i})$ represent the restrictions on i's first-order beliefs that we would like to maintain. Notice that these restrictions can also be about i's opponents' strategies, not just the exogenous uncertainty. The set of Δ-rationalizable profiles, which we continue to denote by $S^{\Delta,\infty}$, can then be defined exactly as in Def. 11, with the understanding that players best-respond to conjectures $\sigma_{-i} \in \Delta(\Theta \times S_{-i})$, rather than in $\Delta(S_{-i})$. The epistemic characterization of $S^{\Delta,\infty}$ via RCBR and common beliefs in the restrictions Δ_i provided in Theorem 2 also extends, provided RCBR is defined as in Sec. 6.2.

[57] Furthermore, for any finite $k \geq 2$, we can modify the above example so that there is k-th order mutual belief in rationality, and still the conjectures do not form a Bayesian Nash equilibrium. Similarly, the necessity of common belief in ψ can be demonstrated. See Pomatto (2011).

[58] The maps ψ_i resemble, but are distinct from, the "books of play" n_i (see Eq. 14) since the former map from Θ-hierarchies, whereas the latter map from S_{-i}-hierarchies.

This framework allows us to model, for example, a situation in which players' ordinal preferences over strategy profiles are fixed (and commonly believed), but their cardinal preferences (i.e., their risk attitudes) are unknown. To study this situation, Börgers (1993) proposes the following notion of rationality: given a complete-information game $(I, (S_i, u_i)_{i\in I})$, s_i is rational if and only if there is a belief $\sigma_{-i} \in \Delta(S_{-i})$ and a function $v_i : S \to \mathbb{R}$ that is a strictly increasing transformation of u_i such that s_i is a best reply to σ_{-i} with utility function v_i. He characterizes this notion of rationality in terms of a novel pure-strategy dominance property, and argues that common belief in his notion of rationality corresponds to the iterated deletion of strategies that are pure-strategy dominated in his sense. As Borgers notes, it is straightforward to formalize this by suitably modifying the definition of the events R_i, and hence RCBR.

Instead of modifying the notion of rationality for complete-information games, we can obtain an alternative epistemic characterization of iterated pure-strategy dominance in Borgers' sense by considering a related game in which players have incomplete information about the risk preferences of their opponents, but know their ordinal rankings. To do so, we retain the usual notion of rationality for incomplete-information games, and consider the implications of RCBR and common belief of the ordinal rankings. We model common belief of the ordinal rankings by a suitable choice of payoff parameter space Θ and commonly-believed restrictions on first-order beliefs Δ. By the incomplete-information analog of Theorem 2, RCBR and common belief of the ordinal rankings characterize the set $S^{\Delta,\infty}$ of Δ-rationalizable profiles. Given the choice of Θ and Δ, $S^{\Delta,\infty}$ is precisely the set of iteratively pure-strategy undominated profiles. To sum up, Borgers relates iterated pure-strategy dominance to RCBR in the original complete-information game, but redefines what it means for a player to be rational. The argument described here relates iterated pure-strategy dominance to RCBR and common belief in the ordinal rankings in an associated incomplete-information game, where rationality has the usual meaning.

To make this precise, we specify the appropriate Θ and Δ. Given the complete-information game $G = (I, (S_i, u_i)_{i\in I})$, let Θ_i be the set of all payoff functions $\theta_i : S \to [0, 1]$, and Θ_i^u be the set of utilities $\theta_i^u \in \Theta_i$ that are ordinally equivalent to u_i. Furthermore, let Δ_i be the set of all finite-support probability measures $\sigma_{-i} \in \Delta(\Theta \times S_{-i})$ such that player i (i) is certain of her own *cardinal* utility, and (ii) is certain of her opponents' *ordinal* preferences: i.e., $\sigma_{-i}(\{\theta_i^u\} \times \Theta_{-i}^u \times S_{-i}) = 1$ for some $\theta_i^u \in \Theta_i^u$. Note that (i) says that Player i is certain of her own risk preferences, but these need not coincide with u_i. Finally, we define the incomplete-information game in the obvious way, that is, $(I, \Theta, (S_i, \upsilon_i)_{i\in I})$, where $\upsilon_i(\theta, s) = \theta_i(s)$ for every $\theta \in \Theta$ and $s \in S$. With these definitions, a strategy $s_i \in S_i$ is a best reply for i in the incomplete-information game $(I, \Theta, (S_i, \upsilon_i)_{i\in I})$ if and only if it is a best reply for i given *some* (complete-information) payoff function $\theta_i : S \to \mathbb{R}$ that is ordinally equivalent to u_i, i.e., if it is not pure-strategy dominated in the sense of Borgers. It follows that $S^{\Delta,\infty}$ coincides with iterated pure-strategy dominance; thus, by the incomplete-information analog of Theorem 2, iterated pure-strategy dominance in the complete-information game $(I, (S_i, u_i)_{i\in I})$ is characterized by RCBR with the restrictions Δ specified here for the incomplete-information game $(I, \Theta, (S_i, \upsilon_i)_{i\in I})$.

6.6 Discussion

The epistemic analysis of games of incomplete information is recent and indeed incomplete. The results above make three points. First, IIR is suitable only when there are no redundant types and no unmodeled correlation. ICR on the other hand is a robust solution concept that corresponds to RCBR. Finally, we saw that equilibrium concepts are difficult to characterize with clean and insightful epistemic conditions. The assumptions required are more demanding than the analogous concepts in complete-information games.[59]

As we do throughout this chapter, in this section we continue to interpret type structures merely as representations of hierarchies of beliefs. As we noted in Section 1.1 type structures are also used to model hard information. If one takes this view, then one would want to replace the independence assumption CI with the statement that, conditional on *types*, beliefs are independent. Together with the other epistemic assumptions in Theorem 14, this would fully characterize IIR. However, we prefer to model such hard information explicitly and distinctly from types. The basic space of uncertainty should include the hard information, and each player's hierarchy of beliefs should be consistent with the hard information received. In such a structure, types associated with different hard information *are* distinguishable – they have different beliefs about the information. In particular, the suitable analogs to type structures $\mathcal{T}^N$ and $\mathcal{T}^R$ are not equivalent in terms of hierarchies, and the types in the latter are not redundant. Consequently, the epistemic assumptions in Theorem 14 would yield different answers for the two structures, consistently with IIR.

We do not discuss almost common belief in the payoff structure within the context of games with incomplete information. This is an issue which, starting from Rubinstein (1989)'s email game, has led to many interesting developments. Among others, these include Monderer and Samet (1989)'s introduction of p-belief in game theory, and the literature on global games and robustness (see, e.g., Carlsson and Van Damme, 1993; Morris and Shin, 2003; Kajii and Morris, 1997a; Weinstein and Yildiz, 2007). Some of these issues (in particular, almost common belief in rationality for games with incomplete information) may benefit from epistemic analysis.

7 Extensive-form games

7.1 Introduction

The results in the previous sections apply *verbatim* to the strategic form of a multistage game. However, merely invoking those results disregards an essential implication of the dynamic nature of the interaction: players may be *surprised* in the course of game play. As is familiar from the textbook presentation of extensive-form solution concepts such as sequential equilibrium (Kreps and Wilson, 1982), and refinements such as the Intuitive Criterion of Cho and Kreps (1987), different assumptions about beliefs at unexpected histories can lead to very different predictions. Epistemic game theory provides a rich framework to analyze such assumptions. We now illustrate this using as an example the debate on the relationship between backward induction and "common belief in rationality."

Consider the three-legged centipede game in Fig. 2. A common, informal—and, it turns out, controversial—argument suggests that "common belief in rationality" implies that the backward-induction (BI) outcome should be played: Player 1, if rational, will choose D at the third node; but then, if Player 2 is rational and believes that 1 is rational, he should choose d at the second node; finally, if Player 1 is rational and anticipates all this, she will choose D at the initial node.

The problem with this informal argument is that it is not immediately obvious what is meant by 'belief' and 'rationality' in an extensive game such as the Centipede. Trivially, if one assumes that players commit to their strategies (perhaps by delegating actual play to agents, or machines) at the beginning of the game, and are rational in the sense of Sec. 3, then one reduces the situation to a game with simultaneous moves. It is easy to verify that, in that strategic-form game, RCBR only eliminates strategy AA for Player 1. Thus, choosing D at the initial node is consistent with

[59] This point, from a different perspective, is consistent with the work on learning in games, which argues for weaker concepts than Nash equilibrium, such as self-confirming equilibrium (see, e.g., Fudenberg and Levine (1993); Dekel, Fudenberg, and Levine (1999); Dekel, Fudenberg, and Levine (2004); and Fudenberg and Kamada (2012)).

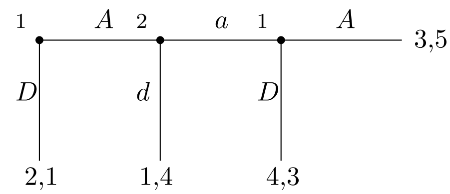


Figure 2: A three-legged centipede game

RCBR, but so is choosing A at the first node and then D at the third node. In particular, player 1 will rationally commit to play strategy AD if she believes that 2 will play a with sufficiently high probability. In turn, player 2 will commit to play a if he expects 1 to actually choose D at the first node, because in this case his choice will not matter. So, the profile (AD, a) is consistent with RCBR.

A possible objection to this argument begins by noting that, if the game actually reaches the second node, player 2 may regret his commitment to a. At the second node, he knows that player 1 did not choose D. Thus, if he continues to believe that 1 has committed to a strategy consistent with RCBR, he concludes that 1 will play D next. In this case, player 2's commitment to a results in a net loss of 1 "util." Then, since 2's choice of a vs. d only matters if play reaches the second node, shouldn't player 2 anticipate all this and commit to d instead?

Despite their intuitive appeal, these considerations pertain to player 2's knowledge, beliefs, and expected payoffs at the second node, and as such are irrelevant from the ex-ante point of view. What then if one abandons this ex-ante perspective? Suppose one takes into account the beliefs that players *would actually hold at different points in the game*, and correspondingly adopts 'sequential rationality' as the appropriate behavioral principle (Kreps and Wilson, 1982). For player 2, this requires that his choice of a or d be optimal given the beliefs she would hold at the second node. The preceding argument now seems to apply: at that node, player 2 knows that 1 did not play D, and it appears that player 2 should conclude that 1 is rationally playing AD, whatever 2's initial beliefs might have been. Thus, it appears that the informal argument supporting the BI solution does in fact apply if one takes the dynamic nature of the game into consideration.

However, in an important contribution, Reny (1992) questions the argument just given. Reny points out that, while this argument allows for the possibility that player 2's beliefs *about 1's play* may change as the game progresses, it implicitly assumes that 2's beliefs *about 1's rationality* will not. If one instead allows 2's beliefs about 1's rationality to change, the BI prediction need not obtain, even though *initially* there is common belief in (sequential) rationality. The intuition is as follows. Suppose that, *before the game is played*, player 2 expects the BI outcome, with the usual rationalization: that is, 2 expects 1 to choose D at the initial node because 1 expects d, which is in turn justified by 2's belief that 1 will choose D at the third node. However, if the second node is reached—an event that 2 does not expect to occur—then player 2 changes his mind about 1's rationality. This is not unreasonable: after all, if 1 expects 2 to choose d, 1 should not play A at the first node, so observing A does provide circumstantial evidence to the effect that 1 is *not* rational. In particular, upon reaching the second node, player 2 revises his beliefs about 1's play and now expects 1 to choose A at the third node as well; this makes his own choice of a sequentially rational. To sum up, player 2's beliefs may *initially* be consistent with BI, and—informally—with RCBR, and yet he may (plan to) play a. Furthermore, suppose player 1 is rational and has correct beliefs about 2's strategy and beliefs. Then, 1 will rationally choose A at the first node, and then D at the third. The point is that, while 2's beliefs about 1's behavior and beliefs are incorrect, both players' *initial* beliefs are consistent with RCBR. Yet, BI does not obtain.[60]

BI can be obtained if we translate into an epistemic model the assumptions that players make conditionally optimal choices at all information sets, and that they believe the same is true of opponents at all subsequent decision nodes; see subsection 7.5. We find it more surprising that it also follows from a strengthening of the notion of belief that is motivated by forward-induction ideas, i.e., from requiring that players "believe in rationality as much as possible": see Corollary 1 in subsection 7.4.

This example illustrates that a careful, explicit modeling of interactive beliefs is extremely important in extensive games. In particular, Reny's argument points out a hidden assumption in the informal argument relating 'common belief in rationality' and BI. These subtle issues suggest that a formal epistemic analysis may be insightful. Indeed, as with simultaneous-move games, epistemic models provide the language and tools to study the implications of common belief in rationality in dynamic games, and to study how modifications of the notion of rationality and belief characterize different solution concepts. Moreover, as before, dynamic epistemic models can and should force the theorist to make all assumptions explicit (or, at least, they can make it easier to spot hidden assumptions).

The present chapter emphasizes this role of epistemic models. We begin by characterizing initial RCBR in multistage games with observed actions, based on the work of Ben-Porath (1997) and Battigalli and Siniscalchi (1999). We then introduce the notion of *strong belief*, and show that, in complete type structures, rationality and common strong belief thereof (RCSBR) is characterized by *extensive-form rationalizability* (Pearce, 1984; Battigalli, 1997). RCSBR captures a (strong) principle of *forward induction* (Kohlberg and Mertens, 1986), as we indicate by means of examples. At the same time, RCSBR in complete type structures implies that the backward-induction outcome will obtain in generic perfect-information games; thus, it provides sufficient conditions for BI (see Battigalli and Siniscalchi (2002)).

7.2 Basic ingredients

For notational convenience, we focus on multistage games with observed actions (Fudenberg and Tirole, 1991a; Osborne and Rubinstein, 1994). These are extensive games in which play proceeds in stages. Perfect-information games are an example, but more generally, two or more players may be active in a given stage, in which case they choose simultaneously. The crucial assumption is that players observe all past choices at the beginning of each stage. Although we shall not do so, it is trivial to add incomplete information to such games (see Subsection 7.7). Most extensive games of interest in applications are in fact multistage games with observed actions (henceforth, "multistage game") and possibly incomplete information.

We follow the definition of multistage games in Osborne and Rubinstein (1994, section 6.3.2), and introduce here only the notation that we need subsequently. We identify a multistage game Γ with the tuple $(I, \mathcal{H}, (\mathcal{H}_i, S_i(\cdot), u_i)_{i\in I})$, where:

- $\mathcal{H}$ is the set of (terminal and non-terminal) histories in Γ. In a perfect-information game, these are (possibly partial) sequences of actions. In the Centipede game of Fig. 2, $\langle A, a, D\rangle$ is a (terminal) history. In a general multistage game, histories are sequences of action profiles. For simplicity, the profiles only indicate actions of players who have non-trivial choices. In the game of Fig. 3, $\langle In\rangle$ and $\langle In, (T, L)\rangle$ are both histories.

[60] One might further conjecture that assuming RCBR at *every* history may deliver the BI solution. However, in many games, RCBR *cannot* possibly hold at every history: this is the case for centipede games, for instance. See Reny (1993) and Battigalli and Siniscalchi (1999) for details and additional results.

- $\mathcal{H}_i$ is the subset of $\mathcal{H}$ where Player i has non-trivial choices, and $\emptyset$ denotes the initial history;
- S_i is the set of strategic-form strategies of Player i: these are maps from histories to actions;
- $\mathcal{H}_i(s_i)$ is the set of histories in $\mathcal{H}_i$ that are *not precluded* if i plays strategy s_i (whether or not any $h \in \mathcal{H}_i(s_i)$ is actually reached thus only depends upon the play of i's opponents);
- $S_i(h)$ is the set of strategies of player i that allow history h to be reached. The Cartesian product $S(h) = \prod_i S_i(h) = S_i(h) \times S_{-i}(h)$ is the set of strategy profiles that reach h.[61]. Note also that $S_i(\emptyset) = S_i$;
- $u_i : S \to \mathbb{R}$ is i's strategic-form payoff function.

A history represents a (possibly partial) path of play. In a perfect-information game, a history is an ordered list of actions

In order to analyze players' reasoning at each point in the game, it is necessary to adopt an expanded notion of probabilistic beliefs, and correspondingly redefine type structures. Specifically, we need a model of *conditional beliefs*. Following Ben-Porath (1997) (see also Battigalli and Siniscalchi, 1999), we adopt the following notion, originally proposed by Rényi (1955).

Definition 21 *Fix a measurable space (Ω, Σ) and a countable collection $\mathcal{B} \subset \Sigma$. A* ***conditional probability system****, or CPS, is a map $\mu : \Sigma \times \mathcal{B} \to [0,1]$ such that:*

1. *For each $B \in \mathcal{B}$, $\mu(\cdot|B) \in \Delta(\Omega)$ and $\mu(B|B) = 1$.*
2. *If $A \in \Sigma$ and $B, C \in \mathcal{B}$ with $B \subset C$, then $\mu(A|C) = \mu(A|B) \cdot \mu(B|C)$.*

The set of CPSs on (Ω, Σ) with conditioning events $\mathcal{B}$ is denoted $\Delta^{\mathcal{B}}(\Omega)$.

A conditional probability system is a belief over a space of uncertainty Ω, together with conditional beliefs over a collection $\mathcal{B}$ of conditioning events $B \subset \Omega$.[62] In the simplest application of this definition we shall consider, we take the point of view of Player i: the domain Ω of her uncertainty is the set S_{-i} of strategy profiles that may be played by her opponents, and, roughly speaking, $\mathcal{B}$ is the set of i's information sets. Formally, if play reaches a (non-terminal) history $h \in \mathcal{H}$, Player i can infer that her opponents played a strategy profile in $S_{-i}(h)$. Thus, the relevant set of conditioning events is $\mathcal{B} \equiv \{S_{-i}(h) : h \in \mathcal{H}\}$, and $\mu(\cdot|S_{-i}(h))$ denotes the conditional belief held by i at history h.

Condition 2 is the essential property of CPSs: it requires that Bayesian updating be applied "whenever possible." Note that player i's initial beliefs are given by $\mu(\cdot|S_{-i}(\emptyset)) = \mu(\cdot|S_{-i})$. Now suppose that, in a given play of the game, the the history reached next is h. If Player i initially considered h possible—that is, if $\mu(S_{-i}(h)|S_{-i}) > 0$—then her beliefs $\mu(\cdot|S_{-i}(h))$ must be obtained from $\mu(\cdot|S_{-i})$ via the usual updating formula. If, on the other hand, Player i initially assigned zero probability to the event that h was reached (more precisely, to opponents' strategies that allow h to be reached), then the Bayesian updating formula clearly cannot apply, and $\mu(\cdot|S_{-i}(h))$ is unconstrained (except for the natural requirement that $\mu(S_{-i}(h)|S_{-i}(h)) = 1$). However, suppose that the history reached after h is h', and $\mu(S_{-i}(h')|S_{-i}(h)) > 0$: in this case, Condition 2 requires

[61] For general extensive games, one can define the sets $S(h)$, $S_i(h)$ and $S_{-i}(h)$; the equality $S(h) = S_i(h) \times S_{-i}(h)$ requires perfect recall. The class of games we consider do satisfy perfect recall, so this equality holds.

[62] For a decision-theoretic analysis of conditional probability systems, see Myerson (1997), Blume et al. (1991), and Siniscalchi (2011).

that Player i derive $\mu(\cdot|S_{-i}(h'))$ from $\mu(\cdot|S_{-i}(h))$ via Bayesian updating. That is, following a surprise event, a player is allowed to revise her beliefs in an essentially unconstrained way; however, once she has done so, she has to conform to Bayesian updating until a new surprise event is observed. Analogous assumptions underlie solution concepts such as sequential equilibrium (Kreps and Wilson, 1982) or perfect Bayesian equilibrium (Fudenberg and Tirole, 1991b); however, these equilibrium concepts add further restrictions on beliefs following surprise events.

For example, consider the centipede game of Fig. 2, and suppose that 2's initial beliefs μ are $\mu(\{AD\}|S_1) = \mu(\{D\}\,|S_1) = 0.5$. Then, conditional upon reaching the second node, player 2 must assign probability one to AD. If instead $\mu(\{D\}\,|S_1) = 1$, then 2's conditional beliefs at the second node must assign probability zero to strategies that choose D at the initial node, but are otherwise unconstrained.

We now define sequential rationality with respect to a CPS over opponents' strategies. Unlike, e.g., Kreps and Wilson (1982), but as in, e.g., Rubinstein (1991), Reny (1992), and Dekel et al. (1999), we do not require that a strategy s_i of Player i be optimal at *all* histories, but only at those that are not ruled out by s_i itself.[63]

Here and subsequently, we denote by $\mathcal{B}_{-i}$ the conditioning events $S_{-i}(h)$, $h \in \mathcal{H}$.

Definition 22 ((Weak) Sequential Rationality) *Fix a player $i \in I$, a CPS $\mu \in \Delta^{\mathcal{B}_{-i}}(S_{-i})$ and a strategy $s_i \in S_i$. Say that s_i is a **sequential best response** to μ iff, for all $h \in \mathcal{H}_i(s_i)$ and all $s_i' \in S_i(h)$,*

$$\mathrm{E}_{\mu(\cdot|S_{-i}(h))}[u_i(s_i,\cdot)] \geq \mathrm{E}_{\mu(\cdot|S_{-i}(h))}[u_i(s_i',\cdot)].$$

*In this case, we say that the CPS μ **rationalizes** the strategy s_i.*

That is, the strategy specified by s_i at every information set it reaches is optimal given the conditional beliefs at that information set.

Finally, we define type structures for multistage games. The definition is analogous to the one for strategic-form games (Def. 3); the key difference is that types are mapped to CPSs (rather than probabilities) over opponents' strategies and types. An essential element of the following definition is the assumption that each type holds beliefs conditional upon reaching every history; thus, the conditioning events are of the form $S_{-i}(h) \times T_{-i}$.

Definition 23 *A type structure for the multistage game $\Gamma = (I, \mathcal{H}, (\mathcal{H}_i, S_i(\cdot), u_i)_{i\in I})$ is a tuple $\mathcal{T} = (I, (\mathcal{C}_{-i}, T_i, \beta_i)_{i\in I})$, where each T_i is a compact metric space,*

1. *$\mathcal{C}_{-i} = \{S_{-i}(h) \times T_{-i} : h \in \mathcal{H}\}$,*
2. *$\beta_i : T_i \to \Delta^{\mathcal{C}_{-i}}(S_{-i} \times T_{-i})$,*

and each β_i is continuous.[64] *We also write $\beta_{i,h}(t_i) = \beta_i(t_i)(\cdot|S_{-i}(h) \times T_{-i})$.*

Note that a type t_i for player i specifies conditional beliefs at histories h where i has non-trivial choices to make, and also histories at which i is essentially not active. In particular, this is true for $h = \emptyset$, the initial history. This simplifies the discussion of assumptions such as "common

[63] Choices specified by s_i at histories precluded by s_i itself are payoff-irrelevant. They are important in equilibrium notions such as Sequential Equilibrium because they represent other players' beliefs about i's play at counterfactual histories; in particular, the requirement that choices at such histories be optimal reflect the assumption that opponents believe i to be rational. For example, in the game of Fig. 2, 1's BI strategy DD encodes 2's belief that 1 will choose D at the last node. But, in an epistemic approach such assumptions can and should be modeled explicitly using opponents' CPSs.

[64] The set $\Delta^{\mathcal{C}_{-i}}(S_{-i} \times T_{-i})$ is endowed with the relative product topology.

belief in rationality at a history" (e.g., initial CBR). It is sometimes convenient to refer to a tuple $(s,t) = (s_i, t_i)_{i\in I}$ as a **state**.

Battigalli and Siniscalchi (1999) construct a type structure for extensive games that is canonical (types are collectively coherent hierarchies of conditional beliefs), embeds all other structures as a belief-closed subset, and is complete. Their construction extends the one we provided in Section 2 for strategic-form games. As in that section, we denote by H_i the set of X_{-i}-based hierarchies of conditional beliefs for player i, and by $\varphi_i : T_i \to H_i$ the **belief hierarchy map** that associates with each type in a type structure $\mathcal{T}$ the hierarchy of conditional beliefs that it generates (cf. Definition 5).

As in the previous sections, it is convenient to introduce explicit notation for first-order beliefs. The first-order beliefs of a type t_i in an epistemic type structure for an extensive game is a CPS on S_{-i}. Thus, given a type structure $(I, (\mathcal{C}_{-i}, T_i, \beta_i)_{i\in I})$ for the extensive game $(I, \mathcal{H}, (\mathcal{H}_i, S_i(\cdot), u_i)_{i\in I})$, the **first-order belief map** $f_i : T_i \to \Delta^{\mathcal{B}_{-i}}(S_{-i})$ for Player i is defined by letting $f_i(t_i)(\cdot|S_{-i}(h)) = \mathrm{marg}_{S_{-i}}\beta_{i,h}(t_i)$ for all $h \in \mathcal{H}$. It can be shown that $f_i(t_i)$ is indeed a CPS on S_{-i} with conditioning events $S_{-i}(h)$, $h \in \mathcal{H}$.

We now define the key ingredients of our epistemic analysis. The following is analogous to Definition 7 in Section 2.

Definition 24 (Rationality and Conditional Belief) *The event "Player i is sequentially rational" is*

$$R_i = \{(s_i, t_i) \in S_i \times T_i : s_i \textit{ is a sequential best reply to } f_i(t_i)\}.^{65}$$

For every measurable subset $E_{-i} \subset S_{-i} \times T_{-i}$ and history $h \in \mathcal{H}$, the event "Player i would believe that E_{-i} if h was reached" is

$$B_{i,h}(E_{-i}) = \{(s_i, t_i) \in S_i \times T_i : \beta_{i,h}(t_i)(E_{-i}) = 1\}.$$

7.3 Initial CBR

We begin with the simplest set of epistemic assumptions that take into account the extensive-form nature of the game, but are still close to strategic-form analysis in spirit.

Following Ben-Porath (1997), we consider the assumption that players are (sequentially) rational and *initially* commonly believe in (sequential) rationality:

$$\begin{aligned} RICBR_i^0 &= R_i, \\ RICBR_i^k &= RICBR_i^{k-1} \cap B_{i,\phi}(RICBR_{-i}^{k-1}) \quad \text{for} \quad k > 0, \\ RICBR_i &= \bigcap_{k\geq 0} RICBR_i^k. \end{aligned} \tag{22}$$

Except for the fact that rationality is interpreted in the sense of Def. 24, these epistemic assumptions are analogous to RCBR in simultaneous-moves games, as defined in Eqs. 9 and 10. Direct restrictions on beliefs are imposed only at the beginning of the game–the "I" in $RICBR_i^k$ refers to this feature. In particular, following a surprise move by an opponent, Player i's beliefs are not constrained.

We now illustrate the above definitions. Table 12 represents a type structure for the Centipede game in Fig. 2. Because we need to represent beliefs at different points in the game, we adopt a more compact notation than in the preceding sections. For each player, we indicate a numbered

[65] This is a slight abuse of notation, because we have used R_i to denote strategic-form rationality in Def. 7.

(non-exhaustive) list of strategy-type pairs; for each such pair (s_i, t_i), we describe the (conditional) beliefs associated with type t_i as a probability vector over the strategy-type pairs of the opponent. By convention, all strategy-type pairs that are not explicitly listed are assigned zero probability at all histories. Furthermore, we omit beliefs at histories that are not relevant for the analysis.

	(s_1,t_1)	$\beta_{1,\phi}(t_1)$	$\beta_{1,\langle A,a\rangle}(t_1)$
1	(D,t_1^1)	(1,0)	(0,1)
2	(AD,t_1^2)	(0,1)	(0,1)
3	(AA,t_1^3)	$(\frac{1}{2},\frac{1}{2})$	(0,1)

	(s_2,t_2)	$\beta_{2,\phi}(t_2)$	$\beta_{2,\langle A\rangle}(t_2)$
1	(d,t_2^1)	(1,0,0)	(0,1,0)
2	(a,t_2^2)	(1,0,0)	(0,0,1)

Table 12: A type structure for the Centipede game of Fig. 2.

To interpret, each row in the two tables corresponds to a pair consisting of a strategy and a type for a player. Each vector in such a row is that type's probability distribution over the rows of the other player's table, conditional on reaching the history described by the column label. For example, consider the row numbered '1' in the table on the left, which corresponds to strategy-type pair (D, t_1^1) of player 1. The vector $(1, 0)$ indicates that, at the initial history ϕ, type t_1^1 believes that player 2 would choose d at the second node, and that 2's type is t_2^1; the vector $(0, 1)$ indicates that at the third node, i.e. after history $\langle A, a\rangle$, type t_1^1 believes that player 2 actually chose a at the second node. The interpretation of the other types is similar. Since we do not list strategy-type pair (D, t_1^2) in the table on the left, player 2 assigns probability 0 to it at every history.

State (D, d, t_1^1, t_2^1) supports the BI prediction. Player 1 chooses D at the initial node ϕ because she expects 2 to play d at the second node $\langle A\rangle$; Player 2 initially expects 1 to play D at ϕ, but indeed plans to choose d at $\langle A\rangle$ because, should he observe A, he would revise his beliefs and conclude that 1 is actually (rationally) playing AD. Instead, state (AD, a, t_1^2, t_2^2) corresponds to Reny's story, as formalized by Ben-Porath: player 1 initially expects 2 to choose a, and thus best-responds with AD; player 2 initially expects 1 to play D and to hold beliefs consistent with backward induction, but upon observing A he revises his beliefs and concludes that 1 is actually irrational, and will continue with A at the third node. Both states are consistent with RICBR. We obtain $RICBR_1^1 = \{(D, t_1^1), (AD, t_1^2)\}$ and $RICBR_2^1 = \{(d, t_2^1), (a, t_2^2)\}$. (Recall that subscripts refer to players and superscripts to iterations.) Note that all strategy-type pairs for 2 are in $RICBR_2^1$; hence, every type for 1 in $RICBR_1^1$ trivially assigns probability one to $RICBR_2^1$ at the initial history, so $RICBR_1^2 = RICBR_1^1$. Moreover, every type of 2 *initially* assigns probability one to (D, t_1^1), which is in $RICBR_1^1$; hence, $RICBR_2^2 = RICBR_2^1$. Repeating the argument shows that $RICBR_i^k = RICBR_i^1$ for all $k \geq 1$. Thus, as claimed in the introduction to this section, the BI prediction is consistent with RICBR, but so is the profile (AD, a).

As is the case for RCBR in simultaneous-move games, RICBR can be characterized via an iterative deletion algorithm (Battigalli and Siniscalchi, 1999) as well as a suitable notion of best reply set. Initial rationalizability (Definition 25) is like rationalizability, in that it iteratively deletes strategies that are not best replies. In each iteration, players' beliefs are restricted to assign positive probability only to strategies that survived the previous rounds. The differences are that here 'best reply' means 'sequential best reply' to a CPS, and only beliefs at the beginning of the game are restricted. Definition 26 is similarly related to best-reply sets, as in Definition 9.

Definition 25 (Initial Rationalizability) *Fix a multistage game $(I, \mathcal{H}, (\mathcal{H}_i, S_i(\cdot), u_i)_{i\in I})$. For every player $i \in I$, let $S_{i,\phi}^0 = S_i$. Inductively, for every $k > 0$, let $S_{i,\phi}^k$ be the set of strategies $s_i \in S_i$ that are sequential best replies to a CPS $\mu \in \Delta^{\mathcal{B}_{-i}}(S_{-i})$ such that $\mu(S_{-i,\phi}^{k-1}|S_{-i}) = 1$. Finally, the set of **initially rationalizable** strategies for i is $S_{i,\phi}^\infty = \bigcap_{k\geq 0} S_{i,\phi}^k$.*

Definition 26 *Fix a multistage game* $(I, \mathcal{H}, (\mathcal{H}_i, S_i(\cdot), u_i)_{i\in I})$*. A set* $B = \prod_{i\in I} B_i \subset S$ *is a* ***sequential best-reply set*** *(or SBRS) if, for every player* $i \in I$*, every* $s_i \in B_i$ *is a sequential best reply to a CPS* $\mu_{-i} \in \Delta^{\mathcal{B}_{-i}}(S_{-i})$ *such that* $\mu(B_{-i}|S_{-i}) = 1$*.*

B *is a* ***full SBRS*** *if, for every* $s_i \in B_i$*, there is a CPS* $\mu_{-i} \in \Delta^{\mathcal{B}_{-i}}(S_{-i})$ *that rationalizes it and such that (i)* $\mu(B_{-i}|S_{-i}) = 1$*, and (ii) all sequential best replies to* μ_{-i} *are also in* B_i*.*

One can easily see that, in any extensive game, S^∞_ϕ is the largest SBRS. We have:

Theorem 16

1. *In any type structure* $\mathcal{T}$*,* $\text{proj}_S RICBR$ *is a full SBRS;*
2. *in any complete type structure* $\mathcal{T}$*,* $\text{proj}_S RICBR = S^\infty_\phi$*;*
3. *for every full SBRS* Q *there exists a type structure* $\mathcal{T}$ *such that* $Q = \text{proj}_S RICBR$*.*

Ben-Porath (1997) shows that, in generic perfect-information games, RICBR characterizes the $S^\infty W$ procedure discussed in Sec. 5. Since S^∞_ϕ coincides with $S^\infty W$ in such games, Theorem 16 generalizes Ben-Porath's. Thus, for generic perfect-information games, Theorem 16 or, equivalently, Ben-Porath's result, provide an alternative, but related, interpretation of the $S^\infty W$ procedure: instead of relying on common p-belief in strategic-form rationality (Definition 7), RICBR imposes common 1-belief in sequential rationality (Definition 22) at the beginning of the game, but no restrictions on beliefs at other points in the game.

7.4 Forward Induction

While the literature has considered a wide variety of 'forward-induction' notions,[66] a common thread emerges: surprise events are regarded as arising out of purposeful choices of the opponents, rather than mistakes or 'trembles.' In turn, this implies that a player may try to draw inferences about future play from a past surprising choice made by an opponent. This leads to restrictions on beliefs conditional upon unexpected histories—precisely the beliefs that RICBR does not constrain.

In this section we consider a particular way to constrain beliefs at unexpected histories, namely iterated strong belief in rationality. We first define strong belief (section 7.4.1) and provide examples showing how strong belief in rationality yields forward and backward induction (section 7.4.2). After providing the characterization results in Section 7.4.3, we discuss important properties of the notion of strong belief in Section 7.4.4.

7.4.1 Strong Belief

Stalnaker (1998) and, independently, Battigalli and Siniscalchi (2002) introduce the notion of 'strong belief' and argue that it is a key ingredient of forward-induction reasoning:

Definition 27 (Strong Belief) *Fix a type structure* $(I, (\mathcal{C}_{-i}, T_i, \beta_i)_{i\in I})$ *for an extensive game* $(I, \mathcal{H}, (\mathcal{H}_i, S_i(\cdot), u_i)_{i\in I})$*. For any player* $i \in I$ *and measurable subset* $E_{-i} \subset S_{-i} \times T_{-i}$*, the event "Player* i ***strongly believes*** *that* E_{-i}*" is*

$$SB_i(E_{-i}) = \bigcap_{h\in\mathcal{H}:[S_{-i}(h)\times T_{-i}]\cap E_{-i}\neq\emptyset} B_{i,h}(E_{-i}).$$

[66] The expression 'forward induction' was coined by Kohlberg and Mertens (1986). The Battle of the Sexes with an outside option is an early example, which Kreps and Wilson (1982) attribute to Kohlberg. See also Cho and Kreps (1987). A recent axiomatic approach is proposed by Govindan and Wilson (2009) (though axioms are imposed on solution concepts, not behavior or beliefs).

In words: if E_{-i} *could* be true in a state of the world where h can be reached,[67] then, upon reaching h, Player i must believe that E_{-i} *is* in fact true. More concisely: *Player i believes that E_{-i} is true whenever possible.*

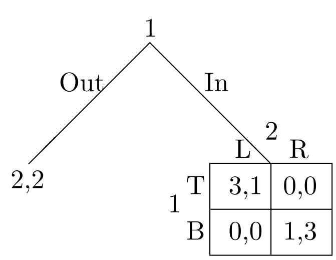
7.4.2 Examples

In this subsection we discuss the implications of strong belief in rationality for the Centipede game (Fig. 2) and the Battle of the Sexes with an outside option (Fig. 3; see Kohlberg and Mertens, 1986).

Example 7 (The Centipede Game) Consider again the type structure in Table 12. As noted above, type t_2^2 of player 2 initially believes that 1 is rational, but becomes convinced that Player 1 is irrational in case 1 chooses A at ϕ. However, note that there is a rational strategy-type pair for player 1 that chooses A at ϕ, namely (AD, t_1^2). Strong belief in 1's rationality, $SB_2(R_1)$, then requires player 2 to believe at the second node, i.e., at history $\langle A\rangle$, that 1 is rational. Therefore, type t_2^2 of player 2 is *not* consistent with strong belief in 1's rationality, because, conditional on $\langle A\rangle$, he assigns probability one to 1 playing the irrational strategy AA. On the other hand, consider now type t_2^1: upon seeing $\langle A\rangle$ type t_2^1 assigns probability one precisely to 1's strategy-type pair (AD, t_1^2); therefore, t_2^1 is the only type of 2 that is consistent with strong belief in 1's rationality.

Since $SB_2(R_1) = \{t_2^1\}$ and t_2^1 expects player 1 to play D at the third node, $R_2 \cap SB_2(R_1) = \{(d, t_2^1)\}$. That is, the joint assumptions that 2 is rational and that he strongly believes in 1's rationality yield the conclusion that 2 should plan to play d. Thus, in the type structure of Table 12, rationality and strong belief in rationality eliminates the non-BI outcome (AD, a). Observe that this is achieved not by arguing that, at the second node, player 2 believes that 1's initial choice of A was a mistake—an unintended deviation from her planned strategy; rather, player 2 interprets prior actions as purposeful, insofar as this is possible. If in addition player 1 is rational and strongly believes $R_2 \cap SB_2(R_1)$, one obtains the *backward*-induction outcome via *forward*-induction reasoning. We will return to this point in subsection 7.4.3. ■

Example 8 (Battle of the Sexes with an outside option) Consider the game in Fig. 3.



	(s_1, t_1)	$\beta_{1,\phi}(t_1)$	$\beta_{1,\langle\text{In}\rangle}(t_1)$
1	$(InB,\, t_1^1)$	0,1,0	0,1,0
2	$(InT,\, t_1^1)$	0,1,0	0,1,0
3	$(Out,\, t_1^1)$	0,1,0	0,1,0
4	$(InT,\, t_1^2)$	0,0,1	0,0,1

	(s_2, t_2)	$\beta_{2,\phi}(t_2)$	$\beta_{2,\langle\text{In}\rangle}(t_2)$
1	$(L,\, t_2^1)$	0,1,0,0	0,1,0,0
2	$(R,\, t_2^2)$	0,0,1,0	1,0,0,0
3	$(L,\, t_2^3)$	0,0,0,1	0,0,0,1

Figure 3: Battle of the Sexes with an outside option and the type structure $\mathcal{T}^{FI}$

An informal forward-induction argument runs as follows: *InB* is a strictly dominated strategy for Player 1, because *Out* yields a strictly higher payoff regardless of 2's choice. "Therefore," if the simultaneous-moves subgame is reached, Player 2 should expect Player 1 to play T, and best-respond with L. But then, if Player 1 anticipates this, she will best-respond with *In*, followed by T (i.e., she will choose strategy InT).

[67] That is, if there is a profile $(s_{-i}, t_{-i}) \in E_{-i}$ such that $s_{-i} \in S_{-i}(h)$.

The right-hand side of Figure 3 displays a type structure, denoted $\mathcal{T}^{FI}$, where strong belief in rationality reflects this reasoning process. Note that InB is irrational regardless of 1's beliefs, and furthermore InT is irrational for 1's type t_1^1, because this type expects 2 to play R. Thus, $R_1 = \{(Out, t_1^1), (InT, t_1^2)\}$; moreover, all strategy-type pairs are rational for Player 2. Neither types t_2^1 nor type t_2^2 are in $SB_2(R_1)$. To see this, first note that, conditional on every history, type t_2^1 assigns probability one to the irrational strategy-type pair (InT, t_1^1). Second, type t_2^2 initially believes that Player 1 rationally chooses Out, but upon observing $\langle In \rangle$, he switches to the belief that 1 plays the irrational strategy InB. On the other hand, t_2^3 is consistent with $SB_2(R_1)$. Since furthermore L is rational for type t_2^3, we have $R_2 \cap SB_2(R_1) = \{(L, t_2^3)\}$. Consequently, if one further assumes that 1 strongly believes that $R_2 \cap SB_2(R_1)$, type t_1^1 of player 1 must be eliminated (because it assigns probability one to $(R, t_2^2) \notin R_2 \cap SB_2(R_1)$ at every history). Thus, $R_1 \cap B_{1,\phi}(R_2 \cap SB_2(R_1)) = \left\{(InT, t_1^2)\right\}$. We have obtained the forward-induction outcome of this game, as claimed. Notice that the assumption that 1 initially believes that $R_2 \cap SB_2(R_1)$ is an assumption on how 1 expects 2 to revise his beliefs in case 2 is surprised: specifically, 1 expects 2 to maintain the belief that 1 is rational as long as possible. ■

In the preceding examples, iterated strong belief in rationality selects backward- and forward-induction outcomes. Theorem 17 and Corollary 1 in section 7.4.3 show that in complete type structures this always holds.[68] The following example shows that, in arbitrary, small type structures, these results need not hold. Section 7.4.4 discusses the reasons for the different conclusions reached in Examples 8 and 9. Theorem 17 also provides a characterization of iterated strong belief in rationality for arbitrary type structures.

Example 9 Consider the type structure in Table 13, denoted $\mathcal{T}^{NFI}$, for the game in Fig. 3.

	(s_1, t_1)	$\beta_{1,\phi}(t_1)$	$\beta_{1,(\text{In})}(t_1)$
1	(InB, t_1^1)	0,1	0,1
2	(InT, t_1^1)	0,1	0,1
3	(Out, t_1^1)	0,1	0,1

	(s_2, t_2)	$\beta_{2,\phi}(t_2)$	$\beta_{2,(\text{In})}(t_2)$
1	(L, t_2^1)	0,1,0	0,1,0
2	(R, t_2^2)	0,0,1	1,0,0

Table 13: Type structure $\mathcal{T}^{NFI}$ for the Battle of the Sexes.

Note that, relative to the type structure $\mathcal{T}^{FI}$ given in the table in Fig. 3, we have removed type t_1^2 for Player 1. As a consequence, now $R_1 = \left\{(Out, t_1^1)\right\}$. This implies that there is *no* rational strategy-type pair of Player 1 who plays In in the type structure $\mathcal{T}^{NFI}$. Therefore, upon observing In, Player 2's beliefs are unconstrained by strong belief; thus, type t_2^2 *is* consistent with $\text{SB}_2(R_1)$ in $\mathcal{T}^{NFI}$. Therefore, repeating the analysis in subsection 7.4.2 now leads to a different conclusion: if Player 1 initially (or strongly) believes $R_2 \cap SB_2(R_1)$, and if she is rational, she will choose the action Out at the initial node, contrary to what the standard FI argument for this game predicts.

7.4.3 RCSBR and Extensive-Form Rationalizability

Battigalli and Siniscalchi (2002) consider the implications of rationality and common strong belief in rationality (RCSBR), which is the strong-belief counterpart of Eq. (8) and (9) for belief and,

[68]Recall that in a complete type structure the belief maps are onto.

respectively, Eq. (18) for p-belief. For every player $i \in I$, let

$$\begin{aligned} RCSBR_i^0 &= R_i, \\ RCSBR_i^k &= RCSBR_i^{k-1} \cap SB_i(RCSBR_{-i}^{k-1}) \quad \text{for} \quad k > 0, \\ RCSBR_i &= \bigcap_{k \ge 0} RCSBR_i^k. \end{aligned} \tag{23}$$

As was the case for iterated p-belief, it *does* matter whether we define mutual and common strong belief as above, or by iterating the strong belief operator as in Eq. (7). Once again, the reason is that strong belief does not satisfy the Conjunction property in Eq. (5). We discuss this further in section 7.4.4.

To see why RCSBR selects the forward-induction outcome, as illustrated by Example 8, we study Eq. (23) in more detail. Consider the two-player case for simplicity, take the point of view of player 1, and focus on $k = 2$ to illustrate:

$$RCSBR_1^2 = R_1 \cap SB_1(R_2) \cap SB_1\big(R_2 \cap SB_2(R_1)\big). \tag{24}$$

Note that, since $R_2 \cap SB_2(R_1) \subset R_2$, every history h in an arbitrary game can fall into one of three categories: (0) histories inconsistent with R_2 and hence a fortiori with $R_2 \cap SB_2(R_1)$; (1) histories consistent with R_2 but not with $R_2 \cap SB_2(R_1)$; and (2) histories consistent with $R_2 \cap SB_2(R_1)$ and hence a fortiori R_2. Eq. (24) requires the following: at histories of type 1, player 1 should assign probability one to R_2; at histories of type 2, she should assign probability one to $R_2 \cap SB_2(R_1)$.[69] Interpreting $R_2 \cap SB_2(R_1)$ as a 'strategically more sophisticated' assumption about 2's behavior and beliefs than R_2, Eq. (24) requires that, at any point in the game, players draw inferences from observed play by attributing the highest possible degree of strategic sophistication to their opponents.[70]

Theorem 17 below shows that RCSBR is characterized by extensive-form rationalizability (Pearce, 1984) in any complete type structure, and by the notion of 'extensive-form best reply set' (Battigalli and Friedenberg, 2010) in arbitrary type structures. Extensive-form rationalizability is similar to initial rationalizability (Definition 25), except that beliefs are restricted to assign positive probability to strategies that survive the previous rounds at *all* histories where it is possible to do so: see Eq. (25). Extensive-form best-reply sets bear the same relationship to SBRSs (Definition 26).

Definition 28 (Extensive-Form Rationalizability) *Fix a multistage game $(I, \mathcal{H}, (\mathcal{H}_i, S_i(\cdot), u_i)_{i\in I})$. For every player $i \in I$, let $\hat{S}_i^0 = S_i$. Inductively, for every $k > 0$, let $\hat{S}_i^k$ be the set of strategies $s_i \in \hat{S}_i^{k-1}$ that are sequential best replies to a CPS $\mu \in \Delta^{\mathcal{B}_{-i}}(S_{-i})$ such that*

$$\text{for every } h \in \mathcal{H}_i, \qquad S_{-i}(h) \cap \hat{S}_{-i}^{k-1} \neq \emptyset \quad \text{implies} \quad \mu(\hat{S}_{-i}^{k-1} | S_{-i}(h)) = 1. \tag{25}$$

*$\hat{S}^\infty = \bigcap_{k\ge 0} \hat{S}^k$ is the set of **extensive-form rationalizable** strategy profiles.*

Definition 29 *Fix a two-player multistage game $(\{1,2\}, \mathcal{H}, (\mathcal{H}_i, S_i(\cdot), u_i)_{i=1,2})$.*[71] *A set $B = B_1 \times B_2 \subset S$ is an **extensive-form best-reply set** (or EFBRS) if, for every player $i = 1, 2$, every*

[69] Because strong belief does not satisfy conjunction, the right-hand side of Eq. (24) is *not* equivalent to $R_1 \cap SB_1(R_2 \cap SB_2(R_1))$.

[70] Battigalli (1996a) calls this the *Best Rationalization Principle*.

[71] We restrict attention to the two-player case to avoid issues of correlation (see Battigalli and Friedenberg, 2010, sec. 9c for additional discussion).

$s_i \in B_i$ is a sequential best reply to a CPS $\mu_{-i} \in \Delta^{\mathcal{B}_{-i}}(S_{-i})$ such that, for every $h \in \mathcal{H}_i$ with $S_{-i}(h) \cap B_{-i} \neq \emptyset$, $\mu(B_{-i}|S_{-i}(h)) = 1$.

*B is a **full EFBRS** if, for every $i = 1, 2$ and $s_i \in B_i$, there is a CPS $\mu_{-i} \in \Delta^{\mathcal{B}_{-i}}(S_{-i})$ that rationalizes s_i, and such that (i) $\mu(B_{-i}|S_{-i}) = 1$ for every $h \in \mathcal{H}_i$ that satisfies $S_{-i}(h) \cap B_{-i} \neq \emptyset$, and (ii) all sequential best replies to μ_{-i} are also in B_i.*

Theorem 17

1. *In any type structure $\mathcal{T}$ for a two-player multistage game,* $\text{proj}_S RCSBR$ *is a full EFBRS;*
2. *in any complete type structure $\mathcal{T}$ for an arbitrary multistage,* $\text{proj}_S RCSBR = \hat{S}^\infty$*;*
3. *for every full EFBRS Q of a two-player multistage game, there exists a type structure $\mathcal{T}$ such that $Q = \text{proj}_S RCSBR$.*

One consequence of Theorem 17 and the notion of strong belief is that, if a history h is reached under a strategy profile $s \in \hat{S}^\infty$, then there is common belief in rationality at h. Thus, while there may be histories where common belief in rationality may fail to hold, it does hold on the path(s) of play predicted by RCSBR. Though he does not use type structures, Reny (1993) defines iterative procedures motivated by the assumption that rationality is common belief at a given history.

Extensive-form rationalizability yields the BI outcome in generic perfect-information games (Battigalli, 1997; Heifetz and Perea, 2013). Combining this with Theorem 17, we obtain the following

Corollary 1 *In any complete type structure $\mathcal{T}$ for a generic perfect-information game Γ, any strategy profile $s \in \text{proj}_S RCSBR$ induces the backward-induction outcome.*

Corollary 1 thus states that RCSBR in a complete type structure provides a *sufficient* epistemic condition for the BI outcome. Note that Corollary 1 does *not* state that $s \in \text{proj}_S RCSBR$ is the (necessarily unique) BI *profile* , but only that such an s induces the BI *outcome*. Indeed, the BI profile may even be inconsistent with RCSBR. Both points are illustrated by the game in Fig. 4, due to Reny (1992).

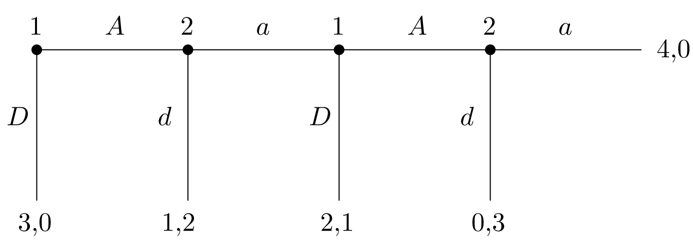


Figure 4: Backward and forward induction

The unique BI profile in this game is (DD, dd). However, the extensive-form rationalizable profiles—hence, the profiles supported by RCSBR in complete type structures—are $(\{DD, DA\}, ad)$. To see this, note that strategy AD is strictly dominated by choosing D at the initial node. Hence, if play reaches the second node, player 2 must conclude that player 1 is playing AA, which makes ad strictly better than choosing d at the second node (note that strategy aa is not sequentially rational). In turn, this leads player 1 to choose D at the first node; hence, the backward-induction outcome obtains. However, RCSBR implies that, conditional upon observing A at the first node, and hence reaching the second node, 2 would expect that 1 will continue with A at the third node. Therefore, RCSBR implies that 2 would play ad, which is not his backward-induction strategy.

7.4.4 Discussion

The different predictions in examples 8 and 9 raise several questions. First, in one case the event *RCSBR* yields the forward-induction outcome, and in the other case it does not. While we noted that the forward-induction conclusion relies on the type structure being sufficiently rich, this merits further discussion. Relatedly, in example 8 type t_2^2 in $\mathcal{T}^{FI}$ is in *RCSBR*, while in example 9 type t_2^2 in $\mathcal{T}^{NFI}$, which has exactly the same hierarchy of beliefs, is not in *RCSBR*. This raises doubts about whether RCSBR depends only on belief hierarchies, or on the type structure as well, i.e., whether it is an elicitable assumption or not. In this section we address these concerns. It is useful to begin with a discussion of properties of strong belief.

Strong belief is not monotonic, and violates conjunction (cf. Equation 5).[72] To see this, consider again the type structure in Table 12 for the Centipede game, and focus on the events

$$SB_1(R_2 \cap SB_2(R_1)) \qquad \text{and} \qquad SB_1(R_2) \cap SB_1(SB_2(R_1)).$$

As shown in Example 7, $R_2 \cap SB_2(R_1) = \{(d, t_2^1)\}$. Now observe that type t_1^1 initially assigns probability one to $\{d, t_2^1)\}$. Furthermore, if player 2 plays d, history $\langle A, a\rangle$ is not reached; hence, strong belief in $R_2 \cap SB_2(R_1) = \{(d, t_2^1)\}$ imposes no restriction on beliefs at $\langle A, a\rangle$. Therefore, type t_1^1 strongly believes $R_2 \cap SB_2(R_1)$, so $SB_1(R_2 \cap SB_2(R_1)) \neq \emptyset$. On the other hand, $SB_1(SB_2(R_1)) = \emptyset$. To see this, note first that $SB_2(R_1) = S_2 \times \{t_2^1\}$. This event is consistent with the third node, i.e. history $\langle A, a\rangle$, being reached; therefore, strong belief in $SB_2(R_1)$ requires that player 1 assign probability one to this event conditional on $\langle A, a\rangle$. However, no type of player 1 in Table 12 assigns positive probability to 2's type t_2^1 at that history. Hence, $SB_1(R_2) \cap SB_1(SB_2(R_1)) = \emptyset \neq SB_1(R_2 \cap SB_2(R_1))$.[73] This failure of Monotonicity and Conjunction plays an important role in the subsequent discussion.

Throughout this paper, we interpret events such as R_i, $B_i(E_{-i})$ or $SB_i(E_{-i})$ defined in a given type structure $\mathcal{T}$ as "player i is rational," "player i believes E_{-i}" or "player i strongly believes E_{-i}." While convenient, this is not quite accurate. Every type structure defines the set of belief hierarchies that are allowed for each player. For instance, consider type structure $\mathcal{T}^{NFI}$ in example 9 and denote by φ_i^{NFI} its belief hierarchy maps (definition 5). The event R_1 in $\mathcal{T}^{NFI}$ should be interpreted as "player 1 is rational <u>and</u> her belief hierarchy is $\varphi_1^{NFI}(t_1^1)$." Similarly, the event $B_{2,\phi}(R_1)$ should be interpreted as "player 2 initially believes that '1 is rational <u>and</u> her belief hierarchy is $\varphi_1^{NFI}(t_1^1)$' <u>and</u> 2's belief hierarchy is either $\varphi_2^{NFI}(t_2^1)$ or $\varphi_2^{NFI}(t_2^2)$." We typically avoid such convoluted statements, but must recognize that the simpler statements "1 is rational" and "2 initially believes that 1 is rational" are precise interpretations of R_1 and $B_{2,\phi}(R_1)$ only if we define these events in a rich type structure—one that generates all hierarchies of (conditional) beliefs.

Fortunately, this is not an issue when interpreting results that use monotonic belief operators. For example, consider Theorem 1. On the one hand, the event *RCBR* is accurately interpreted as "rationality and common belief in rationality," RCBR, only in a complete type structure. Indeed, in smaller type structures, as explained above, the event *RCBR* should be interpreted as RCBR jointly with additional assumptions about hierarchical beliefs. On the other hand, because of Monotonicity,[74] the hierarchies consistent with the event *RCBR* in a small type structure are also

[72] We noted in footnote 10 that Monotonicity is equivalent to the "$\subset$" part of Conjunction. Indeed, that is the only part of Conjunction that strong belief does not satisfy.

[73] It can also be shown that $SB_1(R_2) \cap SB_1(SB_2(R_1))$ is empty in any complete type structure for the Centipede game in Fig. 2.

[74] If a hierarchy of player 2 is consistent with belief in 1's rationality under the assumption that 1's beliefs are constrained in some specific way, then by Monotonicity it is also consistent with belief in 1's rationality without additional restrictions on 1's beliefs (and obviously this argument can be iterated).

consistent with $RCBR$ in a complete type structure. Consequently, every full BRS is a subset of the rationalizable set S^{∞}. Since the latter strategies are the unambiguously interpreted as consistent with RCBR (because $S^{\infty} = \text{proj}_S RCBR$ in a complete type structure), so are the former.

Now consider strong belief. A cumbersome but precise interpretation of the event $SB_2(R_1)$ in the type structure $\mathcal{T}^{NFI}$ is as follows: "player 2 believes that 'player 1 is rational and her belief hierarchy is $\varphi_1^{NFI}(t_1^1)$' at every history that can be reached if this assertion is true, and 2's belief hierarchy is either $\varphi_2^{NFI}(t_2^1)$ or $\varphi_2^{NFI}(t_2^2)$." If instead we define event $SB_2(R_1)$ in the type structure $\mathcal{T}^{FI}$ of Figure 3, its precise interpretation is, "player 2 believes that 'player 1 is rational and her belief hierarchy is either $\varphi_1^{FI}(t_1^1)$ or $\varphi_1^{FI}(t_1^2)$' at every history that can be reached if this assertion is true, and 2's belief hierarchy is $\varphi_2^{FI}(t_2^1)$,$\varphi_2^{FI}(t_2^2)$ or $\varphi_2^{FI}(t_2^3)$."

Observe that these statements are expressed in terms of strategies and hierarchies of conditional beliefs, and hence they may be elicited in principle. Thus, there is no conflict between our goal of elicitation and the notion of strong belief. The apparent conflict arises from an imprecise interpretation of strong belief in small type structures.

That said, the interpretation of the event $RCSBR$ in small type structures is subtle because strong belief does not satisfy Monotonicity. As above, the event $RCSBR$ is accurately interpreted as "rationality and common strong belief in rationality," RCSBR, only in a complete type structure. However, in contrast to the case of RCBR, due to the failure of Monotonicity, the hierarchies consistent with the event $RCSBR$ in a small type structure need *not* be consistent with $RCSBR$ in a larger (a fortiori, in a complete) type structure.[75] Hence, a full EFBRS need not be a subset of the extensive-form rationalizable set $\hat{S}^{\infty}$.[76] Thus, while the latter strategies can accurately be interpreted as consistent with RCSBR, this is not the case for the former. We can say, however, that strategies in an EFBRS are consistent with "rationality plus additional assumptions on hierarchies, and common strong belief thereof."

So, what is the "right" solution concept? If the analyst is interested in the implications of RCSBR, without any additional assumptions, then the answer is $\hat{S}^{\infty}$. (Analogously, for simultaneous-move games, the answer is S^{∞}). If the the analyst wants to impose some particular additional assumption about beliefs, then the answer is a particular EFBRS; which EFBRS depends on the assumption. (For simultaneous-move games, the answer is a particular BRS). Finally, if the analyst wants to be "cautious" and consider the predictions that would arise were she to adopt any possible assumption, then the answer is the (player-by-player) union of all EFRBS's. (For simultaneous-move games, the answer is the union of all BRS's, which in this case is again S^{∞}.) The bottom line is that, when interpreting assumptions involving strong belief, one should be careful to specify whether or not additional assumptions are imposed on players' beliefs.

These considerations apply in particular to the relationship between strong belief and forward-induction reasoning. As we noted, the basic intuition underlying forward induction is that players attempt to maintain the assumption that their opponents are rational as long as possible, in the face of unexpected behavior. This suggests that no a priori constraint is placed on players' attempt to rationalize deviations. In other words, a connection can be established between forward induction and (iterated) strong belief in *complete* type structures. Theorem 17 confirms this. When strong

[75] Recall that, in examples 8 and 9, the hierarchy $\varphi_2^{NFI}(t_2^2) = \varphi_2^{FI}(t_2^2)$ of player 2 is consistent with strong belief in the statement "1 is rational and 1's hierarchy is $\varphi_1^{NFI}(t_1^1)$," but not with strong belief in the statement "1 is rational and 1's hierarchy is either $\varphi_1^{FI}(t_1^1) = \varphi_1^{NFI}(t_1^1)$ or $\varphi_1^{FI}(t_1^2)$."

[76] Consider for instance the game in Fig. 3. As we noted above, $\hat{S}^{\infty} = \{InT\} \times \{L\}$. However, consider the set $Q = \{Out\} \times \{R\}$. The strategy *Out* for player 1 is the unique (sequential) best reply to any CPS that initially assigns probability 1 to R. Furthermore, R is the unique sequential best reply to the CPS that assigns probability one to *Out* at the beginning of the game, and to *InB* conditional on reaching the simultaneous-move subgame. Therefore, Q is an EFBRS, because Q_1 does not reach the subgame and so there are no further restrictions on 2's beliefs. Yet, the sets Q and $\hat{S}^{\infty}$ are disjoint, and induce distinct outcomes.

belief is applied in small type structures, there is an interaction between the rationalization logic of forward induction and whatever assumptions are exogenously imposed on beliefs; this yields EFBRS's, as stated in Theorem 17.[77]

7.5 Backward Induction

As we noted above, RCSBR in a complete type structure yields the backward-induction outcome, but *not* the backward-induction profile. The game of Fig. 4 provides an example. An alternative way to get backward induction is to simply make explicit the assumption inherent in BI: at all nodes where he is active, i makes conditionally optimal choices, i believes that his opponent j does so at subsequent histories where she is active, i believes that j believes that k will also choose optimally at subsequent nodes where k is active, etc.; see Perea (2011b).[78]

Aumann (1995) (see also Balkenborg and Winter, 1997) derives the backward induction profile from 'common knowledge of rationality' in a very different epistemic model.[79] As it does not explicitly incorporate belief revision, Aumann's model lies outside our framework. If we translate Aumann's assumptions into our framework, we obtain the conditions discussed in the previous paragraph.

7.6 Equilibrium

The epistemic analysis of equilibrium concepts for extensive games is largely yet to be developed. In this subsection we briefly describe results on subgame-perfect equilibrium, self-confirming equilibrium in signaling games, and the relationship between EFBRS's and (subgame-perfect) equilibrium.

A basic question is whether sufficient conditions for subgame-perfect equilibrium (Selten, 1975) can be provided by adapting the results of Aumann and Brandenburger (1995). Theorem 5, on two-player games, can be easily adapted. As this is not in the literature, we sketch the steps here.

To avoid introducing new notation, we describe i's play in an extensive-form game using a CPS σ_i on S_i, instead of using behavioral strategies.[80] Thus, a *subgame-perfect equilibrium* (SPE) of a multistage game $(I, \mathcal{H}, (\mathcal{H}_i, S_i(\cdot), u_i)_{i\in I})$ is a profile $(\sigma_i)_{i\in I}$, where $\sigma_i \in \Delta^{\mathcal{B}_i}(S_i)$ is a CPS on S_i with conditioning events $\mathcal{B}_i = \{S_i(h) : h \in \mathcal{H}\}$ for each player i, such that, at every history $h \in \mathcal{H}$, $\big(\sigma_i(\cdot|S_i(h))\big)_{i\in I}$ is a Nash equilibrium of the strategic-form game $\big(I, \big(S_i(h), u_i\big)_{i\in I}\big)$. To clarify how this definition is related to the usual ones, consider the profile (OutB, R) in the game of Fig. 3. Player 1's strategy is represented by any CPS σ_1 such that $\sigma_1(\{\mathit{OutT}, \mathit{OutB}\}|S_1) = 1$ and $\sigma_1(\{\mathit{InB}\}|S_1(h)) = 1$, where h denotes the simultaneous-move subgame. Player 2's strategy is represented by the CPS σ_2 defined by $\sigma_2(\{R\}|S_2) = \sigma_2(\{R\}|S_2(h)) = 1$. It is easy to verify that the profile of CPSs (σ_1, σ_2) satisfies the definition of SPE we have just given.

[77] An equivalent way (in our setting) to incorporate belief restrictions in the analysis is to work in the canonical type structure and explicitly define events C_i that formalize the desired additional assumptions. Then, studying the behavioral implications of events such as $R_i \cap C_i \cap SB_i(R_{-i} \cap C_{-i})$ is the same as studying the implications of the event $R_i \cap SB_i(R_{-i})$ in a type structure that incorporates the desired restrictions on beliefs. A detailed discussion of these issues can be found in Battigalli and Friedenberg (2010) and Battigalli and Prestipino (2011).

[78] Stalnaker (1998) discusses belief revision in dynamic games. In particular, he characterizes backward induction in a similar way to that discussed above, but interprets belief in subsequent optimality as a consequence of a suitable independence assumption on beliefs about future and past play. See also Perea (2008).

[79] See also Aumann (1998) and Samet (1996). Samet (2013) extends Aumann (1995)'s analysis from knowledge to belief.

[80] We use σ_i to denote the CPS on S_i that describes i's *play*, as opposed to i's *beliefs*, denoted μ_i below, which are defined on S_{-i}.

Turning to the epistemic analysis, we adapt the notation from Sec. 4: given a CPS $\mu_i \in \Delta^{\mathcal{B}_{-i}}(S_{-i})$ and a type structure $(I, (\mathcal{C}_{-i}, T_i, \beta_i)_{i\in I})$, let $[\mu_i]$ be the event that i's first-order belief is μ_i; given a profile $(\mu_i)_{i\in I}$, the event $[\mu]$ is defined as the intersection of all $[\mu_i]$ for $i \in I$. Finally, define the event "Player i makes a rational choice at history $h \in \mathcal{H}$" as

$$R_{i,h} = \left\{ (s_i, t_i) : s_i \in \arg \max_{s_i' \in S_i(h)} u_i(s_i', f_{i,h}(t_i)) \right\}.$$ [81]

To see how this is different from event R_i (definition 24), consider strategy InB in the game of Fig. 3. This is strictly dominated, hence (sequentially) irrational in the entire game; however, it does specify a choice in the simultaneous-move subgame that is a best reply to, e.g., the belief that assigns probability one to 2 playing R.

We can now state the counterpart to Theorem 5. Fix a CPS $\mu_i \in \Delta^{\mathcal{B}_{-i}}(S_{-i})$ for $i = 1, 2$. If $[\mu] \cap \bigcap_{h\in\mathcal{H}} B_h(R_h \cap [\mu]) \neq \emptyset$, then $(\sigma_1, \sigma_2) = (\mu_2, \mu_1)$ is a SPE.[82]

As is the case for simultaneous-move games, the situation is more delicate if there are more than two players. One approach is to adapt the definitions of agreement and independence of first-order beliefs in Sec. 4.4, thereby obtaining a counterpart to Theorem 6. Alternatively (cf. Barelli, 2010), one can adapt the notion of common prior (Def. 14) and obtain a counterpart to Theorem 7.[83] In order to adapt the arguments used to prove Theorem 7, the common "prior" for a type structure $(I, (\mathcal{C}_{-i}, T_i, \beta_i)_{i\in I})$ must be defined as a CPS $\mu \in \Delta^{\mathcal{B}}(S \times T)$, where $\mathcal{B} = \{S(h) : h \in \mathcal{H}\}$ such that $\beta_{i,h}(t_i) = \mathrm{marg}_{S_{-i}\times T_{-i}} \mu(\cdot|S(h) \times \{t_i\} \times T_{-i})$ for *all* histories $h \in \mathcal{H}$.[84]

Asheim and Perea (2005) provide an epistemic characterization of sequential equilibrium (Kreps and Wilson, 1982) in two-player games. In their analysis, beliefs are represented using a generalization of lexicographic probability systems.

A different approach is explored by Battigalli and Siniscalchi (2002) in the context of signaling games. We do not formalize it, because doing so would require introducing notation that we do not use anywhere else in this chapter. Roughly speaking, they show that, in any epistemic model, if there is a state in which players' first-order beliefs are consistent with an outcome of the game (that is, a probability distribution over terminal histories), and there is initial mutual belief in rationality and in the event that first-order beliefs are consistent with the outcome, then there exist a self-confirming equilibrium that induces that outcome. They also provide necessary and sufficient epistemic conditions for the outcome to be supported in a self-confirming equilibrium that satisfies the *Intuitive Criterion* of Cho and Kreps (1987).

Finally, Battigalli and Friedenberg (2010) relate EFBRSs, and hence iterated strong belief in rationality, with Nash and subgame-perfect equilibrium in two-person multistage games with observable actions. Every pure-strategy SPE is contained in some EFBRS. Moreover, under a no-relevant ties condition (Battigalli, 1997), a pure-strategy SPE profile *is* an EFBRS. In perfect-information games that satisfy the "transfer of decision-maker indifference" of Marx and Swinkels (1997), if a

[81] Recall that $f_i(t_i)$ is t_i's first-order belief, and $f_{i,h}(t_i)$ is the conditional of the CPS $f_i(t_i)$ given history h.

[82] Alternatively, one could get sufficient epistemic conditions for SPE by assuming that conjectures are given by μ and, at every history, there is mutual belief in $R \cap [\mu]$, where R is the event that all players are sequentially rational in the sense of Def. 24. This is the approach taken by Barelli (2009). However, these conditions rule out the SPE $(OutB, R)$ in the game of Fig. 3. More generally, they may preclude certain SPE in games in which some histories can only be reached if a player plays a strictly dominated strategy.

[83] In this approach, since we continue to assume that players only hold beliefs about their opponents, the independence condition in Def. 15 would also have to be adapted. Note that Barelli (2009) allows players to hold beliefs about their own strategies, and hence does not require any additional condition.

[84] Barelli (2009) notes that this notion is very demanding, because it requires no betting even conditional upon histories that players do not expect to be reached.

state (s,t) in a type structure is consistent with the event $RCSBR$, then s is outcome-equivalent to a Nash equilibrium. Conversely, in games with no relevant ties, for any Nash equilibrium profile (s_1, s_2) such that each strategy s_i is also a sequential best reply to some CPS on S_{-i}, there is a type structure and a profile t of types such that (s,t) is consistent with the event $RCSBR$.[85]

As should be clear from the above, this area is fertile ground for research. For instance, it would be interesting to investigate the implications of strong belief in rationality, and of the best-rationalization principle, in an equilibrium setting. Care is needed; for instance, one cannot assume that—as may appear natural—there is mutual or common belief in the conjectures at *every* information set, because that may be inconsistent with the logic of best rationalization.[86, 87]

7.7 Discussion

Strategies The choice-theoretic interpretation of strategies deserves some comment. In a simultaneous-move game, whether or not a player plays a given strategy is easily observed ex-post. In an extensive game, however, a strategy specifies choices at several histories as play unfolds. Some histories may be mutually exclusive, so that it is simply not possible to observe whether a player actually follows a given strategy. Furthermore, it may be the case that epistemic assumptions of interest imply that a given history h will not be reached, and at the same time have predictions about what the player on the move at h would do if, counterfactually, h was reached. Consider for instance the Centipede game of Figure 2: as we noted above, RCSBR (in the type structure of Table 12) implies that player 1 will choose D at the initial node, and that player 2 would choose d if the second node was reached. Verifying predictions about such unobserved objects seems problematic. This is troublesome both in terms of testing the theory, and because it is not obvious how to elicit players' beliefs about such objects.

One obvious way to avoid this difficulty is to assume that players commit to observable contingent plans at the beginning of the game. While this immediately addresses the issue of verifiability, it seems to do so at the cost of turning the extensive game into a strategic-form game. However, one can impose the requirement that players prefer their plans to be conditionally, not just ex-ante optimal, even at histories they do not expect to be reached.[88] In this case, while players commit to specific plans, the extensive-form structure retains its role. Siniscalchi (2011) develops this approach.

An alternative approach, explored in Battigalli, Di Tillio, and Samet (2011b), is to take as primitives the paths of play, rather than strategy profiles. In this case, at any history, player i chooses an action, given her beliefs about possible continuation paths. Notice that these paths include actions by i's opponents as well as actions that i herself takes. In this respect, such a model requires introspective beliefs about one's future play, in conflict with one of our key desiderata (Sec. 2.6.3). However, this approach does resolve the issue of verifiability of predictions, because these

[85] Proposition 7.1 in Brandenburger and Friedenberg (2010) implies that a similar relationship exists between Nash equilibria of perfect-information games and the epistemic conditions of lexicographic rationality and common assumption thereof, which we analyze in Sec. 8.

[86] For example, consider the game in Figure 2. RCSBR suggests that, on the one hand, player 1 believes that player 2 will choose d at his final decision node. On the other hand, if player 2 observes that player 1 has unexpectedly chosen A at her first move, then RCSBR within a complete type structure requires that player 2 explain this event by believing that player 1 believes that player 2 will choose a at his final decision node. In other words, RCSBR in a rich type structure implies that the equilibrium continuation strategies are *not* commonly believed at the third node.

[87] Reny (1992) introduces a notion of *explicable equilibrium* that is similarly motivated, though his analysis does not employ type structures and is thus closer to Pearce (1984)'s definition of extensive-form rationalizability.

[88] This is related to, but weaker than lexicographic expected utility maximization (Def. 32); for details, see Siniscalchi (2011).

are now observable paths of play and not strategy profiles.[89]

It also enables decomposing the assumption of sequential rationality into the assumptions that (i) the player *expects* her (future) actions to be optimal given her (future) beliefs, and (ii) her *actual* choices at a state coincide with her *planned* actions.[90] This more expressive language can be used to elegantly characterize backwards induction and should also be useful to study environments where players do not correctly forecast their own play (including cases where utility depends on beliefs and are hence not necessarily dynamically consistent).[91]

Incomplete information The definition of multistage games can be easily extended to incorporate payoff uncertainty. As for simultaneous-move games, we specify a set Θ of payoff states or parameters, and stipulate that each player's payoff function takes the form $u_i : \Theta \times S \to \mathbb{R}$. If one assumes that payoff states are not observed, the analysis in the preceding subsections requires only minimal changes.[92] First-order beliefs are modeled as CPSs on $\Theta \times S$, with i's conditioning events being $\mathcal{B}_{-i} = \{\Theta \times S_{-i}(h) : h \in \mathcal{H}\}$. In an epistemic type structure, the conditioning events are $\mathcal{C}_{-i} = \{\Theta \times S_{-i}(h) \times T_{-i} : h \in \mathcal{H}\}$ and the belief maps are defined as functions $\beta_i : T_i \to \Delta^{\mathcal{C}_{-i}}(\Theta \times S_{-i} \times T_{-i})$. Chen (2011) and Penta (2012) extend the notion of ICR to dynamic games. It is also straightforward to adapt the notion of Δ-rationalizability to allow for incomplete information; one obtains versions of initial or strong rationalizability that incorporate commonly-believed restrictions on first-order beliefs. Epistemic characterizations adapting Theorems 16 and 17 may be found in Battigalli and Siniscalchi (2007) and Battigalli and Prestipino (2011).

8 Admissibility

We now return to strategic-form analysis to analyze epistemic conditions and solution concepts related to *admissibility*, i.e., ruling out weakly dominated strategies. In particular, we will discuss epistemic conditions for *iterated admissibility*. This continues the analysis in Sec. 5: as noted therein, there is a conceptual inconsistency between the 'everything is possible' logic behind admissibility, and common-belief conditions. In Sec. 5 we introduced the notion of p-belief to resolve this inconsistency, and weakened the notion of common belief accordingly. This section explores an alternative approach: we replace probabilistic beliefs with the richer concept of a *lexicographic probability system* (LPS). These are related to the CPSs introduced in Section 7 to study extensive-form solution concepts; we elaborate on the connection in Section 8.4. We saw that common p-belief in rationality yields $S^\infty W$. We shall now see that suitable epistemic conditions characterize iterated admissibility (and its best-reply set analog, 'self-admissible sets'). The main idea (Brandenburger et al., 2008) is to introduce an analog to the notion of strong belief (Def. 27) for LPSs, called *assumption*.

A lexicographic probability system is a finite array $\mu_0, \ldots, \mu_K$ of probabilistic beliefs over, say, opponents' strategy profiles; μ_k is the *k-th level* of the LPS (distinct from a k-th *order* belief in a belief hierarchy). The lowest-level beliefs are the most salient, in the sense that, if a strategy s_i yields a strictly higher expected utility than another strategy s'_i with respect to μ_0, then s_i is preferred to s'_i. If, however, s_i and s'_i have the same μ_0–expected utility, then Player i computes μ_1–expected utilities, and so on. Thus, higher-level (less salient) probabilities are used to break

[89]A related model of conditional beliefs in dynamic games is considered by Di Tillio, Halpern, and Samet (2012).

[90]See also Bach and Heilmann (2011).

[91]The characterization is elegant in that it obtains backward induction by weakening the assumption of correct forecasting (which is a way to model "trembles").

[92]Even if there is private information, the changes required to adapt the analysis are only notational.

ties.[93] In order to formalize the notion of 'common assumption in lexicographic rationality,' we need to modify our notion of type structure: types will now be mapped to LPSs over opponents' strategies and types.

8.1 Basics

We begin by defining LPSs and lexicographic type spaces.

Definition 30 (Blume et al. 1991; Brandenburger et al. 2008) *A* ***lexicographic probability system*** *(or LPS) $\sigma = (\mu_0, \dots, \mu_{n-1})$ on a compact metric space Ω is a sequence of length $n < \infty$ in $\Delta(\Omega)$.*

An LPS $\sigma = (\mu_0, \dots, \mu_{n-1})$ has ***full support*** *if $\bigcup_\ell \operatorname{supp} \mu_\ell = \Omega$. The set of LPSs on Ω is denoted $\mathcal{L}(\Omega)$. The set of full-support LPSs on Ω is denoted $\mathcal{L}^+(\Omega)$.*

Definition 31 *A* ***lexicographic type structure*** *for the strategic-form game $G = (I, (S_i, u_i)_{i \in I})$ is $\mathcal{T} = (N, (T_i, \beta_i)_{i \in I})$ where each T_i is a compact metric space and each $\beta_i : T_i \to \mathcal{L}(S_{-i} \times T_{-i})$ is continuous.*

In order to define best replies, we first recall the lexicographic (i.e., "dictionary") order on vectors.

Definition 32 *Fix vectors $x = (x_\ell)_{\ell=0}^{n-1}, y = (y_\ell)_{\ell=0}^{n-1} \in \mathbb{R}^n$, write $x \geq^L y$ iff*

$$y_j > x_j \text{ implies } x_k > y_k \text{ for some } k < j.^{94} \tag{26}$$

Given a strategic-form game $G = (I, (S_i, u_i)_{i \in I})$, a strategy s_i of player i is a ***(lexicographic) best reply*** *to an LPS $\sigma_{-i} = (\mu_0, \dots, \mu_{n-1})$ on S_{-i} if $(\pi_i(s_i, \mu_\ell))_{\ell=0}^{n-1} \geq_L (\pi_i(s_i', \mu_\ell))_{\ell=0}^{n-1}$ for all $s_i' \in S_i$.*

It is easy to see that a strategy is admissible if and only if it is a lexicographic best reply to a *full-support* LPS.

Given a type structure, we define "rationality" as usual; we also define "full-support beliefs" analogously to Def. 16.

Definition 33 *Fix a lexicographic type structure $\mathcal{T} = (I, (T_i, \beta_i)_{i \in I})$ for the game G. The event that Player i is* rational *is*

$$R_i = \left\{ (s_i, t_i) \ : \ s_i \text{ is a lexicographic best reply to } \left(\operatorname{marg}_{S_{-i}} \mu_\ell \right)_{\ell=0}^{n-1}, \text{ where } \beta_i(t_i) = (\mu_0, \dots, \mu_{n-1}) \right\}.^{95} \tag{27}$$

The event that Player i has full-support beliefs *is*

$$FS_i = \left\{ (s_i, t_i) \ : \ \beta_i(t_i) \in \mathcal{L}^+(S_{-i} \times T_{-i}) \right\}.^{96} \tag{28}$$

[93] A behavioral characterization of lexicographic expected utility maximization is provided by Blume et al. (1991).

[94] That is: either $x = y$, or there exists $k \in \{0, \dots, n-1\}$ such that $x_j = y_j$ for $j = 0, \dots, k-1$, and $x_k > y_k$.

[95] As in Def. 24, the repeated use of R_i is a slight abuse of notation.

[96] Note that, in Def. 16, the event FS_i required full support of the beliefs over opponents' strategies only; here we follow Brandenburger et al. (2008) and require that the beliefs on strategies *and* types have full support. We do not know whether full-support first-order beliefs would be enough to obtain the results in this section.

8.2 Assumption and Mutual Assumption of Rationality

We can now introduce the notion of assumption.

Definition 34 *Fix a lexicographic type structure $\mathcal{T} = (I, (T_i, \beta_i)_{i \in I})$ and an event $E_{-i} \subset S_{-i} \times T_{-i}$. Then (s_i, t_i)* ***assumes*** *E_{-i}, written $(s_i, t_i) \in A_i(E_{-i})$, iff $\beta_i(t_i) = (\mu_0, \ldots, \mu_{n-1})$ has full support and there is $\ell^* \in \{0, \ldots, n-1\}$ such that:*

(i) $\mu_\ell(E_{-i}) = 1$ for $\ell \leq \ell^*$;

(ii) $E_{-i} \subseteq \bigcup_{\ell \leq \ell^*} \operatorname{supp} \mu_\ell$;[97]

(iii) for every $\ell > \ell^*$ there exist numbers $\alpha_1, \ldots, \alpha_{\ell^*} \in \mathbb{R}$ such that, for every event $F_{-i} \subseteq E_{-i}$ such that $\mu_\ell(F_{-i}) > 0$, $\mu_\ell(F_{-i}) = \sum_{k \leq \ell^*} \alpha_k \mu_k(F_{-i})$.

Assumption captures the notion that E_{-i} and all its subsets are infinitely more likely than the complement of E_{-i}. The level-zero measure must assign probability one to E_{-i}, although its support may be a strict subset of E_{-i}. If it is a strict subset, then the remainder of E_{-i} must receive probability one in the next level, and so on, until all of E_{-i} has been "covered." For those measures that assign positive probability outside E_{-i}, i.e. those after level ℓ^*, their restriction to E_{-i} is behaviorally irrelevant. To elaborate, in any LPS on a set Ω, a measure that is a linear combination of lower-level measures can be removed without changing lexicographic expected-utility rankings. Therefore, part (iii) of Definition 34 states that, at levels $\ell > \ell^*$, either $\mu_\ell(E_{-i}) = 0$, or $\mu_\ell(\cdot|E_{-i})$ is a linear combination of lower-level conditionals, and hence is irrelevant on E_{-i}. For example, if Ω consists of three points, the LPS given by $(\frac{1}{2}, \frac{1}{2}, 0)$, $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ will assume the event consisting of the first two points; the LPS given by $(\frac{1}{2}, \frac{1}{2}, 0)$, $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ will not.

Strong belief in an event also captures the notion that it is infinitely more likely than its complement; we discuss the connection between assumption and strong belief in Section 8.4. In view of this connection, it should come as no surprise that assumption also violates both Monotonicity and Conjunction (cf. Eq. (5)). As for strong belief, this implies that care must be taken when iterating the assumption operator. Furthermore, as for RCSBR, the behavioral implications of rationality and common assumption of rationality will not be monotonic with respect to the type structure. The discussion of these and related issues in Sec. 7.4.4 apply here *verbatim*.

We can now define the events "admissibility and mutual or common assumption thereof."

$$\begin{aligned} ACAA_i^0 &= R_i \cap FS_i; \\ ACAA_i^k &= ACAA_i^{k-1} \cap A_i(ACAA_{-i}^{k-1}) \quad \text{for } k > 0. \end{aligned} \tag{29}$$

The event that admissibility and common assumption of admissibility hold is $ACAA = \bigcap_{k \geq 0} ACAA^k$.

8.3 Characterization

Just like we need sufficiently rich type structures for RCSBR to yield forward induction (more precisely, extensive-form rationalizability: see Section 7.4.4), now we need sufficiently rich structures to obtain iterated admissibility from mutual or common assumption of admissibility. Adapting arguments from Brandenburger et al. (2008), one can readily show that there exists a complete lexicographic type structure.[98]

[97] Since the support of a measure is the smallest closed set with measure 1, this condition implies that the notion of 'assumption' depends upon the topology; see also Sec. 7.7.

[98] Specifically, one can adapt the proof of Proposition 7.2 in Brandenburger et al. (2008) (p. 341). Their argument goes further because they restrict attention to a subset of LPSs (see Sec. 8.4). We suspect, but have not proved, that a *canonical* construction à la Mertens and Zamir (1985) or Brandenburger and Dekel (1993) is also possible for LPSs. Ganguli and Heifetz (2012) show how to construct a non-topological "universal" type structure for LPSs, such that every other such LPS-based type structure can be uniquely embedded in it.

We recall the definitions of admissibility with respect to a Cartesian product of strategy sets and iterated admissibility. We then introduce a suitable analog of best-reply sets. As in Brandenburger et al. (2008), we restrict attention to two-player games.

Definition 35 *Fix $B_1 \times B_2 \subset S_1 \times S_2$. An action $s_i \in B_i$ is* ***weakly dominated with respect to*** *$B_1 \times B_2$ if there is $\mu_i \in \Delta(B_i)$ such that $u_i(\mu_i, s_{-i}) \geq u_i(s_i, s_{-i})$ for all $s_{-i} \in B_{-i}$, and $u_i(\mu_i, s^*_{-i}) > u_i(s_i, s^*_{-i})$ for some $s^*_{-i} \in B_{-i}$. The action $s_i \in B_i$ is* ***admissible with respect to*** *$B_1 \times B_2$ if it is not weakly dominated with respect to $B_1 \times B_2$.*

Definition 36 (Iterated Admissibility) *Fix a two-player strategic-form game $(I, (S_i, u_i)_{i \in I})$. For every player $i \in I$, let $W^0_i = S_i$. For $k > 0$, let $s_i \in W^k_i$ iff $s_i \in W^{k-1}_i$ and s_i is admissible w.r.to $W^{k-1}_1 \times W^{k-1}_2$. The set of* ***iteratively admissible*** *strategies is W^∞.*

We need an additional definition. Say that a strategy $s'_i \in S_i$ of player i *supports* $s_i \in S_i$ if there exists a mixed strategy $\sigma_i \in S_i$ for i that duplicates s_i and has s'_i in its support: that is, $u_i(\sigma_i, s_{-i}) = u_i(s_i, s_{-i})$ for all $s_{-i} \in S_{-i}$, and $\sigma_i(s'_i) > 0$.

Definition 37 *Fix a two-player strategic-form game $(I, (S_i, u_i)_{i \in I})$. A set $B = \prod_{i \in I} B_i \subset S$ is a* ***self-admissible set*** *(or SAS) if, for every player $i \in I$, every $s_i \in B_i$ is admissible with respect to both $S_i \times S_{-i}$ and $S_i \times B_{-i}$;[99] it is a* ***full SAS*** *if, in addition, for every player $i \in I$ and strategy $s_i \in B_i$, if s'_i supports s_i then $s'_i \in B_i$.*

In the definition of full SAS, including in the set B_i a strategy s'_i that supports some other strategy $s_i \in B_i$ plays the same role as including all best replies to a belief that justifies some element of a full BRS. For additional discussion, see Brandenburger et al. (2008).

As is the case for extensive-form rationalizability and EFBRSs, the set W^∞ is a full SAS; however, it is not the largest (full) SAS, and indeed there may be games in which a full SAS is disjoint from the IA set. For example, in the strategic form of the game in Fig. 3, the unique IA profile is (InT, L); however, $B = \{OutT, OutB\} \times \{R\}$ is also a full SAS.

The characterization result is as follows.

Theorem 18 *Fix a two-person game $G = (I, (S_i, u_i)_{i \in I})$.*

1. *In any lexicographic type structure $(I, (S_i, T_i, \beta_i)_{i \in I})$ for G, $\mathrm{proj}_S ACAA$ is a full SAS.*

2. *In any complete lexicographic type structure $(I, (S_i, T_i, \beta_i)_{i \in I})$ for G, and for every $k \geq 0$, $\mathrm{proj}_S ACAA^k = W^{k+1}$.*

3. *For every full SAS B, there exists a finite lexicographic type structure $(I, (S_i, T_i, \beta_i)_{i \in I})$ for G such that $\mathrm{proj}_S ACAA = B$.*

[99] To see why we need admissibility with respect to both $S_i \times B_{-i}$ and $S_i \times S_{-i}$, consider the following two-person games (only Player 1's payoffs are indicated).

	L	R
T	0	0
B	0	1

	L	R
T	0	1
B	1	0

In the game on the left, T is admissible with respect to $S_1 \times \{L\}$, but not with respect to $S_1 \times S_2$. On the other hand, in the game on the right, T is admissible with respect to $S_1 \times S_2$, but not with respect to $S_1 \times \{L\}$.

8.4 Discussion

We start by discussing three issues in the characterization of IA. These are the relationship to the characterization in Brandenburger et al. (2008), the full-support assumption, and common vs. mutual assumption of admissibility. We then discuss the relationship between the current section and the extensive-form analysis of Section 7. In particular, we relate LPSs to CPSs, assumption to strong belief, and admissibility to sequential rationality.

Before turning to these issues, we note that, as is the case for strong belief, since assumption violates Monotonicity and Conjunction, its interpretation in small type structures is somewhat delicate. We do not repeat the discussion of these issues here, as the treatment in Sec. 7.4.4 regarding RCSBR applies verbatim here.

8.4.1 Issues in the characterization of IA

Relationship with Brandenburger et al. (2008) Our presentation differs from Brandenburger et al. (2008) in that their main results are stated for LPSs with disjoint supports; following Blume et al. (1991), we call these "lexicographic *conditional* probability systems," or LCPSs. We choose to work with LPSs to avoid certain technical complications that arise with LCPSs (for example, the definition and construction of a complete type structure). The proof of Theorem 18 can be found in Dekel, Friedenberg, and Siniscalchi (2013a).

Full-support beliefs The characterization of IA focuses on types that commonly assume rationality and full-support beliefs. This raises the question whether one could incorporate the full-support assumption in the definition of lexicographic type structures. That is, could we assume that *all* types have full-support beliefs, or at least full-support first-order beliefs? Recall that, in the characterization of $S^{\infty}W$ in Section 5, we also focus on types that commonly p-believe in both rationality and full-support beliefs. There, we could restrict attention to type structures where each type's belief over the opponents' strategies have full support. The following example demonstrates that we cannot do this in the current environment.

Example 10 (Figure 2.6 in Brandenburger et al. (2008), attributed to P. Battigalli) Consider the strategic-form game in Table 14. The IA set is $\{U, M, D\} \times \{C, R\}$.

	L	C	R
U	4,0	4,1	0,1
M	0,0	0,1	4,1
D	3,0	2,1	2,1

Table 14: Iterated admissibility and ACAA

Fix an arbitrary lexicographic type structure. Note first that, since L is strictly dominated for player 2, $(L, t_2) \notin R_2$ for any type t_2 of 2; a fortiori, $(L, t_2) \notin R_2 \cap FS_2$. Moreover, C and R always yield a payoff of 1, and hence both $(C, t_2) \in R_2 \cap FS_2$ and $(R, t_2) \in R_2 \cap FS_2$ hold if and only if type t_2 has full-support beliefs.

Now consider a type t_1 of player 1 such that $(D, t_1) \in R_1 \cap FS_1 \cap A_1(R_2 \cap FS_2)$, and let $\beta_1(t_1) = (\mu_0, \ldots, \mu_{n-1})$. Since the definition of assumption (Def. 34) requires full-support beliefs, as t_1 assumes $R_2 \cap FS_2$, this type must have full-support beliefs; in particular, there must be an order k with $\mu_k(\{L\} \times T_2) > 0$. Furthermore, since t_1 assumes $R_2 \cap FS_2$, and L is irrational for 2, it must be the case that $k > 0$.

Next, by lexicographic utility maximization, for all $\ell = 0, \ldots, k-1$, we must have $\mu_\ell(\{C\}\times T_2) = \mu_\ell(\{R\} \times T_2) = \frac{1}{2}$ for $0 \leq \ell \leq k-1$: otherwise, D could not be a best reply. But then, U and M are also best replies to $\mathrm{marg}_{S_2}\mu_\ell$, $\ell = 0, \ldots, k-1$. In other words, D ties with U and M against the beliefs $\mathrm{marg}_{S_2}\mu_0, \ldots, \mathrm{marg}_{S_2}\mu_{k-1}$. Then, the optimality of D requires that D also be a best response to the k-th level belief $\mathrm{marg}_{S_2}\mu_k$.

Finally, for this to be the case, we must have $\mu_k(\{R\}\times T_2) > 0$. Moreover, as $\mu_k(\{L\}\times T_2) > 0$ and L is not rational for 2, $\mu_k(R_2) < 1$, hence $\mu_k(R_2 \cap FS_2) < 1$. However, t_1 assumes $R_2 \cap FS_2$. Therefore, by the definition of assumption, $\mu_k(R_2 \cap FS_2)$ must equal either 1 or 0. Hence, it must be the case that $\mu_k(R_2 \cap FS_2) = 0$. On the other hand, $\mu_k(\{R\} \times T_2) > 0$, so there must be types t_2 of 2 for whom $(R, t_2) \notin R_2 \cap FS_2$. That is, because 1's kth-level belief assigns zero probability to 2 being rational and having full-support beliefs, and positive probability to 2 playing R, it must be that 1 expects those types of 2 who are playing R to hold beliefs that either do not rationalize R, or do not have full support. But, since R is a best reply against any beliefs, the only way this can hold is if 1 expects 2's type to not have full-support beliefs. This means that the type structure under consideration *must* contain types for player 2 that do not have full-support beliefs.

Common vs. mutual assumption of admissibility Finally, there is an additional subtlety. Note that Theorem 18 does not characterize *common* assumption of admissibility for complete type structures–merely finite-order assumption of admissibility. Indeed, Brandenburger et al. (2008) show that, under completeness and restricting attention to LCPSs, $\bigcap_{k\geq 0} ACAA_i^k$ is empty. Admissibility and common assumption of admissibility thus cannot hold in a complete, LCPS-based type structure. We believe (but have not proved) that the same is true when beliefs are represented by LPSs.

This is a puzzling result. In a recent paper, Lee and Keisler (2011) demonstrate that the problem arises out of the requirement in Def. 31 that the belief maps β_i be *continuous*. If one drops this requirement, and merely asks that they be measurable, it is possible to construct a complete, LCPS-based type structure in which $\mathrm{proj}_S ACAA$ equals IA, so that ACAA *is* possible (and characterizes iterated admissibility).[100]

8.4.2 Extensive-form analysis and strategic-form refinements

LPSs and CPSs LPSs and CPSs are clearly similar. CPSs are also collections of probabilities, that also may differ in terms of saliency (lower-saliency beliefs come into play as unexpected events are encountered). However, there are also differences, due to the fact that the former are strategic-form objects whereas the latter are defined for extensive-form games.[101] Probabilities in an LPS are completely ordered, whereas in a CPS the order is partial. For example, consider a game in which Player 1 can choose T, M or B, and Player 2 (who moves immediately after 1) is initially certain of T. Then, Player 2's conditional beliefs following M and B are not ranked in terms of their salience, although they are less salient than Player 2's initial belief. Second, the supports of any two probabilities in a CPS are either disjoint, or one is included in the other; in an LPS, the supports can overlap arbitrarily. In addition, a technical distinction in the context of type structures is that, for a finite extensive game, the number of probabilities in a CPS is fixed and equal to the number of non-terminal histories in the game; on the other hand, in general there is no upper bound on the number of levels in an LPS.

[100] Other papers that provide epistemic conditions related to IA include Asheim and Dufwenberg (2003), Barelli and Galanis (2011), Yang (2011) and Perea (2012).

[101] In fact, CPSs are no different from regular probabilities for extensive forms of simultaneous-move games.

Strong belief and assumption As we noted above, both strong belief and assumption capture the notion that an event and its subsets are infinitely more likely than its complement. Recall that player i assumes E_{-i} if she assigns probability one to it or some subset of it in each of the first ℓ^* levels of her LPS, until all of E_{-i} has been given probability 1 at some level; furthermore, higher-level measures either assign probability zero to E_{-i}, or are behaviorally irrelevant conditional on E_{-i}. Analogously, if player i strongly believes E_{-i} in an extensive game, then her initial beliefs assign probability one to E_{-i} or some subset thereof. Moreover, so long as E_{-i} has not been contradicted by observed play, when player i revises her beliefs,[102] she continues to assign probability one to some subset of E_{-i}. Once E_{-i} has been contradicted, it must receive probability zero. Thus, with strong belief, the 'level' at which E_{-i} is no longer believed is objective, while in the case of assumption, the level at which i no longer believes E_{-i} is subjective. Nevertheless, assumption and strong belief are quite similar. Specifically, for finite spaces Ω, there is a one-to-one mapping between LCPSs (but not arbitrary LPSs) and CPSs in which the set of conditioning events consists of all non-empty subsets of Ω. Furthermore, an LCPS λ 'assumes' an event E_{-i} (analogously to Def. 34) if and only if the corresponding CPS μ 'strongly believes' E_{-i}.[103]

Admissibility and sequential rationality Brandenburger (2007) shows that, in single-person, dynamic choice problems, admissibility is equivalent to sequential rationality in all decision trees that have the same strategic form, up to the addition or deletion of strategies that are convex combinations of other strategies (i.e., trees that have the same *fully reduced* normal form in the sense of Kohlberg and Mertens, 1986). Nevertheless, Brandeburger's result is about single-person problems; adding or deleting convex combinations of existing strategies in an extensive *game* may affect the players' strategic reasoning (see e.g. Hillas (1994) and Govindan and Wilson (2009)).

References

D.S. Ahn. Hierarchies of ambiguous beliefs. *Journal of Economic Theory*, 136(1):286–301, 2007.

C.D. Aliprantis and K.C. Border. *Infinite dimensional analysis: a hitchhiker's guide*. Springer Verlag, 2007.

F.J. Anscombe and R.J. Aumann. A definition of subjective probability. *The annals of mathematical statistics*, 34(1):199–205, 1963.

W. Armbruster and W. Böge. Bayesian game theory. *Game theory and related topics, North-Holland, Amsterdam*, 1979.

G.B. Asheim and M. Dufwenberg. Admissibility and common belief. *Games and Economic Behavior*, 42(2):208–234, 2003.

G.B. Asheim and A. Perea. Sequential and quasi-perfect rationalizability in extensive games. *Games and Economic Behavior*, 53(1):15–42, 2005.

Geir B. Asheim. Proper rationalizability in lexicographic beliefs. *International Journal of Game Theory*, 30:453–478, 2002. ISSN 0020-7276. URL http://dx.doi.org/10.1007/s001820200090. 10.1007/s001820200090.

[102]When we say that she "revises" her beliefs, we allow for both Bayesian belief updating, following positive-probability observations, as well as formulating entirely new beliefs, following zero-probability observations.

[103]For the case of infinite sets Ω, see Brandenburger, Friedenberg, and Keisler (2007).

Geir B. Asheim, M. Voorneveld, J. W Weibull, et al. Epistemically stable strategy sets. *Memorandum*, 2009.

R. Aumann and A. Brandenburger. Epistemic conditions for Nash equilibrium. *Econometrica: Journal of the Econometric Society*, 63(5):1161–1180, 1995.

R.J. Aumann. Agreeing to disagree. *The annals of statistics*, 4(6):1236–1239, 1976.

R.J. Aumann. Correlated equilibrium as an expression of Bayesian rationality. *Econometrica*, 55 (1):1–18, 1987. ISSN 0012-9682.

R.J. Aumann. Backward induction and common knowledge of rationality. *Games and Economic Behavior*, 8(1):6–19, 1995. ISSN 0899-8256.

R.J. Aumann. On the centipede game. *Games and Economic Behavior*, 23:97–105, 1998.

R.J. Aumann. Interactive epistemology i: knowledge. *International Journal of Game Theory*, 28 (3):263–300, 1999a.

R.J. Aumann. Interactive epistemology ii: Probability. *International Journal of Game Theory*, 28 (3):301–314, 1999b.

R.J. Aumann and J.H. Dreze. Assessing strategic risk. mimeo, 2004.

R.J. Aumann and J.H. Dreze. Assessing strategic risk. *American Economic Journal: Microeconomics*, 1(1):1–16, 2009.

Christian W. Bach and C. Heilmann. Agent connectedness and backward induction. *International Game Theory Review*, 13(02):195–208, 2011.

D. Balkenborg and E. Winter. A necessary and sufficient epistemic condition for playing backward induction. *Journal of Mathematical Economics*, 27(3):325–345, 1997. ISSN 0304-4068.

P. Barelli. Consistency of beliefs and epistemic conditions for nash and correlated equilibria. *Games and Economic Behavior*, 67(2):363–375, 2009.

P. Barelli. Consistent beliefs in extensive form games. *Games*, 1(4):415–421, 2010.

Paulo Barelli and Spyros Galanis. Admissibility and event-rationality. 2011.

K. Basu and J.W. Weibull. Strategy subsets closed under rational behavior. *Economics Letters*, 36 (2):141–146, 1991.

P. Battigalli. Strategic rationality orderings and the best rationalization principle. *Games and Economic Behavior*, 13(2):178–200, 1996a.

P. Battigalli. On rationalizability in extensive games. *Journal of Economic Theory*, 74(1):40–61, 1997.

P. Battigalli and A. Friedenberg. Forward induction reasoning revisited. *Theoretical Economics*, 2010. ISSN 1555-7561.

P. Battigalli and A. Prestipino. Transparent restrictions on beliefs and forward induction reasoning in games with asymmetric information. Technical report, 2011.

P. Battigalli and M. Siniscalchi. Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games. *Journal of Economic Theory*, 88(1):188–230, 1999.

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

P. Battigalli and M. Siniscalchi. Rationalization and incomplete information. *Advances in Theoretical Economics*, 3(1):1073–1073, 2003.

P. Battigalli and M. Siniscalchi. Interactive epistemology in games with payoff uncertainty. *Research in Economics*, 61(4):165–184, 2007.

P. Battigalli, A. Di Tillio, E. Grillo, and A. Penta. Interactive epistemology and solution concepts for games with asymmetric information. working paper, Università Bocconi, 2010.

P. Battigalli, A. Di Tillio, E. Grillo, and A. Penta. Interactive epistemology and solution concepts for games with asymmetric information. *The BE Journal of Theoretical Economics*, 11(1), 2011a.

P. Battigalli, A. Di Tillio, and D. Samet. Strategies and interactive beliefs in dynamic games. 2011b.

Pierpaolo Battigalli. Comportamento razionale ed equilibrio nei giochi e nelle situazioni sociali. *unpublished undergraduate dissertation, Bocconi University, Milano*, 1987.

Pierpaolo Battigalli. Comment to on the decision theoretic foundation of game theory by m. mariotti. In K.J. Arrow, E. Colombatto, M. Perlman, and C. Schmidt, editors, *The Foundations of Rational Economic Behavior*. London: McMillan, 1996b.

E. Ben-Porath. Rationality, Nash equilibrium and backwards induction in perfect-information games. *The Review of Economic Studies*, pages 23–46, 1997.

D. Bergemann and S. Morris. Correlated equilibrium in games with incomplete information. *Cowles Foundation Discussion Papers*, (1822), 2011.

B.D. Bernheim. Rationalizable strategic behavior. *Econometrica: Journal of the Econometric Society*, 52(4):1007–1028, 1984. ISSN 0012-9682.

P. Billingsley. *Probability and measure*. Wiley-India, 2008.

L. Blume, A. Brandenburger, and E. Dekel. Lexicographic probabilities and choice under uncertainty. *Econometrica: Journal of the Econometric Society*, 59(1):61–79, 1991.

W. Böge and T. Eisele. On solutions of bayesian games. *International Journal of Game Theory*, 8 (4):193–215, 1979.

G. Bonanno and K. Nehring. How to make sense of the common prior assumption under incomplete information. *International Journal of Game Theory*, 28(3):409–434, 1999.

T. Börgers. Pure strategy dominance. *Econometrica: Journal of the Econometric Society*, 61(2): 423–430, 1993. ISSN 0012-9682.

Tilman Borgers. Weak dominance and approximate common knowledge. *Journal of Economic Theory*, 64(1):265–276, 1994. URL http://econpapers.repec.org/RePEc:eee:jetheo:v:64:y:1994:i:1:p:265-276.

A. Brandenburger. The power of paradox: some recent developments in interactive epistemology. *International Journal of Game Theory*, 35(4):465–492, 2007.

A. Brandenburger and E. Dekel. Rationalizability and correlated equilibria. *Econometrica*, pages 1391–1402, 1987a.

A. Brandenburger and E. Dekel. Rationalizability and correlated equilibria. *Econometrica*, 55(6): 1391–1402, 1987b. ISSN 0012-9682.

A. Brandenburger and E. Dekel. Hierarchies of beliefs and common knowledge. *Journal of Economic Theory*, 59:189–189, 1993.

A. Brandenburger and A. Friedenberg. Intrinsic correlation in games. *Journal of Economic Theory*, 141(1):28–67, 2008.

A. Brandenburger and A. Friedenberg. Self-admissible sets. *Journal of Economic Theory*, 145(2): 785–811, 2010. ISSN 0022-0531.

A. Brandenburger, A. Friedenberg, and H.J. Keisler. Admissibility in games. *Econometrica*, 76(2): 307, 2008.

Adam Brandenburger. Lexicographic probabilities and iterated admissibility. In Partha Dasgupta, Douglas Gale, Oliver Hart, and Eric Maskin, editors, *Economic Analysis of Markets and Games*, pages 282–290. MIT Press, 1992.

Adam Brandenburger and Eddie Dekel. Tba. 1986.

Adam Brandenburger, Amanda Friedenberg, and Jerome Keisler. Notes on the relationship between strong belief and assumption. mimeo, 2007.

H. Carlsson and E. Van Damme. Global games and equilibrium selection. *Econometrica: Journal of the Econometric Society*, 61(5):989–1018, 1993.

Y-C. Chen. A Structure Theorem for Rationalizability in Dynamic Games. working paper, National University of Singapore, 2011.

Y.C. Chen. Universality of the epstein-wang type structure. *Games and Economic Behavior*, 68 (1):389–402, 2010.

Y.C. Chen, N.V. Long, and X. Luo. Iterated strict dominance in general games. *Games and Economic Behavior*, 61(2):299–315, 2007.

I.K. Cho and D.M. Kreps. Signaling games and stable equilibria. *The Quarterly Journal of Economics*, 102(2):179, 1987.

Vincent P Crawford, Miguel A Costa-Gomes, and Nagore Iriberri. Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature*, 2012.

Bruno De Finetti. *Foresight: Its logical laws, its subjective sources*, volume Breakthroughs in Statistics: Foundations and Basic Theory, pages 134–174. Springer-Verlag, 1992.

E. Dekel and D. Fudenberg. Rational behavior with payoff uncertainty. *Journal of Economic Theory*, 52(2):243–267, 1990. ISSN 0022-0531.

E. Dekel, D. Fudenberg, and D.K. Levine. Payoff information and self-confirming equilibrium. *Journal of Economic Theory*, 89(2):165–185, 1999.

E. Dekel, D. Fudenberg, and D.K. Levine. Learning to play bayesian games. *Games and Economic Behavior*, 46(2):282–303, 2004.

E. Dekel, D. Fudenberg, and S. Morris. Topologies on types. *Theoretical Economics*, 1:275–309, 2006.

E. Dekel, D. Fudenberg, and S. Morris. Interim correlated rationalizability. *Theoretical Economics*, 2(1):15–40, 2007.

Eddie Dekel, Amanda Friedenberg, and Marciano Siniscalchi. Admissibility revisited. in progress, 2013a.

Eddie Dekel, Luciano Pomatto, and Marciano Siniscalchi. Epistemic game theory: Online appendix. 2013b.

C. Dellacherie and PA Meyer. Probability and potential. *Paris: Hermann*, 1978.

A. Di Tillio. Subjective expected utility in games. *Theoretical Economics*, 3(3):287–323, 2008.

Alfredo Di Tillio, Joseph Halpern, and Dov Samet. Conditional belief types. January 2012.

Songzi Du. Correlated equilibrium and higher order beliefs about play. 2011.

M. Dufwenberg and M. Stegeman. Existence and uniqueness of maximal reductions under iterated strict dominance. *Econometrica*, 70(5):2007–2023, 2002.

J.C. Ely and M. Peski. Hierarchies of belief and interim rationalizability. *Theoretical Economics*, 1(1):19–65, 2006.

L.G. Epstein and T. Wang. " Beliefs about beliefs" without probabilities. *Econometrica*, 64(6): 1343–1373, 1996. ISSN 0012-9682.

R. Fagin and J.Y. Halpern. Reasoning about knowledge and probability. *Journal of the ACM (JACM)*, 41(2):340–367, 1994.

R. Fagin, J.Y. Halpern, and N. Megiddo. A logic for reasoning about probabilities. *Information and computation*, 87(1):78–128, 1990.

R. Fagin, J. Halpern, Y. Moses, and M. Vardi. Reasoning about knowledge. *Cambridge, Massachusetts: The MIT Press*, 1995.

Y. Feinberg. Characterizing common priors in the form of posteriors* 1. *Journal of Economic Theory*, 91(2):127–179, 2000.

F. Forges. Five legitimate definitions of correlated equilibrium in games with incomplete information. *Theory and decision*, 35(3):277–310, 1993.

F. Forges. Correlated equilibrium in games with incomplete information revisited. *Theory and decision*, 61(4):329–344, 2006.

A. Friedenberg. When do type structures contain all hierarchies of beliefs? *Games and Economic Behavior*, 68(1):108–129, 2010.

D. Fudenberg and Y. Kamada. Rationalizable partition-confirmed equilibrium. Technical report, mimeo, 2011.

D. Fudenberg and D.K. Levine. Self-confirming equilibrium. *Econometrica*, 61(3):523–545, 1993.

D. Fudenberg and J. Tirole. *Game theory*. MIT Press, 1991a.

D. Fudenberg and J. Tirole. Perfect bayesian equilibrium and sequential equilibrium. *Journal of Economic Theory*, 53(2):236–260, 1991b.

D. Fudenberg, D.M. Kreps, and D.K. Levine. On the robustness of equilibrium refinements. *Journal of Economic Theory*, 44(2):354–380, 1988.

Drew Fudenberg and Yuichiro Kamada. Rationalizable partition-confirmed equilibrium. working paper, 2012.

D. Gale and S. Sherman. Solutions of finite two-person games. In H. Kuhn and A. Tucker, editors, *Contributions to the theory of games*. Princeton University Press, 1950.

Jayant Ganguli and Aviad Heifetz. Universal interactive preferences. *Available at SSRN 2174371*, 2012.

J. Geanakoplos, D. Pearce, and E. Stacchetti. Psychological games and sequential rationality. *Games and Economic Behavior*, 1(1):60–79, 1989.

S. Govindan and R. Wilson. On forward induction. *Econometrica*, 77(1):1–28, 2009. ISSN 1468-0262.

F. Gul. Rationality and coherent theories of strategic behavior. *Journal of Economic Theory*, 70 (1):1–31, 1996.

Faruk Gul and Wolfgang Pesendorfer. Interdependent preference models as a theory of intentions. Technical report, Princeton University Working Paper, 2010.

J.C. Harsanyi. Games with incomplete information played by "Bayesian" players, I-III. Part I. The basic model. *Management Science*, pages 159–182, 1967.

John C Harsanyi. Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory*, 2(1):1–23, 1973.

A. Heifetz. The positive foundation of the common prior assumption. *Games and Economic Behavior*, 56(1):105–120, 2006.

A. Heifetz and P. Mongin. Probability logic for type spaces. *Games and economic behavior*, 35 (1-2):31–53, 2001.

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

Aviad Heifetz. The bayesian formulation of incomplete information the non-compact case. *International Journal of Game Theory*, 21:329–338, 1993. ISSN 0020-7276. URL http://dx.doi.org/10.1007/BF01240148. 10.1007/BF01240148.

Aviad Heifetz and Andrés Perea. On the outcome equivalence of backward induction and extensive form rationalizability. mimeo, Maastricht University, 2013.

John Hillas. How much of forward induction is implied by backward induction and ordinality. mimeo, University of Auckland, 1994.

T.W. Hu. On p-rationalizability and approximate common certainty of rationality. *Journal of Economic Theory*, 136(1):379–391, 2007.

A. Kajii and S. Morris. The robustness of equilibria to incomplete information. *Econometrica: Journal of the Econometric Society*, pages 1283–1309, 1997a.

Atsushi Kajii and Stephen Morris. Refinements and higher order beliefs: A unified survey. Northwestern University CMS-EMS discussion paper n. 1197, 1997b. URL http://www.kellogg.northwestern.edu/research/math/papers/1197.pdf.

Willemien Kets. Bounded reasoning and higher-order uncertainty. *Available at SSRN 2116626*, 2012.

E. Kohlberg and J.F. Mertens. On the strategic stability of equilibria. *Econometrica: Journal of the Econometric Society*, 54(5):1003–1037, 1986.

D.M. Kreps and R. Wilson. Sequential equilibria. *Econometrica: Journal of the Econometric Society*, 50(4):863–894, 1982.

Byung-Soo. Lee and H. J. Keisler. Common Assumption of Rationality. working paper, University of Toronto, Rotman School of Management, 2011.

Q. Liu. On redundant types and Bayesian formulation of incomplete information. *Journal of Economic Theory*, 2009.

Q. Liu. Higher-order beliefs and epistemic conditions for nash equilibrium. 2010.

Qingmin Liu. Correlation and common prior in games with incomplete information. August 2011.

M. Mariotti. Is bayesian rationality compatible with strategic rationality? *The Economic Journal*, pages 1099–1109, 1995.

L.M. Marx and J.M. Swinkels. Order independence for iterated weak dominance. *Games and Economic Behavior*, 18(2):219–245, 1997.

J.F. Mertens and S. Zamir. Formulation of Bayesian analysis for games with incomplete information. *International Journal of Game Theory*, 14(1):1–29, 1985.

P. Milgrom and N. Stokey. Information, trade and common knowledge. *Journal of Economic Theory*, 26(1):17–27, 1982.

D. Monderer and D. Samet. Approximating common knowledge with common beliefs. *Games and Economic Behavior*, 1(2):170–190, 1989.

S. Morris. Trade with heterogeneous prior beliefs and asymmetric information. *Econometrica: Journal of the Econometric Society*, pages 1327–1347, 1994.

S. Morris. *Alternative definitions of knowledge*, volume Epistemic Logic and the Theory of Games and Decisions, pages 217–233. Kluwer, 1997.

S. Morris. Typical types. *mimeo, Department of Economics. Princeton University.[448, 449, 469]*, 2002.

S. Morris and H.S. Shin. Global games: theory and applications. In *advances in Economics and Econometrics: theory and applications, Eighth world Congress*, volume 1, pages 56–114, 2003.

Stephen E. Morris and Satoru Takahashi. Common Certainty of Rationality Revisited. *SSRN eLibrary*, 2011. doi: 10.2139/ssrn.1789773.

R.B. Myerson. Refinements of the nash equilibrium concept. *International Journal of Game Theory*, 7(2):73–80, 1978.

R.B. Myerson. *Game theory: analysis of conflict*. Harvard Univ Pr, 1997.

R.F. Nau and K.F. McCardle. Coherent behavior in noncooperative games. *Journal of Economic Theory*, 50(2):424–444, 1990. ISSN 0022-0531.

M.J. Osborne and A. Rubinstein. *A course in game theory*. The MIT press, 1994.

D.G. Pearce. Rationalizable strategic behavior and the problem of perfection. *Econometrica: Journal of the Econometric Society*, 52(4):1029–1050, 1984. ISSN 0012-9682.

Antonio Penta. Higher order uncertainty and information: Static and dynamic games. *Econometrica*, 80(2):631–660, 2012.

A. Perea. A one-person doxastic characterization of nash strategies. *Synthese*, 158(2):251–271, 2007.

A. Perea. Minimal belief revision leads to backward induction. *Mathematical Social Sciences*, 56 (1):1–26, 2008.

Andrés Perea. An algorithm for proper rationalizability. *Games and Economic Behavior*, 72(2): 510–525, 2011a.

Andres Perea. Belief in the opponents' future rationality. mimeo, 2011b.

Andres Perea. *Epistemic Game Theory: reasoning and choice*. Cambridge University Press, 2012.

Alex Peysakhovich. Correlation without signals. Technical report, 2011.

Luciano Pomatto. Epistemic conditions for bayesian nash equilibrium. mimeo, Northwestern University, 2011.

Matthew Rabin. Incorporating behavioral assumptions into game theory. In James W. Friedman, Warren J. Samuels, and William Darity, editors, *Problems of Coordination in Economic Activity*, volume 35 of *Recent Economic Thought Series*, pages 69–87. Springer Netherlands, 1994. ISBN 978-94-011-1398-4.

P.J. Reny. Backward induction, normal form perfection and explicable equilibria. *Econometrica: Journal of the Econometric Society*, pages 627–649, 1992.

P.J. Reny. Common belief and the theory of games with perfect information. *Journal of Economic Theory*, 59:257–257, 1993.

A. Rényi. On a new axiomatic theory of probability. *Acta Mathematica Hungarica*, 6(3):285–335, 1955. ISSN 0236-5294.

A. Rubinstein. The Electronic Mail Game: Strategic Behavior Under "Almost Common Knowledge". *The American Economic Review*, 79(3):385–391, 1989.

A. Rubinstein. Comments on the interpretation of game theory. *Econometrica: Journal of the Econometric Society*, pages 909–924, 1991.

A. Rubinstein and A. Wolinsky. On the logic of 'agreeing to disagree' type results. *Journal of Economic Theory*, 51(1):184–193, 1990.

Ariel Rubinstein and Asher Wolinsky. Rationalizable conjectural equilibrium: between nash and rationalizability. *Games and Economic Behavior*, 6(2):299–311, 1994.

T. Sadzik. Beliefs revealed in bayesian-nash equilibrium. *New York University*, 2011.

D. Samet. Hypothetical knowledge and games with perfect of of information. *Games and economic behavior*, 17:230–251, 1996.

D. Samet. Common priors and separation of convex sets. *Games and economic behavior*, 24: 172–174, 1998a.

D. Samet. Iterated expectations and common priors. *Games and Economic Behavior*, 24:131–141, 1998b.

Dov Samet. Common belief of rationality in games of perfect information. *Games and Economic Behavior*, (0):–, 2013. ISSN 0899-8256. doi: 10.1016/j.geb.2013.01.008. URL http://www.sciencedirect.com/science/article/pii/S0899825613000195.

L. Samuelson. Dominated strategies and common knowledge. *Games and Economic Behavior*, 4 (2):284–313, 1992.

L.J. Savage. *The foundations of statistics*. Dover Pubns, 1972.

Burkhard Schipper. The unawareness bibliography, 2013. URL http://www.econ.ucdavis.edu/faculty/schipper/unaw.htm.

F. Schuhmacher. Proper rationalizability and backward induction. *International Journal of Game Theory*, 28(4):599–615, 1999. ISSN 0020-7276.

R. Selten. Reexamination of the perfectness concept for equilibrium points in extensive games. *International journal of game theory*, 4(1):25–55, 1975.

Marciano Siniscalchi. A revealed-preference theory of sequential rationality. mimeo, Northwestern University, November 2011.

R. Stalnaker. Belief revision in games: forward and backward induction. *Mathematical Social Sciences*, 36(1):31–56, 1998.

T.C.C. Tan and S.R. da Costa Werlang. The Bayesian foundations of solution concepts of games. *Journal of Economic Theory*, 45(2):370–391, 1988. ISSN 0022-0531.

Q. Tang. Interim partially correlated rationalizability. 2011.

O. Tercieux. p-best response set. *Journal of Economic Theory*, 131(1):45–70, 2006.

Eric Van Damme. *Refinements of the Nash equilibrium concept*. Springer-Verlag, 1983.

J. Weinstein and M. Yildiz. A structure theorem for rationalizability with application to robust predictions of refinements. *Econometrica*, 75(2):365–400, 2007. ISSN 1468-0262.

Chih-Chun Yang. Weak assumption and iterative admissibility. 2011.