Comprehensive Rationalizability[*]

Aviad Heifetz[†]   Martin Meier[‡]   Burkhard C. Schipper[§]

February 25, 2019

Abstract

We present a new solution concept for strategic games called *comprehensive rationalizability* that embodies a new version of "common cautious belief in rationality" based on a sound epistemic characterization in a universal type space. Differently from existing iterative strategy elimination procedures in the literature, it should rather be viewed as an iterative strategy *demotion* procedure as it requires some memory of previously discarded strategies. It refines rationalizability, but it neither refines nor is refined by iterated admissibility. Nevertheless, it coincides with iterated admissibility in many relevant economic applications.

Keywords: Common assumption of rationality, common belief in rationality, iterated admissibility, rationalizability, lexicographic belief systems.

JEL-Classification: C72.

[†]The Economics and Management Department, The Open University of Israel. Email: aviadhe@openu.ac.il

[‡]Department of Economics, University of Bath, and Institut für Höhere Studien, Wien. Email: m.meier@bath.ac.uk

[§]Department of Economics, University of California, Davis. Email: bcschipper@ucdavis.edu

1   Introduction

What are the behavioral implications of "common cautious belief of rationality"? Consider the following example of a strategic game.[1]

Colin

		a	b	c
	x	$4,0$	$4,1$	$0,2$
Rowena	y	$0,0$	$0,1$	$4,2$
	z	$3,0$	$2,2$	$2,1$

There are two players, Rowena and Colin, each of them possessing three actions. Colin's action a is strictly dominated by any mixture of his other actions. But after eliminating action a, no other action is weakly dominated. Hence, the maximal reduction under iterated admissibility or iterated elimination of weakly dominated strategies is $\{x, y, z\} \times \{b, c\}$. This is also the prediction of "common cautious belief in rationality" if iterated admissibility is taken as embodying this assumption. At a second glance, the only reason why we cannot eliminate action z of Rowena is that it is rationalizable with a full support belief on Colin's actions $\{b, c\}$ that remain after eliminating a in the first round.[2] When Rowena assigns probability $\frac{1}{2}$ both to actions b and c of Colin, she is indifferent between all of her actions. Yet, a cautious player should never completely rule out any action of the opponent. In particular because of her indifference, Rowena may consider what happens if her primary belief in $\{b, c\}$ is contradicted and Colin plays action a nevertheless. Conditional on such a contradiction, Rowena strictly prefers x over z. Thus, she may use this secondary belief in a as a criterion to select among her actions in case of indifference with respect to her primary belief. After eliminating z, we can eliminate successively actions b and x by iterated strict dominance. So, contrary to iterated admissibility, already eliminated actions of an opponent can still serve as tie-breaker between actions of a player later on.

The procedure just described is an instance of a novel solution concept to strategic games that we call *comprehensive rationalizability*. The maximal reduction of comprehensive rationalizability in the game above is $\{(y, c)\}$. In this example, it is a strict

---

[1]This game is a variant of a game discussed in Brandenburger, Friedenberg, and Keisler (2008, p. 313) who attributed it to Pierpaolo Battigalli.
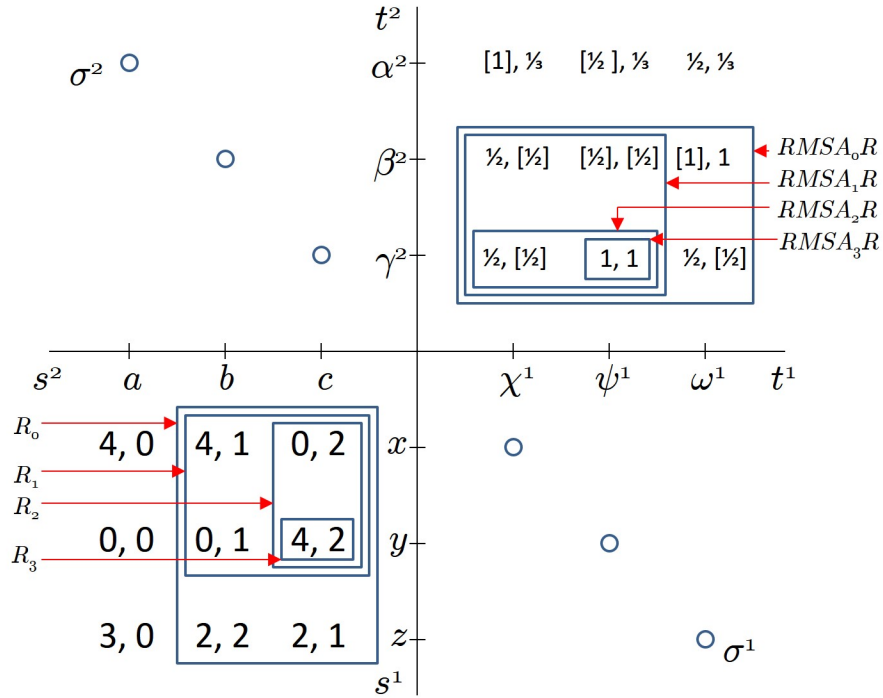
[2]Note that we can also not eliminate z by weak dominance in the first round since it is rationalizable for instance with the belief $\left(\frac{1}{2}, 0, \frac{1}{2}\right)$ on $\{a, b, c\}$, respectively.

refinement of iterated admissibility. However, we will show that this is not the case in general. Surprisingly, comprehensive rationalizability neither refines nor is refined by iterated admissibility. Yet, in many applications in the literature it coincides with iterated admissibility.

Besides defining comprehensive rationalizability, showing some of its properties, and exploring it in some applications, we provide an epistemic characterization of comprehensive rationalizability by rationality and common *strategic assumption* of rationality in a universal lexicographic beliefs type space. To this end, we append the game with types for each player, each type specifying a lexicographic belief system over opponents' types. A lexicographic belief system is a finite sequence of probability measures interpreted as the type's primary belief, secondary belief etc. over the opponents' types. Moreover, we define for each player a strategy map, that is, a function that assigns to each type of the player a pure strategy. Each type's lexicographic belief over the opponents' types together with opponents' strategy maps induce then a full-support lexicographic belief over opponents' profiles of strategies, that is, a sequence of beliefs over opponents' profiles of strategies whose supports are disjoint, and whose union of supports cover the entire strategy space of opponents.

To illustrate the approach informally, consider again the previous example. A lexicographic beliefs type space is depicted in the upper right orthant of Figure 1. For each player, we introduce just three types. The types of Rowena are located on the x-axis while the types of Colin are located on the y-axis. The tuples of numbers indicate the lexicographic beliefs of each type; the lexicographic beliefs of Rowena's types are arranged in columns of first components in each tuple while the lexicographic beliefs of Colin's types are arranged in rows of second components in each tuple. Secondary beliefs are printed in "[...]", while primary beliefs are printed without those brackets. For instance, type χ^1 of Rowena (player 1) assigns primary probability $\frac{1}{2}$ each to types γ^2 and β^2 of Colin (player 2) while assigning secondary probability 1 to type α^2 of Colin. Thus, type χ^1 of Rowena views types γ^2 and β^2 infinitely more likely than type α^2 of Colin. Type β^2 of Colin assigns primary probability 1 to type ω^1 of Rowena and secondary probability $\frac{1}{2}$ each to types χ^1 and ψ^1 of Rowena. The lower right and upper left orthants depict the strategy maps σ^i that map for each player $i = 1, 2$ types into pure strategies. For instance, type χ^1 of Rowena plays $\sigma^1(\chi^1) = x$ while type β^2 of Colin plays $\sigma^2(\beta^2) = b$. The lower left orthant depicts the strategic game above. Note that the lexicographic beliefs of each player's types together with the opponent's strategy map induce a full-support lexicographic belief over the opponent's strategies. For instance, type χ^1 of Rowena assigns

Figure 1: Type Structure

t^2

σ^2 ○

α^2 — [1], ⅓   [½], ⅓   ½, ⅓

○

β^2 — ½, [½]   [½], [½]   [1], 1   ← $RMSA_oR$
← $RMSA_1R$
← $RMSA_2R$
← $RMSA_3R$

○

γ^2 — ½, [½]   1, 1   ½, [½]

| s^2 | a | b | c | | χ^1 | ψ^1 | ω^1 | t^1 |

R_o →
R_1 →
R_2 →
R_3 →

	a	b	c
x	4, 0	4, 1	0, 2
y	0, 0	0, 1	4, 2
z	3, 0	2, 2	2, 1

χ^1 ○

ψ^1 ○

ω^1 ○ σ^1

s^1

primary probability $\frac{1}{2}$ each to Colin playing c and b while assigning secondary probability 1 to Colin playing a.

To sketch our epistemic characterization, consider again type ω^1 of Rowena. With her full-support lexicographic belief, her action z is rational. In fact, the only type in this type structure, who is not rational is Colin's type α^2. His conjecture over Rowena's strategies does not rationalize his action a. In fact, no belief over Rowena's strategies could rationalize playing his action a. The rational type profiles are indicated by the rectangle labeled $RMSA_0R$. Their corresponding level-1 comprehensive rationalizable strategies are labeled by R_0 in the lower left orthant. Consider now again type χ^1 of Rowena. She considers it infinitely more likely that Colin is of the rational type β^2 or γ^2 than him being of the irrational type α^2. Thus, we say that type χ^1 *assumes* that Colin is rational. Similarly, for instance Colin's type β^2 assumes Rowena's rational type ω^1. Note, however, that Rowena's type ω^1 does not assume that Colin is rational since she assigns primary probability $\frac{1}{2}$ to Colin being of the irrational type α^2. Any player's type in the rectangle labeled by $RMSA_1R$ assumes rational types of the opponent. $RMSA_1R$ stands for the event of "rationality and level 1 mutual strategic assumption of rational-

3

ity". It does not imply Rowena's strategy z since Rowena's type ω^1 who plays z does not assume Colin to be rational. The strategies corresponding to types in $RMSA_1R$ are strategies in the rectangle labeled R_1 in the lower left orthant. Continuing in this fashion level by level, we identify the types in $RMSA_3R$ who play comprehensive rationalizable strategies in R_3. This is the set of types in the event "rationality and common strategic assumption of rationality". Note that the type structure depicted in Figure 1 is just a simple example. Our main result characterizes comprehensive rationalizability in the universal lexicographic beliefs type space: The set of comprehensive rationalizable strategies is the set of strategies played by the types in the event "rationality and common strategic assumption of rationality" in the universal space.

There is a growing literature on approaches that characterize various notions of "common cautious belief in rationality". Brandenburger (1992), Samuelson (1992), Börgers and Samuelson (1992), Börgers (1994), Stahl (1995), Ben Porath (1997), Asheim (2001), Ewerhart (2002), Asheim and Dufwenberg (2003), Brandenburger, Friedenberg, and Keisler (2008), Halpern and Pass (2009), Perea (2012), Barelli and Galanis (2013), Keisler and Lee (2015), Yang (2015), Lee (2016), and Catonini and De Vito (2018a, 2018b) all focus on iterated admissibility. Closest to our work is Stahl (1995), Brandenburger, Friedenberg, and Keisler (2008), Keisler and Lee (2015), Yang (2015), Lee (2016), and Catonini and De Vito (2018a). Stahl (1995) defines a notion of rationalizability with lexicographic beliefs and shows that it characterizes iterated admissibility similarly to the characterization of iterated elimination of strictly dominated strategies by (correlated) rationalizability (Bernheim, 1984, Pearce, 1984). In contrast to our approach, he does not require lexicographic beliefs to satisfy mutual singularity. Stahl's analysis remains "pre-epistemic" in the sense that he has no explicit type-structure to define the event that a player is rational, the event that a player believes that another is rational etc. Brandenburger, Friedenberg, and Keisler (2008) add a type structure that allows them to formalize rationality and common assumption of rationality. In their setting, a player assumes an event if she considers the event infinitely more likely than the complement. They show that in their set up rationality and common assumption of rationality does not characterize iterated admissibility but the more general solution concept of self-admissible sets (see also Brandenburger and Friedenberg, 2010). They actually show a negative result according to which there is no complete and continuous type structure that allows for rationality and common assumption of rationality. Keisler and Lee (2015) provide a positive result by dropping the continuity assumption. They also conclude that the negative result of Brandenburger, Friedenberg, and Keisler (2008) is due to the

fact that players are "too cautious" about assuming events in the type space. In some sense, we avoid this problem by allowing types to have lexicographic beliefs with non-full supports over opponents' types while requiring lexicographic beliefs with full-support over opponents' strategies. More precisely, we require that the union of supports of the lexicographic belief on opponents' strategies must cover the entire opponents' strategy space while the union of supports of the lexicographic belief on opponents' types may not cover the entire space of opponents' type profiles. We think that such a formulation is natural as cautiousness should be foremost with respect to others' behavior rather than more abstract constructs such as types. We also assume, as Brandenburger, Friedenberg, and Keisler (2008) do, that any lexicographic belief satisfy disjoint supports, which again we think is natural to assume when beliefs in the sequence are interpreted as alternative hypotheses. The assumption of disjoint supports of the full-support lexicographic belief on others' strategies is akin to the notion of a *basis* of a consistent conditional probability system (Siniscalchi 2016), which by definition has disjoint supports as well.[3]

The assumption of disjoint supports has recently been criticized by Dekel, Friedenberg, and Siniscalchi (2016) as having no "compelling behavioral rationale, or a clear intuitive interpretation."[4] Our use of the disjoint support assumption is not just technical but contributes crucially to the conceptual difference with iterated admissibility. Although comprehensive rationalizability bears similarity to iterated admissibility, the two solution concepts embody two opposing philosophical views.[5] Although a player whose reasoning is consistent with iterated admissibility might believe that the opponent plays an iteratively admissible action because he is rational, she nevertheless entertains the possibility - though infinitesimal it may be - that the opponent will play the same iteratively admissible action when he is irrational. In contrast, the mutual singularity assumption together with our notion of strategic assumption implies that a player whose reasoning is consistent with comprehensive rationalizability bases her choice on the premise that whenever her opponent chooses a comprehensive rationalizable strategy, he does so because he is rational and not by fluke.

This is also the difference to two concurrent papers that also require - as we do -

---

[3]Moreover, assuming that the player unravels in her mind the elimination process of others' strategies, in a dynamic process in which the other players have to choose between their eliminated and surviving strategies, our notion of strategic assumption is akin to structural preference (Siniscalchi 2016) in the unraveled game.

[4]In private communication, one of the authors, Marciano Siniscalchi, assured us that the authors "were concerned with the logic of IA, and effectively just point out that the mutual-singularity assumption of BFK is spurious." Their paper has "nothing to say about" our "interesting and different idea."

[5]We are grateful to a reviewer for pointing this out so eloquently.

5

that the marginal lexicographic beliefs on opponents' actions have full support. Yang (2015) uses it to characterize iterated admissibility in a continuous beliefs-complete type space. He drops mutual singularity of lexicographic beliefs but assumes full support. In contrast to our setting, a player in Yang (2015) may entertain the possibility that an irrational opponent plays an iteratively admissible action. For instance, according to the characterization of weak assumption in Lemma 1 in Yang (2015), a type of Rowena in the above example may play z and satisfy rationality and weak assumption of Colin's rationality by weakly assuming that Colin is rational (and plays either b or c, with equal probabilities), with a second-order belief of Rowena that Colin is irrational and plays either a or c, with equal probabilities, and where c is played by an irrational type of Colin who is certain that Rowena plays z (vis-a-vis which c is actually inferior to b). In contrast, such a second-order belief of Rowena is not compatible with $RMSA_1R$ in our proposed notion of strategic assumption of rationality, a notion that requires any order of belief of Rowena to attribute rationality to a type of Colin who plays c (because there do exist rational types of Colin who play c).

Catonini and De Vito (2018a) also present an epistemic characterization of iterative admissibility while requiring that marginal lexicographic beliefs on opponents' actions have full support. Differently to Yang (2015) they drop the assumption of full support lexicographic beliefs. Moreover, in contrast to our setting, they also drop mutual singularity. The same type of Rowena mentioned above (playing z with a second order belief that Colin is irrational and plays either a or c with equal probabilities) exhibits cautious rationality and assumption of Colin's cautious rationality (Catonini and De Vito 2018a), thus demonstrating how also this notion differs from our proposed notion of $RMSA_1R$.

We view Yang (2015), Catonini and De Vito (2018a), and our paper as complementing each other. Together these papers demonstrate what can be achieved under different assumptions imposed on lexicographic beliefs of players.

Almost all of the above mentioned papers seek an epistemic justification for an existing solution concept and - like in the case of iterated admissibility in Brandenburger, Friedenberg, and Keisler (2008) may fail to justify it by "natural" conditions in their type structure. In contrast, we start with a type structure and sensible epistemic conditions and obtain as a result a new solution concept that may be of interest independently from its epistemic characterization. Moreover, almost all of the above mentioned papers seek an epistemic justification for a "memoryless" procedure of iteratively deleting actions in the sense that only the game remaining from the last round is needed to perform the next round of deletions. In contrast, our solution concept belongs to a new, separate cat-

egory of solution concepts that requires some memory of previously deleted actions. It is an iterative action *demotion* procedure rather than an iterative elimination procedure.[6] Despite the conceptual and technical differences between comprehensive rationalizability and iterated admissibility, we demonstrate with a number of examples that these differences do not play a role in many applications. We conclude that while it might be more difficult to provide epistemic characterizations for iterated admissibility, it is quite natural to provide epistemic characterizations of a related solution concept that also embodies cautiousness with respect to opponents' actions. At the same time, the applied game theorist need not despair because in many applications our solution concept gives rise to the same outcome as iterative admissibility.

The paper is organized as follows: In Section 2 we introduce some preliminary definitions of lexicographic belief systems, conjectures, domination, and best replies. This is followed by Section 3 in which we define comprehensive rationalizability and show that a comprehensive rationalizable outcome exists in every finite strategic game. In Section 4 we define lexicographic beliefs type spaces. In Section 5 we provide an epistemic characterization of comprehensive rationalizability – the main result of the paper. In Section 6 we relate comprehensive rationalizability to rationalizability, iterated admissibility, and an iterative solution concept introduced by Dekel and Fudenberg (1990). In Section 7 we discuss applications. Some of the proofs are relegated to the appendix.

2   Lexicographic Beliefs and Best Replies

In this section, we start to introduce basic notions of lexicographic beliefs, full-support lexicographic beliefs, and lexicographic best replies.

For a non-empty standard Borel space X, $\Delta(X)$ is the space of probability measures on X, with the σ-algebra generated by the events $\{\mu \in \Delta(X) \mid \mu(E) \geq p\}$ for measurable E and $p \in [0, 1]$.

Definition 1 (Lexicographic belief) *A lexicographic belief μ over a standard Borel space X is a finite sequence of probability measures on X*

$$\mu = (\mu_1, \ldots, \mu_n) \in (\Delta(X))^n$$

---

[6]Again, we are grateful to a reviewer for pointing this out very clearly.

7

which are mutually singular:

$$\mu_\ell \perp \mu_{\ell'} \text{ for } \ell \neq \ell'.$$

That is, there are disjoint measurable events $E_1, ..., E_n \subseteq X$ that partition X such that $\mu_\ell(E_\ell) = 1$, $\ell = 1, ..., n$. Hence, $\mu_\ell(E_{\ell'}) = 0$, for $\ell \neq \ell'$.

Blume, Brandenburger, and Dekel (1991a) axiomatize versions of subjective expected utility with lexicographic beliefs in the context of individual choice under uncertainty. In particular, they also axiomatize a version in which probabilities of the sequence of probabilities measures have pairwise disjoint supports and the union of their supports cover the entire space of underlying uncertainties. Such lexicographic beliefs we call full-support lexicographic beliefs. More formally, we define:

Definition 2 (Full-support lexicographic belief) *A full-support lexicographic belief μ over a finite space X is a lexicographic belief with the additional property that the union of the supports of μ cover X:*

$$\bigcup_{\ell \leq n} \text{supp } \mu_\ell = X.$$

We emphasize that the notion of full-support lexicographic beliefs is defined here over a finite space X only. This suffices for our purpose as we focus on finite games with finite sets of strategies.

For any lexicographic belief $\mu = (\mu_1, \ldots, \mu_n) \in (\Delta(X))^n$ we call n the length of the lexicographic belief μ.

We denote the set of lexicographic beliefs on X by $\mathcal{L}(X)$. For a finite space X we denote the set of full-support lexicographic beliefs on X by $\mathcal{L}^+(X)$. Note that in this case we have $\mathcal{L}^+(X) \subseteq \mathcal{L}(X)$.

If Y and Z are two disjoint measurable subsets of X, we say that the lexicographic belief $\mu = (\mu_1, \ldots, \mu_n)$ deems Y as *infinitely more likely* than Z if $\mu_\ell(Y) > 0$ for some $\ell \in \{1, \ldots, n\}$, and whenever $\mu_{\ell'}(Z) > 0$ it is the case that $\mu_{\ell''}(Y) = 0$ for all $\ell'' \geq \ell'$. If Y' and Z' are two subsets of X, not necessarily disjoint, we say that the lexicographic belief $\mu = (\mu_1, \ldots, \mu_n)$ deems Y' as *infinitely more likely* than Z' if it deems $Y' \setminus Z'$ infinitely more likely than Z'.

We say that Y is *assumed* by the lexicographic belief $\mu = (\mu_1, \ldots, \mu_n)$ if it deems Y as infinitely more likely than $X \setminus Y$.

The notion of infinitely more likely is due to Blume, Brandenburger, and Dekel (1991a) in the finite case. Brandenburger, Friedenberg, and Keisler (2008) generalize it to topological spaces. They also introduced the notion of assumption.[7]

The short proof of the following lemma is contained in the appendix. The lemma relates the notion of assumption to the sequence of probabilities of a full-support lexicographic belief. It is a finite version of Proposition 5.1 in Brandenburger, Friedenberg, and Keisler (2008).[8]

Lemma 1 *Let X be a nonempty finite space and $\mu = (\mu_1, ..., \mu_n)$, $n \geq 1$, be a full-support lexicographic belief on X. Let $E \subseteq X$ be such that μ assumes E. Then there is an ℓ with $1 \leq \ell \leq n$ such that $\mu_j(E) = 1$ for all $j \leq \ell$ and $\mu_j(E) = 0$ for $n \geq j > \ell$. If $X \neq E$, we must have $\ell < n$.*

Consider a game Γ with finitely many players $i \in I$ and finite strategy sets $(S^i)_{i \in I}$ equipped with the discrete topology, and utility functions $(u^i)_{i \in I}$.

If $\mu^i = (\mu^i_1, \ldots, \mu^i_n) \in (\Delta(S^{-i}))^n$ is a full-support lexicographic belief over the other players' strategy profiles $S^{-i} = \prod_{j \neq i} S^j$, and $s^i \in S^i$, we say that s^i is a *lexicographic best reply* to μ^i if

$$
\left(\sum_{s^{-i} \in S^{-i}} u^i\left(s^i, s^{-i}\right) \mu^i_\ell(\{s^{-i}\}) \right)^n_{\ell=1} \geq^L \left(\sum_{s^{-i} \in S^{-i}} u^i\left(\hat{s}^i, s^{-i}\right) \mu^i_\ell(\{s^{-i}\}) \right)^n_{\ell=1}
$$

for every $\hat{s}^i \in S^i$.[9]

---

[7]A referee suggested that instead of imposing mutual singularity of the probability measures as the defining feature of a (full-support) lexicographic belief, we could have assumed instead a Lexicographic Probability System (LPS) (Blume, Brandenburger, and Dekel, 1991a) with the additional feature that the space X is partitioned into disjoint members $X_1, ..., X_n$ each of which is deemed infinitely more likely than its successor; a conjecture is that each probability of the LPS would then be a convex combination of its predecessors and an additional probability measure whose support is the next partition member, and that this LPS would therefore represent the same preferences as those representable by a lexicographic belief in our sense, thus replacing a technical condition in the definition of lexicographic beliefs (i.e., mutual singularity) with a condition motivated by a feature of preferences. The conjecture may hinge on additional assumptions (e.g., richness of implications/outcomes), and is therefore left for future exploration.

[8]Because of the finiteness assumption here we do not require condition (iii) in Proposition 5.1 of Brandenburger, Friedenberg, and Keisler (2008).

[9]$(x_\ell)^n_{\ell=1} \geq^L (y_\ell)^n_{\ell=1}$ if and only if $y_j > x_j$ implies $x_k > y_k$ for some $k < j$.

3   Comprehensive Rationalizability

We now define our solution concept that we will obtain from our epistemic characterization. In what follows we will use the notational convention $Y^{-i} = \prod_{j \in I, j \neq i} Y^j$.

Definition 3 (Comprehensive Rationalizability) *Let $C^i_{-1} = \mathcal{L}^+ (S^{-i})$ and $R^i_{-1} = S^i$. Define inductively*

$$
\begin{aligned}
C^i_{k+1} &= \left\{ \mu^i \in C^i_k \mid R^{-i}_k \text{ is assumed by } \mu^i \right\} \\
R^i_{k+1} &= \left\{ s^i \in S^i \mid \exists \mu^i \in C^i_{k+1} \text{ for which } s^i \text{ is a lexicographic best reply} \right\}
\end{aligned}
$$

Player i's comprehensive rationalizable strategies are

$$
R^i_\infty = \bigcap_{k=0}^{\infty} R^i_k.
$$

Two points are worth emphasizing. First, at each level k, players form full-support lexicographic beliefs over other players' strategy profiles. That is, at each level k, a player i's lexicographic belief over opponents' strategies is such that its union of supports covers the opponents' strategy space. Second, while comprehensive rationalizability is defined as a reduction procedure on full-support lexicographic beliefs, it implies immediately a reduction procedure on strategies.

Remark 1 $R^i_k \subseteq R^i_{k-1}$ *for every $k \geq 1$.*

For finite games, we can show that comprehensive rationalizable strategies always exist. The proof is by induction on the levels of full-support lexicographic beliefs.

Proposition 1 (Existence) *For all $i \in I$, $C^i_k \neq \emptyset$ and $R^i_k \neq \emptyset$ for any $k \geq 0$.*

Remark 2 *A direct consequence of the above proposition and of the finiteness of the strategy spaces is that there exists an $r \geq 0$ be such that $C^i_{r+n} = C^i_r$ and $R^i_{r+n} = R^i_r$, for all $i \in I$ and $n \geq 0$.*

Our next goal will be to provide an epistemic characterization of comprehensive rationalizable strategies. To this effect we will first define in the next section a *lexicographic beliefs type space.*

4   Lexicographic Beliefs Type Space

For a given game with a finite set of players I and a finite space of strategy profiles $S = \prod_{i \in I} S^i$,

$$T = \left\langle T^i \right\rangle_{i \in I}$$

is a *lexicographic beliefs type space* if for all $i \in I$, $T^i = \bigcup_{n \geq 1} T_n^i$ and each T_n^i is a standard Borel space, with measurable mappings[10]

$$\sigma^i : T^i \to S^i$$

specifying each type's strategy, defined by the measurable mappings

$$\sigma_n^i : T_n^i \to S^i, \quad n \geq 1$$

and the measurable mappings

$$\tau^i : T^i \to \mathcal{L}\left(T^{-i}\right)$$

specifying each type's state of mind regarding the other players' types, as defined by the measurable mappings

$$\tau_n^i : T_n^i \to \left(\Delta\left(T^{-i}\right)\right)^n \cap \mathcal{L}\left(T^{-i}\right), \quad n \geq 1$$

having the property that the n-tuple of beliefs $\tau_n^i\left(t_n^i\right)_{|S^{-i}}$ of $t_n^i \in T_n^i$ on S^{-i} defined by

$$\tau_n^i\left(t_n^i\right)_{|S^{-i}}(\cdot) := \tau_n^i\left(t_n^i\right)\left(\left(\left(\sigma^j\right)_{j \neq i}\right)^{-1}(\cdot)\right)$$

constitutes a full-support lexicographic belief over S^{-i} (i.e., these beliefs are mutually singular and the union of their supports is S^{-i}). Note that this implies that the maps σ^i are onto, for $i \in I$. We can view τ_n^i as the restriction of τ^i to T_n^i.

We denote the subset of $\mathcal{L}\left(T^{-i}\right)$ with this property by $\mathcal{L}^{\oplus}\left(T^{-i}\right)$. Note that the range of τ^i is in fact $\mathcal{L}^{\oplus}\left(T^{-i}\right)$.

Some (but not all) of the T_n^i may be empty. Actually the condition on the marginals on S^{-i} implies that T_n^i is empty for $n > |S^{-i}|$.

Notice that in this definition, only the beliefs of the types t_n^i on the other players'

---

[10]We endow T^i with the following σ-algebra: $A \subseteq T^i$ is measurable iff $A \cap T_n^i$ is measurable in T_n^i, for all $n \geq 1$.

strategies S^{-i} are required to form a full-support lexicographic belief; in contrast, the beliefs of the types t_n^i on the other players' types T^{-i} are not required to have supports whose union cover T^{-i}.

The motivation for this distinction is that if a player's expected utilities under the primary belief coincide for two strategies, then the player considers opponents' strategies outside the support of the primary belief. To be more precise, note that according to Blume, Brandenburger, and Dekel's (1991a, Theorem 5.3) decision theoretic axiomatization of lexicographic expected utility for a finite space, full-support lexicographic beliefs with non-overlapping supports have an interesting interpretation: The primary belief μ_1 can be interpreted as prior belief. If the expected utilities under μ_1 from two strategies are the same, then the player considers opponents' strategies in $S^{-i} \setminus supp\ \mu_1$. The secondary belief μ_2 takes then the place of the "posterior" conditional on the event $S^{-i} \setminus supp\ \mu_1$. Inductively, for $\ell > 1$ the ℓ-th order belief takes the place of "posterior" belief conditional on the event $S^{-i} \setminus \left(\bigcup_{k=1}^{\ell-1} supp\ \mu_k \right)$ in the case that the expected utilities under $\mu_{\ell'}$ for $\ell' < \ell$ from two strategies are the same.

An alternative motivation for this distinction is that in a prior unmodeled stage before playing the game, player i may potentially get surprising verifiable evidence that her primary belief on the other players' strategy profile S^{-i} was wrong (i.e., that the other players' strategy profiles not ruled out by that evidence was assigned probability zero by the primary belief), in which case she resorts to her secondary belief, and so forth. In contrast, *no direct verifiable evidence is feasible regarding the other players' beliefs.* Hence, prior to playing the game there cannot arise a *necessity* for player i to replace her primary belief about the other players' beliefs, and therefore player i need not necessarily entertain an exhaustive arsenal of mutually singular alternative beliefs on the other players' types. This does not preclude, of course, that a switch of player i to a secondary (or ternary, etc.) belief about the other players' strategies may be *correlated* with a corresponding switch in belief about the other players' types.

For our epistemic characterization, it is desirable to capture all hierarchies of lexicographic beliefs whose marginals on strategies correspond to full-support lexicographic beliefs. That is, it is desirable to use a "rich" lexicographic type space. To this extent we will, introduce the universal lexicographic beliefs type space.

Fix type spaces $T = \langle T^i \rangle_{i \in I}$ and $\tilde{T} = \left\langle \tilde{T}^i \right\rangle_{i \in I}$. Let $h^i : T^i \rightarrow \tilde{T}^i$ be measurable for all $i \in I$. Then $h = (h^i)_{\in I}$ is a *type morphism* if for all $i \in I$ and for all $t^i \in T^i$, $\sigma^i (t^i) = \tilde{\sigma}^i (h^i (t^i))$ and for every measurable $E \subseteq \tilde{T}^{-i}$, $\tau^i (t^i) (h^{-i} (E)) = \tilde{\tau}^i (h^i (t^i)) (E)$.

The lexicographic beliefs type space $\bar{T}$ is *universal* if every type space T admits a unique type morphism to $\bar{T}$.

Lee (2015, Corollary 7.4) proved that a universal space $\bar{T} = \left\langle \bar{T}^i \right\rangle_{i \in I}$ exists and is unique (up to type isomorphism), where $\bar{T}^i = \bigcup_{n=1}^{\infty} \bar{T}_n^i$ and $\bar{T}_n^i$ are, each, the projective limit of coherent hierarchies $\left(\bar{T}_n^{i,m}\right)_{m=1}^{\infty}$ of lexicographic beliefs of length n on $\prod_{j \neq i} \left(\bigcup_{k=1}^{\infty} \left(S^j \times \prod_{\ell=1}^{m-1} \bar{T}_k^{j,\ell} \right) \right)$.

5   Epistemic Characterization

First, we qualify the notion of assumption. A type strategically assumes a subset of opponents' types E if she assumes E and for any strategy profile played by some profile of types in this set E, she deems the opponents' profile of types in E who play this strategy profile infinitely more likely than type-profiles that are not in E but nevertheless play this strategy profile.

The role of "strategic assumption" (in particular the role of the qualifier "strategic") in conjunction with the lexicographic beliefs type space is the following. We do not only want sufficiently rational types of a player to believe that other players are rational and play rational strategies but also that if a rational strategy is played, it is played for rational reasons. Loosely speaking, if a rational type of a player finds a new manuscript of a Dostoyevsky novel written on a typewriter, we want him to believe that it was a Dostoyevsky who wrote that novel and not a monkey that just randomly played with the keyboard of the typewriter typing that manuscript by chance.

Definition 4 (Strategic Assumption) *We say $\tau^i(t^i)$ strategically assumes $E^{-i} \subseteq T^{-i}$ if $\tau^i(t^i)$ assumes E^{-i} and for every $s^{-i} \in \sigma^{-i}(E^{-i})$, $\tau^i(t^i)$ deems $\{t^{-i} \in E^{-i} \mid \sigma^{-i}(t^{-i}) = s^{-i}\}$ as infinitely more likely than $\{t^{-i} \notin E^{-i} \mid \sigma^{-i}(t^{-i}) = s^{-i}\}$.*

Remark 3 Strategic *assumption of E^{-i} by $\tau^i(t^i)$ together with the property that $\tau^i(t^i)_{|S^{-i}}(\cdot)$ is a full-support lexicographic belief implies that for any $s^{-i} \in \sigma^{-i}(E^{-i})$, $\{t^{-i} \notin E^{-i} \mid \sigma^{-i}(t^{-i}) = s^{-i}\}$ gets probability 0 at all levels of $\tau^i(t^i)$.*[11]

The property that the lexicographic marginal on strategy profiles of other players are mutually disjoint, implies that each such profile gets positive probability at most one

---

[11]Note that we could have used alternatively Remark 2 as the definition of strategic assumption. The equivalence of the notion of strategic assumption of Definition 4 and the notion of Remark 3 relies on non-overlapping supports. Although the notion of strategic assumption of Definition 4 makes sense even without non-overlapping supports, the equivalence to the notion in Remark 3 would break down.

level of the lexicographic order. The property that those marginals form a full-support lexicographic belief (i.e., union of supports cover the entire space) implies that such profiles get positive probability at exactly one level. This means the following: If $\tau^i(t^i)$ strategically assumes E^{-i}, then the type profiles in E^{-i} are the sole explanation of type t^i for why strategy profiles in $\sigma^{-i}(E^{-i})$ might be played. Alternative explanations, namely types in $T^{-i} \setminus E^{-i}$ that also play such a profile are not only deemed infinitely less likely, but are discarded altogether.

Without the qualification "strategic" in strategic assumption, the following situation might occur. Type t^i assumes that other players are rational, assumes that others assume that others are rational etc., and yet explains a profile of actions played by such very rational types of players as being played by very irrational types of other players!

In light of these comments, we should clarify that in our approach if a rational player finds a new manuscript of a Dostoyevsky novel, then she entertains not even an infinitesimal probability that a monkey typed it randomly on the keyboard. This distinguishes our approach from epistemic avenues towards iterated admissibility à la Brandenburger, Friedenberg, and Keisler (2008) in which if a rational player finds a new manuscript of a Dostoyevsky novel, then she must entertain an infinitesimal but non-zero probability that a monkey typed it randomly on the keyboard.[12]

In the appendix, we state and prove technical lemmata that are useful for our characterization. First, in any type space, the set of types that deem some measurable subset of opponents' types infinitely more likely than another measurable subset of opponents' types is itself a measurable subset of any given lexicographic beliefs type space (lemma 2). Second, the set of types that strategically assumes a measurable subset of opponents' types is a measurable subset of any given lexicographic beliefs type space (lemma 3).

Our main result seeks to characterize comprehensive rationalizability by an "epistemic analogue" that makes the epistemic conditions behind comprehensive rationalizability transparent. To this end, let $\langle T^i \rangle_{i \in I}$ be a lexicographic beliefs type space for a given game with strategy profiles $S = \prod_{i \in I} S^i$.

Definition 5 (Rationality and Common Strategic Assumption of Rationality)
Define the following sequence of events of player i's Rationality and Mutual Strategic Assumption of degree $k \geq 0$ of Rationality:

$$RMSA_0 R^i := \left\{ t^i \in T^i \mid \sigma^i \left(t^i \right) \ \text{is a lexicographic best reply to } \tau^i \left(t^i \right)_{|S^{-i}} \right\}$$

---

[12]We thank a reviewer for suggesting this clarification.

and inductively

$$RMSA_{k+1}R^i := \left\{ t^i \in RMSA_k R^i \mid \tau^i \left(t^i \right) \text{ strategically assumes } RMSA_k R^{-i} \right\}$$

Furthermore, define the event of i's rationality and common strategic assumption of rationality *to be*

$$RCSAR^i = \bigcap_{k=0}^{\infty} RMSA_k R^i.$$

In the epistemic characterization in Theorem 1 below we show the comprehensive rationalizable strategies are exactly those strategies played by the types in the universal space characterized by rationality and common strategic assumption of rationality. This relates the algorithmic notion of comprehensive rationalizability with its fundamental epistemic counterpart. The proof appears in the appendix.

Theorem 1 (Epistemic Characterization) *In the universal lexicographic beliefs type space $\bar{T}$, for each $k \geq 0$ the strategies $\sigma^i \left(RMSA_k R^i \right)$ played by the types in $RMSA_k R^i$ are the strategies in R_k^i,*

$$R_k^i = \sigma^i \left(RMSA_k R^i \right)$$

and also in the limit, the comprehensive rationalizable strategies of player i are the strategies played by i's types in the event of i's rationality and common strategic assumption of rationality,

$$R_\infty^i = \sigma^i \left(RCSAR^i \right).$$

The difficult direction in the proof requires, for every comprehensive rationalizable strategy, a construction of a specific type in the universal space characterized by rationality and common strategic assumption of rationality and who plays this strategy. This construction is carried out by induction at the beginning of the proof. A sequence of lemmata then shows how the so constructed types do indeed have the required properties. The motivation and logic of the development from one lemma to the next are detailed along the proof in the appendix.

6   Relationship to other Iterative Solution Concepts

6.1   Comprehensive Rationalizability versus Iterated Admissibility

Iterated admissibility has a long tradition in game theory and its applications. For instance, Kohlberg and Mertens (1986) argue that it is a necessary condition for a satisfactory solution to any game. The earliest applications seem to go back to Farquharson (1969), Brams (1975), and Moulin (1979). Iterated admissibility is appealing for several reasons: First, it is easy to apply since it is defined in terms of an algorithm that successively eliminates weakly dominated actions. Moreover, it yields relatively sharp predictions. Second, it is not an equilibrium concept and consequently it does not presume the existence of an equilibrium convention. Third, admissible strategies are equivalent to best responses to full support beliefs (Pearce, 1984, Lemma 4). Thus, iterated admissibility captures optimizing under some form of cautious beliefs. In this section, we compare comprehensive rationalizability with iterated admissibility.

An action $s^i \in S^i$ is *weakly dominated* with respect to $X \times Y \subseteq S^i \times S^{-i}$ if there exist $\alpha^i \in \Delta(S^i)$ with $\alpha^i(X) = 1$ such that $\sum_{\tilde{s}^i \in X} \alpha^i(\tilde{s}^i) u^i(\tilde{s}^i, s^{-i}) \geq u^i(s^i, s^{-i})$ for every $s^{-i} \in Y$ and $\sum_{\tilde{s}^i \in X} \alpha^i(\tilde{s}^i) u^i(\tilde{s}^i, s^{-i}) > u^i(s_i, s^{-i})$ for some $s^{-i} \in Y$. Otherwise, we say that s^i is *admissible* with respect to $X \times Y$.

Let $S^i_{-1} = S^i$ and define for $k \geq 0$

$$S^i_{k+1} = \left\{ s^i \in S^i_k \mid s^i \text{ is admissible with respect to } S^i_k \times S^{-i}_k \right\}.$$

The set of *iteratively admissible* actions of player i is

$$S^i_\infty = \bigcap_{k=0}^{\infty} S^i_k.$$

Similar to Moulin (1979) we say that a game is *solvable* by iterated admissibility if the maximal reduction $S_\infty = \times_{i \in I} S^i_\infty$ is nonempty and for every player $i \in I$, the payoff function u^i is constant with respect to s^i on all outcomes in $S_\infty = \times_{j \in I} S^j_\infty$, i.e., for all $s^i, \tilde{s}^i \in S^i_\infty$, $u^i(s^i, s^{-i}) = u^i(\tilde{s}^i, s^{-i})$ for all $s^{-i} \in S^{-i}_\infty$.[13] Analogously, a game is solvable by comprehensive rationalizability if R_∞ is nonempty and for every player $i \in I$, the payoff function is constant with respect to s^i on all outcomes in $\times_{j \in I} R^j_\infty$. Clearly, if a game is

---

[13]Note though that different from Moulin (1979) we allow for domination by mixed strategies.

solvable by iterated admissibility (or comprehensive rationalizability, respectively), then S_∞ (R_∞, respectively) is a subset of its Nash equilibria.

In the introductory example we have shown that there are games in which the set of comprehensive rationalizable strategies strictly refines the set of iterated admissible strategies. As we show in the next example, this is not generally the case.

Example 1 *Surprisingly, comprehensive rationalizability is* not *a refinement of iterated admissibility as the following example demonstrates.*

	a	b	c
x	$4,0$	$4,1$	$0,1$
y	$0,0$	$0,1$	$4,1$
z	$3,0$	$2,2$	$2,1$
w	$5,0$	$0,1$	$0,2$

The order of elimination under iterated admissibility is a, w, c and then both y and z. The maximal reduction or iterative admissible actions are $\{(x, b)\}$.

For comprehensive rationalizability, the order of elimination is a and then both z and w. The set of comprehensive rationalizable profiles is $\{x, y\} \times \{b, c\}$.

That every first level admissible strategy is first level comprehensive rationalizable and vice versa follows from Blume, Brandenburger and Dekel (1991b, Proposition 1) and Pearce (1984, Lemma 4).

Remark 4 *Every first level admissible strategy is first level comprehensive rationalizable and vice versa. I.e., for any $i \in I$, $S_0^i = R_0^i$.*

Both the introductory example and Example 1 show that this observation does not extend to higher levels. What are sufficient conditions for the equivalence of iteratively admissible strategies and comprehensive rationalizability?

Consider first one of the most studied classes of games in game theory, the class of 2×2 games. For 2×2 games we are able to show that comprehensive rationalizability is equivalent to iterative admissibility. In fact, we can show this equivalence more generally for any $2 \times n$ game (resp. $n \times 2$ game).

Proposition 2 *Let $n \geq 1$. For any $n \times 2$ game (resp. $2 \times n$ game), $S_k^i = R_k^i$, for all $k \geq 0$ and $i = 1, 2$. Hence, the set of comprehensive rationalizable strategy profiles coincides with the set of iterative admissible strategy profiles, i.e., $S_\infty^i = R_\infty^i$, for $i = 1, 2$.*

The proof of the proposition is in the appendix.

Together with the example of the 3×3 game in the Introduction, Proposition 2 implies that the class of $n \times 2$ games (resp. $2 \times n$ games) is a "maximal" class of games where the equivalence between comprehensive rationalizability and iterated admissibility holds without further restrictions.

What can be said about sufficient conditions for equivalence beyond $n \times 2$ games? Consider again Example 1. The reason why comprehensive rationalizability does not refine iterated admissibility there is that no full support belief of player 1 on the first level admissible strategies of player 2 exists such that z is a strict best reply. E.g., for player 1, playing both x or y is as good as playing z against the belief that assigns equal probability on both b and c. The following proposition further develops this observation.

Proposition 3 *If for all $i \in I$, $k \geq 1$, and $s^i \in S_k^i$ there exists a full support belief $\eta^i \in \Delta(S_{k-1}^{-i})$ for which s^i is the unique best reply amongst the actions in S_k^i, then $S_k^i = R_k^i$, for all $i \in I$ and $k \geq -1$. Consequently, $S_\infty^i = R_\infty^i$, for all $i \in I$.*

The proof of the proposition is in the appendix.

6.2 Rationalizability

Bernheim (1984) and Pearce (1984) (see also Spohn, 1982) introduced rationalizability. Let $P_{-1}^i = S^i$ and define for $k \geq -1$, $B_{k+1}^i = \Delta(P_k^{-i})$ and $P_{k+1}^i = \{s^i \in S^i : s^i$ is a best reply to some $\mu \in B_{k+1}^i\}$. The set of *rationalizable actions* is $P_\infty^i = \bigcap_{k=0}^\infty P_k^i$.[14]

Proposition 4 *For every $i \in I$ and $k \geq 0$, $R_k^i \subseteq P_k^i$. Moreover, every comprehensive rationalizable action is rationalizable.*

PROOF.   By definition, $R_{-1}^i = P_{-1}^i = S^i$ for all $i \in I$. Suppose that $R_k^i \subseteq P_k^i$ for all $i \in I$ and some $k \geq -1$.

Let $s^i \in R_{k+1}^i$. Then, s^i is a lexicographic best reply to some full-support lexicographic belief $(\mu_1, ..., \mu_n) \in C_{k+1}^i$ that assumes R_k^{-i}. In particular, it is a best reply to μ_1,

---

[14] "Best reply" refers to the pure action best reply.

18

and by Lemma 1 the support of μ_1 is contained in $R_k^{-i} \subseteq P_k^{-i}$. Hence, s^i is a best reply to some belief (namely μ_1) on P_k^{-i} and therefore $s^i \in P_{k+1}^i$. $\square$

Since well-known games such as *Guess-the-Average* are solvable by rationalizability, the result implies that they are solvable by comprehensive rationalizability as well.

6.3 One round elimination of weakly dominated actions followed by iterative elimination of strictly dominated actions

Dekel and Fudenberg (1990) introduce one round elimination of weakly dominated actions followed by iterative elimination of strictly dominated actions as solution concept, which has been characterized epistemically by Brandenburger (1992), Börgers (1994), Ben Porath (1997), and most recently by Catonini and De Vito (2018b). Let WS_∞ denote the maximal reduction of this procedure.

Proposition 5 *For every player $i \in I$, $R_\infty^i \subseteq WS_\infty^i$.*

PROOF. By definition, we have $S_0^i = WS_0^i$. Hence, by Remark 2, $R_0^i = WS_0^i$. By Pearce (1984, Lemma 3), every action not strictly dominated is a best reply to some belief over opponents' actions and vice versa. Thus the result follows exactly like in the proof of Proposition 4. $\square$

7 Economic Applications

In many economically relevant examples the condition of Proposition 3 is satisfied and comprehensive rationalizability coincides with iterated admissibility. We discuss some of the examples in sequel. The first example concerns voting with a president.

Example 2 (Voting with a president) *Consider an example of majority voting with a president (Moulin, 1986, p. 73-74). Three players have to select one of three alternatives $\{a, b, c\}$. If a majority votes for an alternative, it will be implemented. Otherwise, the alternative selected by player 1, the president, is selected. For each player, the ranking of the alternatives from most preferred to least preferred is given by numbers $3, 2$, and 1, respectively. The strategic form is given by the following three matrixes where player 1 chooses matrices, player 2 rows and player 3 columns.*

	a	b	c
a	$3,2,1$	$3,2,1$	$3,2,1$
b	$3,2,1$	$2,1,3$	$3,2,1$
c	$3,2,1$	$3,2,1$	$1,3,2$

a

	a	b	c
a	$3,2,1$	$2,1,3$	$2,1,3$
b	$2,1,3$	$2,1,3$	$2,1,3$
c	$2,1,3$	$2,1,3$	$1,3,2$

b

	a	b	c
a	$3,2,1$	$1,3,2$	$1,3,2$
b	$1,3,2$	$2,1,3$	$1,3,2$
c	$1,3,2$	$1,3,2$	$1,3,2$

c

At the first round, the only admissible strategy of player 1 is a. For player 2 it is the set $\{a,c\}$, since for her b is weakly dominated by c. For player 3 it is $\{b,c\}$, since for him, a is weakly dominated by b. Hence the game is reduced to

	b	c
a	$3,2,1$	$3,2,1$
c	$3,2,1$	$1,3,2$

a

One more round of elimination of weakly dominated strategies leads to (a,c,c), the only iterative admissible profile. Interestingly, in this profile the president faces his lowest ranked alternative.

Since S_1^i is a singleton for $i = 1,2,3$, Proposition 3 trivially applies and the unique comprehensive rationalizable strategy profile is (a,c,c) as well.

Example 3 (Dividing Money) *The following game is due to Brams, Kilgour, and Davis (1993), see also Osborne (2004, p. 38). Two players use the following procedure to divide $\$10$ between themselves. Each person names a number of dollars (a nonnegative integer), at most equal to $\$10$. If the sum is at most $\$10$, then each person receives the amount of money she named and the remainder is burned. If the sum exceeds $\$10$ and the players named different amounts, then the person who named the smaller amount receives that amount and the other player receives the remaining money. If the sum exceeds $\$10$ and the amounts named by the players are the same, then each player receives $\$5$.*

In the first round, every amount weakly lower than $\$5$ is weakly dominated by $\$6$. If the opponent names more than $\$6$, then the sum exceeds $\$10$ and the player receives $\$6$. If the opponent names exactly $\$6$, then the sum also exceeds $\$10$, and both players receive $\$5$. If the opponent names any amount strictly less than $\$5$, then the sum is at most $\$10$, and the player receives $\$6$. Any other amount $a \in \{\$6, ..., \$9\}$ is a strict best response to

a + 1, and $10 is a strict best response to $0.

In the following rounds, the highest remaining amount is weakly dominated by $6 and all other amounts are a strict best response to that amount plus 1. Thus, the maximal reduction under iterated admissibility is ($6, $6) yielding a payoff of $5 to each player. Note that since every amount not eliminated at the previous round is a strict best reply to some full support belief on the amounts that survived till the round before, Proposition 3 applies and the game is solvable by comprehensive rationalizability yielding the same outcome as the IA-procedure.

7.1   Price Competition

Consider a symmetric Bertrand duopoly in which each firm is restricted to choose integer prices. Let p be the price and c the cost (also in integers), and

$$D(p) = \begin{cases} \alpha - p & \text{if } p \leq \alpha \\ 0 & \text{if } p > \alpha \end{cases}$$

and assume $c + 1 < \alpha$. The profit function of firm $i \neq j$ is given by

$$\pi(p_i, p_j) = \begin{cases} (p_i - c)D(p_i) & \text{if } p_i < p_j \\ \frac{1}{2}(p_i - c)D(p_i) & \text{if } p_i = p_j \\ 0 & \text{if } p_i > p_j \end{cases}$$

Assume that the monopoly price is unique. (If $\alpha + c$ is even, this is the case and the monopoly price is $0.5(\alpha + c)$).

Proposition 6 *In the Bertrand duopoly, $p = c + 1$ is the unique comprehensive rationalizable price. It is also the unique iterative admissible price.*

The proof is in the appendix.

7.2   Second Price Common Value Auctions

Consider a second price common value auction in which each of the n bidders receives a privately observed signal x^i about the value of the object to be auctioned off. The signals x^i, $i = 1, ..., n$ are independently and identically distributed over some finite set $X \subset \mathbb{N}$

(integers) and we assume that every signal may be drawn with positive probability. Let $x_{\max}$ denote the realization of the highest of these n signals (i.e., first order statistic). The common value of the object to each bidder is $x_{\max}$. Each bidder submits a bid in a sealed envelope. The highest bidder wins and pays the second highest bid. In case of a tie, each highest bidder obtains the object with equal probability. Let $b^i : X \longrightarrow \mathbb{R}$ denote the bid function of player i.

Bidding your value is the unique bidding function that is obtained after two rounds of iterated admissibility (see Harstad and Levin, 1985). Comprehensive rationalizability yields the same outcome.

Proposition 7 *In the second price common value sealed bid auction, $R_k^i = \{b(x^i) = x^i\}$ for all $k \geq 2$ and $i \in I$.*

The proof is in the appendix.

7.3 Comprehensive Rationalizable Implementation

Many economic problems take the following form: How to design without knowledge of the preferences of the individuals an institution in which individuals interact such that any outcome of their interaction satisfies certain desirable properties such as efficiency etc.? Such problems are referred to as implementation problems as they "implement" those outcomes in some solution concept of the game to be designed (i.e., the institution). What outcomes can be implemented in comprehensive rationalizable strategies?

Let X denote the set of simple lotteries (i.e., with finite support) over an arbitrary set of alternatives. Each player $i \in I$ has now preferences over lotteries represented by $u^i : X \times \mathbb{R} \times \Psi^i \longrightarrow \mathbb{R}$, where Ψ^i is a finite set of utility parameters for player i. Distinct parameters in Ψ^i are associated with distinct preferences orderings over $X \times \mathbb{R}$. Moreover, we assume that a player is never indifferent between all lotteries in X. The function is linear in its first argument.

We assume that the preference profile $\psi \in \Psi = \times_{i \in I} \Psi^i$ is common knowledge among players in I. Yet, the social planner does not know ψ and wants to implement some lottery over alternatives in X.

A social choice function $f : \Psi \longrightarrow X$ associates with each preference profile a lottery over alternatives. We consider a finite mechanism with transfers $\langle M^1, ..., M^n, g, t \rangle$ defined by a finite action set M^i for each player i, an outcome function $g : M \longrightarrow X$ that

associates with each action profile in $M = \times_{i \in I} M^i$ a lottery over outcomes, and a transfer rule $t = (t^i)_{i \in I} : M \longrightarrow \mathbb{R}^n$ that for each player associates with each action profile a fine. (The second argument in each player's utility function refers to the fines. Less fines are preferred to more.) A mechanism $\langle M^1, ..., M^n, g, t \rangle$ and a preference profile in Ψ define a strategic game with complete information.

A mechanism exactly implements a social choice function f in comprehensive rationalizable strategies with fines bounded by $\bar{t} > 0$ if and only if $| t^i(m) | \leq \bar{t}$ for all $m \in M$ and $i \in I$, and for any $\psi \in \Psi$, there exists $m^*(\psi) \in M$ such that $g(m^*(\psi)) = f(\psi)$, $t(m^*(\psi)) = 0$, and $R_\infty(\psi) = \{m^*(\psi)\}$. A social choice function f is exactly implementable in comprehensive rationalizable strategies with small fines if for all $\bar{t} > 0$, there exists a mechanism which exactly implements f with fines bounded by $\bar{t}$.

Proposition 8 *Suppose that there are at least three players. Then any social choice function is exactly implementable in comprehensive rationalizable strategies with small fines.*

The proof is in the appendix.

A   Proofs

Proof of Proposition 1

Both R^i_{-1} and C^i_{-1} are nonempty for all $i \in I$. Assume R^i_k and C^i_k are nonempty for all $i \in I$.

Claim: There exists a $\sigma^i_{k+1} \in C^i_k$ that assumes R^{-i}_k.

Proof of the claim: Fix $\sigma^i_k = (\mu^i_{1,k}, ..., \mu^i_{m,k}) \in C^i_k$. Note that by the induction hypothesis σ^i_k assumes R^{-i}_ℓ for $\ell = -1, ..., k-1$. Again, by the induction hypothesis R^{-i}_k is nonempty and by Remark 1, $R^{-i}_k \subseteq R^{-i}_{k-1}$.

Since R^{-i}_k is nonempty by the induction hypothesis, and σ^i_k is a full-support lexicographic belief (in particular it is a full support sequence), we have a $j \leq m$ such that $\mu^i_{j,k}(R^{-i}_k) > 0$. Without loss of generality assume $R^{-i}_k \subsetneq R^{-i}_{k-1}$, since otherwise σ^i_k already assumes R^{-i}_k.

Let $\mu^i_{n_1,k}, ..., \mu^i_{n_p,k}$ be all the $\mu^i_{j,k}$ such that $\mu^i_{j,k}(R^{-i}_k) > 0$ with $n_r < n_{r+1}$, for all $r = 1, ..., p-1$. Moreover, let $\mu^i_{m_1,k}, ..., \mu^i_{m_q,k}$ be all the $\mu^i_{j,k}$ such that $\mu^i_{j,k}(R^{-i}_{k-1} \setminus R^{-i}_k) > 0$.

Since σ_k^i assumes R_{k-1}^{-i}, there exists an $\ell \leq m$ such that $\mu_{j,k}^i(R_{k-1}^{-i}) = 1$ for all $j = 1, ..., \ell$ and $\mu_{j,k}^i(R_{k-1}^{-i}) = 0$ for $j > \ell$. Note that we have $\{n_1, ..., n_p\} \cup \{m_1, ..., m_q\} = \{1, ..., \ell\}$.

Now, define

$$\sigma_{k+1}^i = \left(\nu_{1,k+1}^i, ..., \nu_{p+q+m-\ell,k+1}^i \right)$$

where

$$\nu_{j,k+1}^i = \begin{cases} \mu_{n_j,k}^i(\cdot \mid R_k^{-i}) & \text{for } j = 1, ..., p \\ \mu_{m_{j-p},k}^i(\cdot \mid R_{k-1}^{-i} \setminus R_k^{-i}) & \text{for } j = p+1, ..., p+q \\ \mu_{j-p-q+\ell,k}^i & \text{for } j = p+q+1, ..., p+q+m-\ell. \end{cases}$$

By construction σ_{k+1}^i assumes R_k^{-i} and R_{k-1}^{-i}, but also R_j^{-i} for $j < k-1$. To see this, let $j < k-1$. Since σ_k^i assumes R_j^{-i}, there exists $\ell' \geq 1$ such that $\mu_{r,k}^i(R_j^{-i}) = 1$ for $r \leq \ell'$ and $\mu_{r,k}^i(R_j^{-i}) = 0$ for all $r > \ell'$. But since $R_k^{-i} \subseteq R_j^{-i}$, we have $\ell' \geq \ell$. If $R_k^{-i} = R_j^{-i}$, there is nothing to show. If $R_k^{-i} \subsetneq R_j^{-i}$ then, by the fact that σ_k^i is a full-support lexicographic belief (in particular, a full support sequence), we must have $\ell' > \ell$. Then, $\nu_{r,k+1}^i(R_j^{-i}) = 1$ for $r = 1, ..., p+q$, but also for all $j = p+q+1, ..., p+q-\ell+\ell'$. And we have by construction, $\nu_{r,k+1}^i(R_j^{-i}) = 0$ for $r > p+q-\ell+\ell'$. So, σ_{k+1}^i assumes also R_j^{-i}. This finishes the proof of the claim.

Since C_{k+1}^i is nonempty and the game is finite, we must have that R_{k+1}^i is nonempty. $\square$

Proof of Lemma 1

If $\mu_j(X \setminus E) = 0$ for all j with $1 \leq j \leq n$, then $\mu_j(E) = 1$ for all j. Otherwise, let k be the smallest $k \leq n$ such that $\mu_k(X \setminus E) > 0$. Then by the definition of infinitely more likely and assumption $\mu_j(E) = 0$ for all $j \geq k$. Since each μ_j is a probability measure, $\mu_j(X \setminus E) = 0$ for $j < k$ implies for $\mu_j(E) = 1$ for $j < k$. By the definition of infinitely more likely and assumption, we cannot have $\mu_1(X \setminus E) > 0$, hence $\mu_1(E) = 1$. If $E \neq X$, the fact μ has full support implies that there is a j with $\mu_j(X \setminus E) > 0$. Hence $k \leq j$ and $\ell = k - 1 < n$. $\square$

Lemmata on Measurability of Subsets of Types

Lemma 2 *In every lexicographic beliefs type space, for any measurable sets $Y \subseteq T^{-i}$ and $Z \subseteq T^{-i}$ with $Y \cap Z = \emptyset$ we have $\{t^i \mid \tau^i(t^i) \text{ deems } Y \text{ infinitely more likely than } Z\}$ is measurable.*

PROOF. Let $Y, Z \subseteq T^{-i}$ be measurable and $Y \cap Z = \emptyset$.

For any n, T_n^i is measurable.

For $\ell \leq n$ the set $B_\ell^{i,>0}(Y) := \{t^i \in T_n^i \mid \tau^i(t^i)_\ell(Y) > 0\}$ as well as the set $\{t^i \in T_n^i \mid \tau^i(t^i)_\ell(Z) = 0\}$ are measurable by definition.

Hence $B_{I,\ell}^i(Y, Z)^n = \{t_n^i \mid \tau^i(t_n^i)_\ell(Y) > 0, \tau^i(t_n^i)_k(Y) = 0, \text{ for } k > \ell, \tau^i(t_n^i)_j(Z) = 0, \text{ for } j \leq \ell\}$ is measurable, since it is an intersection of measurable sets.

The set $B_I^i(Y, Z)^n := \{t_n^i \in T_n^i \mid \tau^i(t_n^i) \text{ deems } Y \text{ as infinitely more likely than } Z\}$ is $\bigcup_{\ell=1,\dots,n} B_{I,\ell}^i(Y, Z)^n$, and therefore measurable.

Hence the set $\bigcup_{n \geq 1} B_I^i(Y, Z)^n$ is measurable. $\qquad\square$

Lemma 3 *In every lexicographic beliefs type space T, for any measurable event $E^{-i} \subseteq T^{-i}$, the event $\{t^i \mid \tau^i(t^i) \text{ strategically assumes } E^{-i}\}$ is measurable in T^i.*

PROOF. For all $s^{-i} \in \sigma^{-i}(E^{-i})$ the set

$$\{t^i \mid \tau^i(t^i) \text{ deems } (\sigma^{-i})^{-1}(\{s^{-i}\}) \cap E^{-i} \text{ as infinitely more likely than } ((\sigma^{-i})^{-1}(\{s^{-i}\})) \backslash E^{-i})\}$$

is measurable by Lemma 2.

The set

$$\{t^i \mid \tau^i(t^i) \text{ stratgically assumes } E^{-i}\}$$

is an intersection of the measurable set

$$\{t^i \mid \tau^i(t^i) \text{ assumes } E^{-i}\}$$

and finitely many measurable sets

$$\{t^i \mid \tau^i(t^i) \text{ deems } (\sigma^{-i})^{-1}(\{s^{-i}\}) \cap E^{-i} \text{ as infinitely more likely than } ((\sigma^{-i})^{-1}(\{s^{-i}\})) \backslash E^{-i})\},$$

where $s^{-i} \in \sigma^{-i}(E^{-i})$. $\qquad\square$

Proof of Theorem 1

Within this proof, all events are subsets of the universal lexicographic beliefs type space $\bar{T} = \langle \bar{T}^i \rangle_{i \in I}$ for the given game Γ with strategy profiles $S = \prod_{i \in I} S^i$.

The theorem will be proved by induction. Given the induction hypothesis for all levels up to k, it will not be difficult to show that every type of player i in $RMSA_{k+1}R^i$, i.e. a type that mutually strategically assumes rationality up to level $k + 1$, necessarily plays some strategy in R^i_{k+1}, i.e. a strategy that survives the demotion procedure up to level $k + 1$. The more challenging direction will be to show the converse, namely that for every strategy in R^i_{k+1} there exists *some* type in $RMSA_{k+1}R^i$ in the universal type space that plays this strategy. The following construction is aimed at achieving just that. To wit, we construct a finite belief-closed subspace of the universal space within which each strategy of each player is played by exactly one type, who is as rational as possible given that it plays this strategy.

Construction Let $r \geq 0$ be the minimal natural number such that $C^i_{r+n} = C^i_r$ and $R^i_{r+n} = R^i_r$, for all $i \in I$ and $n \geq 0$. Such an r exists in view of Remark 2.

Stage 1. For each $s^i \in R^i_r$ choose $\mu^i_1(s^i) \in C^i_r$ such that s^i is a lexicographic best reply to the full-support lexicographic belief $\mu^i_1(s^i) = (\beta^i_{1,1}(s^i), ..., \beta^i_{1,n(s^i)}(s^i))$ on the other players' strategy profiles. Likewise, for any $s^i \notin R^i_r$, choose a $\mu^i_1(s^i) \in C^i_m$ with m maximal $< r$ such that s^i is a lexicographic best reply to $\mu^i_1(s^i) = (\beta^i_{1,1}(s^i), ..., \beta^i_{1,n(s^i)}(s^i))$, if such a $\mu^i_1(s^i)$ exists. Otherwise, if s^i is not a lexicographic best reply to any full-support lexicographic belief on the other players' strategies, let $\mu^i_1(s^i) = (\beta^i_{1,1}(s^i), ..., \beta^i_{1,n(s^i)}(s^i))$ be any full-support lexicographic belief.

Stage 2. Define the second-order lexicographic belief $\mu^i_2(s^i) = (\beta^i_{2,1}(s^i), ..., \beta^i_{2,n(s^i)}(s^i))$ of player i on the other players' profiles of strategies and first-order beliefs by

$$\beta^i_{2,\ell}(s^i)(\cdot) := \sum_{s^{-i} \in S^{-i}} \beta^i_{1,\ell}(s^i)(\{s^{-i}\}) \cdot \delta_{(s^j, \mu^j_1(s^j))_{j \neq i}}(\cdot)$$

for every level $\ell = 1, ..., n(s^i)$ of the lexicographic $\mu^i_1(s^i)$, and where $s^{-i} = (s^j)_{j \neq i}$. In words, the first order belief $\mu^i_1(s^i)$ is now extended with a certainty that for every given strategy s^j of every one of the other players $j \neq i$, if this strategy is played then it is played by a type with first-order belief $\mu^j_1(s^j)$ defined in stage 1. By definition $\text{marg}_{|S^{-i}} \beta^i_{2,\ell}(s^i)(\cdot) = \beta^i_{1,\ell}(s^i)(\cdot)$ for every $\ell = 1, ..., n(s^i)$.

Inductive step. Let $m \geq 1$. Assume for each $i \in I$ and $s^i \in S^i$, we have al-

ready defined by induction $t^{i,m}(s^i) = \left(s^i, (\beta^i_{1,\ell}(s^i), ..., \beta_{m,\ell}(s^i))_{\ell \leq n(s^i)}\right) \in \bar{T}^{i,m}_{n(s^i)}$. Then define $\beta^i_{m+1,\ell}(s^i)(\cdot) := \sum_{s^{-i} \in S^{-i}} \beta^i_{1,\ell}(s^i)(\{s^{-i}\}) \cdot \delta_{(t^{j,m}(s^j))_{j \neq i}}(\cdot)$. By the induction hypothesis $\beta^i_{m,\ell}(s^i) = \sum_{s^{-i} \in S^{-i}} \beta^i_{1,\ell}(s^i)(\{s^{-i}\})\delta_{(t^{i,m-1}(s^j))_{j \neq i}}$ and we have that $\mathrm{marg}_{|\bar{T}^{-i}_m}\delta_{(t^{j,m}(s^j))_{j \neq i}}(\cdot) = \delta_{(t^{j,m-1}(s^j))_{j \neq i}}(\cdot)$ since $t^{j,m}(s^j)$ extends $t^{j,m-1}(s^j)$ by construction.

Then by construction $t^i(s^i) = (s^i, (\beta^i_{1,\ell}(s^i), ...)_{\ell \leq n(s^i)}) \in \bar{T}^i_{n(s^i)} \subset \bar{T}^i$ is a type in the universal space, whose belief is $\bar{\tau}^i(t^i(s^i))_\ell = \sum_{s^{-i} \in S^{-i}} \beta^i_{1,\ell}(s^i)(\{s^{-i}\})\delta_{(t^j(s^j))_{j \neq i}}(\cdot)$. This finishes the construction.

This construction implies a preliminary result. Types in the just constructed subset of the universal space, who play m-level comprehensive rationalizable strategies, strategically assume those sets of the just constructed opponents' types that play p-level comprehensive rationalizable strategies, for all $p < m$. This is shown formally in Lemma 4.

Lemma 4 *For all $i \in I$, for all $m \geq 0$, for all $p < m$, and for all $s^i \in R^i_m$, $\tau^i(t^i(s^i))$ strategically assumes $\{t^{-i}(s^{-i}) \mid s^{-i} \in R^{-i}_p\}$.*

PROOF. By definition, if $s^i \in R^i_m$, then μ^i_1 assumes R^{-i}_p for all $p < m$. Note, this is also true for $m \geq r+1$, since $C^i_r = C^i_{r+n}$, for all $n \geq 1$. Therefore, by construction of $t^i(s^i)$, $\tau^i(t^i(s^i))$ assumes $\{t^{-i}(s^{-i}) \mid s^{-i} \in R^{-i}_p\} =: t^{-i}(R^{-i}_p)$. But also, by construction, if $s^{-i} \in \sigma^{-i}(t^{-i}(R^{-i}_p)) = R^{-i}_p$, then $t^i(s^i)$ deems $\{t^{-i}(\hat{s}^{-i}) \mid \sigma^{-i}(\hat{s}^{-i}) = s^{-i}$ and $\hat{s}^{-i} \in R^{-i}_p\} = \{t^{-i}(s^{-i})\}$ as infinitely more likely than $\{t^{-i} \notin t^{-i}(R^{-i}_p) \mid \sigma^{-i}(t^{-i}) = s^{-i}\}$. Note, by construction this latter set gets probability 0 at every level of $\tau^i(t^i(s^i))$, while $\tau^i(t^i(s^i))_\ell(\{t^{-i}(s^{-i})\}) > 0$ if and only if $\beta^i_{1,\ell}(s^i)(\{s^{-i}\}) > 0$. But since $\mu^i_1(s^i)$ is a full-support lexicographic belief on S^{-i} this happens for some $\ell \leq n(s^i)$. $\square$

Moreover, we can also show that a type in the universal type space who is rational and mutually strategically assumes rationality at the m-th level must assume that opponents play strategies consistent with rationality and mutual strategic assumption of rationality at the $m-1$-th level. This is made precise in Lemma 5.

Lemma 5 *For all $i \in I$ and $m \geq 1$, if $t^i \in RMSA_m R^i$, then $\tau^i(t^i)_{|S^{-i}}$ assumes $\sigma^{-i}(RMSA_{m-1}R^{-i})$.*

PROOF. Let $t^i \in RMSA_m R^i$. Since $\tau^i(t^i)$ strategically assumes $RMSA_{m-1}R^{-i}$, there is an index $\ell \geq 1$ such that $\tau^i(t^i)_{\ell'}(RMSA_{m-1}R^{-i}) = 1$ for all $\ell' \leq \ell$ and

$\tau^i(t^i)_{\ell'} (RMSA_{\ell'} R^{-i}) = 0$ for all $\ell' > \ell$. This implies that if $(\tau^i(t^i)_{\ell'})_{|S^{-i}}(\{s^{-i}\}) > 0$, for some $\ell' \leq \ell$, then $s^{-i} \in \sigma^{-i}(RMSA_{m-1}R^{-i})$, since $RMSA_{m-1}R^{-i} \subseteq (\sigma^{-i})^{-1}(\sigma^{-i}(RMSA_{m-1}R^{-i}))$.

Conversely, let $s^{-i} \in \sigma^{-i}(RMSA_{m-1}R^{-i})$. By the second condition of the definition of strategic assumption and the definition of infinitely more likely, we have $\tau^i(t^i)_{\ell'} (\{t^{-i} \in RMSA_{m-1}R^{-i} \mid \sigma^{-i}(t^{-i}) = s^{-i}\}) > 0$ for some ℓ' and since $\tau^i(t^i)_{\ell''}(RMSA_{m-1}R^{-i}) = 0$ for all $\ell'' > \ell$, we have $\ell' \leq \ell$. But this implies that $(\tau^i(t^i)_{\ell'})_{|S^{-i}}(\{s^{-i}\}) > 0$. Together with the first part of the induction step, we have that $\sigma^{-i}(RMSA_{m-1}R^{-i})$ is assumed by $\tau^i(t^i)_{|S^{-i}}$. $\qquad\square$

In Lemmata 6 and 7 below we show that if the strategies played by rational types in the universal space who also mutually strategically assume rationality at level p are exactly the p-level comprehensive rational strategies for $0 \leq p < m$, then the strategies played by rational types in the universal space who also mutually strategically assume rationality at the m-th level must be m-level comprehensive rationalizable. Lemma 6 shows the easier direction, while lemma 7 requires the construction.

Lemma 6 *For all $m \geq 0$, if $\sigma^i(RMSA_p R^i) = R_p^i$, for all $0 \leq p < m$ and all $i \in I$, then $\sigma^i(RMSA_m R^i) \subseteq R_m^i$, for all $i \in I$.*

PROOF. Let $s^i \in \sigma^i(RMSA_p R^i)$. Then there is a $t^i \in RMSA_m R^i$ with $\sigma^i(t^i) = s^i$. By Lemma 5, $\tau^i(t^i)_{|S^{-i}}$ assumes $\sigma^{-i}(RMSA_p R^{-i}) = R_p^{-i}$, for all $p < m$. Since $t^i \in RMSA_0 R^i$, s^i is a lexicographic best reply to $\tau^i(t^i)_{|S^{-i}}$. Hence $s^i \in R_m^i$. $\qquad\square$

Lemma 7 *For all $m \geq 0$, if it is the case that $s^i \in R_p^i$ if and only if $t^i(s^i) \in RMSA_p R^i$, for all $0 \leq p < m$ and all $i \in I$, then $s^i \in R_m^i$ implies $t^i(s^i) \in RMSA_m R^i$, for all $i \in I$.*

PROOF. Let $s^i \in R_m^i$ for some $m \geq 0$. $\tau^i(t^i(s^i))_\ell$ assigns probability 1 to $\{t^{-i}(s^{-i}) \mid s^{-i} \in S^{-i}\}$, for all $\ell \leq n(s^i)$. Hence, $\tau^i(t^i(s^i))$ strategically assumes $RMSA_p R^{-i}$ iff $\tau^i(t^i(s^i))$ strategically assumes $RMSA_p R^{-i} \cap \{t^{-i}(s^{-i}) \mid s^{-i} \in S^{-i}\}$, for $0 \leq p < m$.

If $m \geq 0$, by construction s^i is a lexicographic best reply to $\tau(t^i(s^i))_{|S^{-i}}$.

Now let $m > 0$ and $s^i \in R_m^i$. By the induction hypothesis, for all j we have $\{t^j(s^j) \mid s^j \in R_p^j\} = \{t^j(s^j) \mid t^j(s^j) \in RMSA_p R^j\}$, for $0 \leq p < m$.

Now, by the above observation, for all $\ell \leq n(s^i)$ we have $\tau^i(t^i(s^i))_\ell (\{t^{-i}(s^{-i}) \mid s^{-i} \in R_p^{-i}\}) = \tau^i(t^i(s^i))_\ell(RMSA_p R^{-i})$, for $0 \leq p < m$.

By Lemma 4 $\tau^i(t^i(s^i))$ strategically assumes $\left(\{t^{-i}(s^{-i}) \mid s^{-i} \in R_p^{-i}\}\right)$, for $0 \leq p < m$, and hence strategically assumes $RMSA_p R^{-i}$, for $0 \leq p < m$. Hence, since s^i is a lexicographic best reply to $\tau(t^i(s^i))_{|S^{-i}}$, we have shown that $t^i(s^i) \in RMSA_m R^i$. $\qquad\square$

These preliminary observations then allow us to provide a characterization for every finite level of mutual strategic assumption of rationality/comprehensive rationalizability and the limit.

Lemma 8 *For all $m \geq 0$ and all $i \in I$: $\sigma^i(RMSA_m R^i) = R_m^i$ and $s^i \in R_m^i$ if and only if $t^i(s^i) \in RMSA_m R^i$.*

PROOF.   By induction on $m \geq 0$. Let $m = 0$. By Lemma 6, $\sigma^i(RMSA_0 R^i) \subseteq R_0^i$. In particular, since $\sigma^i(t^i(s^i)) = s^i$ by construction, $t^i(s^i) \in RMSA_0 R^i$ implies $s^i \in R_0^i$. By Lemma 7, $s^i \in R_0^i$ implies that $t^i(s^i) \in RMSA_0 R^i$. Hence, we also have $R_0^i \subseteq \sigma^i(RMSA_0 R^i)$.

Let $m \geq 1$ and assume that $\sigma^i(RMSA_p R^i) = R_p^i$ and that $s^i \in R_p^i$ if and only if $t^i(s^i) \in RMSA_p R^i$, for all p with $0 \leq p \leq m - 1$. Then, by Lemma 6, $\sigma^i(RMSA_m R^i) \subseteq R_m^i$. By Lemma 7, $s^i \in R_m^i$ implies $t^i(s^i) \in RMSA_m R^i$.

Hence, $\sigma^i(RMSA_m R^i) = R_m^i$. $\square$

Using this characterization result for all $m \geq 0$, we can now prove that the characterization holds also in the limit.

Lemma 9 *For all $i \in I$, $R_\infty^i = \sigma^i(RMSA_\infty R^i)$.*

PROOF.   By Lemma 8 we have for all $m \geq 0$ that $s^i \in R_m^i$ if and only if $t^i(s^i) \in RMSA_m R^i$.

We have $s^i \in R_\infty^i$ if and only if $s^i \in R_m^i$, for all $m \geq 0$, if and only if $t^i(s^i) \in RMSA_m R^i$, for all $m \geq 0$, if and only if $t^i(s^i) \in RMSA_\infty R^i$. Since $t^i(s^i)$ plays s^i, we have that $R_\infty^i \subseteq \sigma^i(RMSA_\infty R^i)$. On the other hand, if $t^i \in RMSA_\infty R^i$ it follows that $t^i \in RMSA_m R^i$, for all $m \geq 0$, and therefore $\sigma^i(t^i) \in R_m^i$, for all $m \geq 0$, by Lemma 8. But $\sigma^i(t^i) \in R_m^i$, for all $m \geq 0$, implies $\sigma^i(t^i) \in R_\infty^i$. Hence we have show that $\sigma^i(RMSA_\infty R^i) = R_\infty^i$, for all $i \in I$. $\qquad\square$

Our characterization theorem now summarizes Lemmata 8 and 9. Recall from Proposition 1 that $R_\infty^i \neq \emptyset$, for all $i \in I$. This finishes the proof of Theorem 1. $\qquad\square$

Proof of Proposition 2

Let $S^2 = \{b_1, b_2\}$.

Note that in the two procedures of comprehensive rationalizability and iterated admissibility, a player can only eliminate some additional actions in stage $m+1$ if the other player has eliminated some action(s) in stage m.

Since, by Remark 4 every first level admissible strategy is first level comprehensive rationalizable and vice versa, if iterative admissibility does not eliminate any action in stage 0, then so does comprehensive rationalizability and vice versa. In such a case, both procedures stop at stage 0 and $S^i_k = R^i_k$, for all $k \geq 0$ and $i = 1, 2$, and $S^i_\infty = R^i_\infty$, for $i = 1, 2$.

Note that player 2 cannot eliminate an action later than in stage 1, since player 1 can at most once (namely in stage 0) eliminate actions without player 2 having eliminated any action.

Since player 2 can eliminate at most one action (either in stage 0 or in stage 1), player 1 can only eliminate actions in at most two stages (1st in stage 0 and 2nd in stage 1, if player 2 has eliminated an action in stage 0, resp. 2nd in stage 2 if player 2 has eliminated an action in stage 1). But of course, it is also possible that player 1 does not eliminate any action in stage 0, or only in stage 0.

Consider the case that player 1 eliminates some actions in stage 0 according to iterated admissibility (and hence also to comprehensive rationalizability) and player 2 does not eliminate any action in any stage according to iterated admissibility. Since $S^2_0 = R^2_0$, player 2 cannot eliminate any action according to comprehensive rationalizability in stage 0.

Now, assume by contradiction, that player 2 eliminates an action, say b_2 according to comprehensive rationalizability in stage 1. Then, b_1 is a strictly better reply than b_2 to any conjecture on S^1 that assumes S^1_0. It cannot be that there is an action $a \in S^1_0$ to which b_2 is a strict better reply than b_1. Otherwise there is a full support belief μ_1 (giving enough weight to action a) on S^1_0 to which b_2 is a strict better reply than b_1 and hence by considering any full support belief μ_2 on $S^1 \setminus S^1_0$, we have that (μ_1, μ_2) is a full-support lexicographic belief on S^1 assuming S^1_0 to which b_2 is the unique best reply. But then b_2 could not have been eliminated according to comprehensive rationalizability in stage 1. If there is an action in S^1_0 to which b_1 is a strict better reply than b_2, since there is no action in S^1_0 to which b_2 is a strict better reply than b_1, conditional on S^1_0, b_1 weakly dominates b_2, and hence b_2 would be eliminated according to IA in stage 1, a

contradiction. Hence, against any action of player 1 in S_0^1, both actions of player 2 give the same payoff. Therefore, since b_1 is a strictly better reply than b_2 to any full-support lexicographic belief on S^1 that assumes S_0^1, b_1 must be a strictly better reply than b_2 to any full-support lexicographic belief on $S^1 \setminus S_0^1$, and hence in particular to any full support belief on $S^1 \setminus S_0^1$. But this implies, since both actions are equivalent conditional on S_0^1, that b_1 must be a strictly better reply than b_2 to any full support belief on S^1. But then, by Pearce (1984) Lemma 4, b_2 must have been eliminated according to iterated admissibility in stage 0, again a contradiction.

If player 2 eliminates some action according to iterated admissibility in stage 0, Remark 4 (and the fact that there is nothing any more to eliminate) implies that $S_k^2 = R_k^2$, for all $k \geq 0$.

Since $S_0^2 = R_0^2$ in any case, if player 2 eliminates an action according to iterated admissibility in stage 1, since then S_1^2 is a singleton, Proposition 3 below implies that $S_k^2 = R_k^2$, for all $k \geq 0$.

Taken altogether, we have now shown that $S_k^2 = R_k^2$, for all $k \geq 0$, and since we know by Remark 4 that $S_0^1 = R_0^1$, it remains to show that $S_k^1 = R_k^1$, for all $k \geq 1$.

If player 2 does not eliminate any action, then player 1 cannot (according to neither comprehensive rationalizability nor iterated admissibility) eliminate any action in stages $k \geq 1$, and it follows that $S_k^1 = R_k^1$, for all $k \geq 1$.

So, assume that player 2 eliminates one action, say b_2 in either stage $m = 0$ or in stage $m = 1$.

If $m = 1$, then player 1 cannot eliminate any action in stage 1 according to iterated admissibility or comprehensive rationalizability. Hence, in both cases ($m = 0, 1$) we have $S_m^1 = S_0^1 = R_0^1 = R_m^1$.

According to iterated admissibility, in stage $m + 1$ player 1 eliminates exactly those actions in S_0^1 that do not give the maximal payoff against b_1 amongst the actions in S_0^1. If an action $a \in S_0^1$ of player 1 gives the maximal payoff against b_1 amongst the actions in S_0^1, but gives a lower payoff against b_2 than some other action $a' \in S_0^1$, where a' also gives the maximal payoff against b_1 amongst the actions in S_0^1, then a is weakly dominated by a' and so a must have been already eliminated in stage 0, a contradiction.

According to comprehensive rationalizability, in stage $m+1$ player 1 assumes action b_1 of player 2. Hence his actions in R_{m+1}^1 are exactly those actions in S^1 that give maximal payoff against b_1 and among the actions with this property those that give maximal payoff against b_2. Note that the difference of R_{m+1}^1 to the set S_{m+1}^1 is only that here the

surviving actions satisfy a certain optimality criterion (lexicographic optimality against b_1 and only the against b_2) globally (that is in comparison to all other actions in S^1), while in S^1_{m+1} the same criterion is applied locally (amongst the actions in S^1_0). But note that such an action in R^1_{m+1} is obviously not weakly dominated in S^1 and so could not have been eliminated according to iterated admissibility in stage 0. So, any action that is in R^1_{m+1} is in S^1_0. Hence, the global optimum is in the set considered locally, and hence the sets R^1_{m+1} and S^1_{m+1} coincide. $\qquad\square$

Proof of Proposition 3

For all $i \in I$, $S^i_0 = R^i_0$ follows from Remark 2. (Note that trivially, we also have $S^i_{-1} = R^i_{-1}$, for all $i \in I$.)

Induction Step: Let $k \geq 0$ and $i \in I$. We show that if for all $j \in I$, $S^j_k = R^j_k$ and for every $s^i \in S^i_{k+1}$ there exists a full support belief $\eta^i \in \Delta(S^{-i}_k)$ for which s^i is the strict best reply in S^i_{k+1}, then $S^i_{k+1} = R^i_{k+1}$.

First, we show $S^i_{k+1} \subseteq R^i_{k+1}$. Let $\eta^i \in \Delta(S^{-i}_k)$ be full support on S^{-i}_k such that $s^i \in S^i_{k+1}$ is the unique best reply in S^i_{k+1} to η^i. Then s^i is also the unique best reply in S^i_k to η^i, since any best reply amongst the actions in S^i_k to η^i must be in S^i_{k+1} already. Consider any full-support lexicographic belief $\mu = (\mu^i_1, ..., \mu^i_n) \in C^i_k$, for some $n \geq 1$. Let $\mu^i_{n_1}, ..., \mu^i_{n_p}$ be all the μ^i_m such that $\mu^i_m(S^{-i} \backslash S^{-i}_k) > 0$, ordered such that $n_r < n_{r+1}$, for all $r = 1, ..., p-1$. Now, set $\nu^i_1 := \eta^i$ and $\nu^i_m := \mu_{n_{m-1}}(\cdot \mid S^{-i} \backslash S^{-i}_k)$, for $m = 2, ..., p+1$. Then $\nu^i = (\nu^i_1, ..., \nu^i_{p+1})$ assumes $S^{-i}_k = R^{-i}_k$. Since $\mu^i \in C^i_k$, it follows easily (like in the proof of Proposition 1) that ν^i assumes R^{-i}_q, for $q = -1, ..., k-1$. Hence, $\nu^i \in C^i_{k+1}$. Let $\tilde{s}^i$ be any lexicographic best reply in S^i to ν^i. Since $\nu^i \in C^i_{k+1}$, we have that $\tilde{s}^i \in R^i_{k+1} \subseteq R^i_k = S^i_k$. In particular, $\tilde{s}^i$ is at least as good a reply in S^i_k to $\eta^i = \nu^i_1$ as s^i is. By the uniqueness assumption, we have $\tilde{s}^i = s^i$. We conclude that $s^i \in R^i_{k+1}$.

Next, we show that $R^i_{k+1} \subseteq S^i_{k+1}$. Let $s^i \in R^i_{k+1}$. There exists a full-support lexicographic belief $\mu^i = (\mu^i_1, ..., \mu^i_n) \in C^i_{k+1}$, for some $n \geq 1$ such that s^i is a lexicographic best reply to μ^i. Since μ^i assumes R^{-i}_k, by Lemma 1, there is an $\ell \in \{1, ..., n\}$ such that $\mu^i_j(R^{-i}_k) = 1$ for $1 \leq j \leq \ell$ and $\mu^i_j(R^{-i}_k) = 0$ for $n \geq j \geq \ell$. By the induction hypothesis, $S^{-i}_k = R^{-i}_k$. By Blume, Brandenburger, and Dekel (1991b, Proposition 1), applied to $(\mu^i_1, ..., \mu^i_\ell)$, there is a full support belief η^i on S^{-i}_k such that s^i is a best reply (even in S^i) to η^i. Hence, since $s^i \in R^i_{k+1} \subseteq R^i_k = S^i_k$, it follows that $s^i \in S^i_{k+1}$.

It follows that $S^i_k = R^i_k$, for all $k \geq -1$ and all $i \in I$. Hence, $S^i_\infty = \bigcap^\infty_{k=-1} S^i_k = \bigcap^\infty_{k=-1} R^i_k = R^i_\infty$ for every $i \in I$. $\qquad\square$

Proof of Proposition 6

We consider iterative admissibility and then apply Proposition 3 to show that it coincides with comprehensive rationalizability. In the first round every price in excess of the monopoly price is weakly dominated by the monopoly price. If the opponent sets a price weakly higher than the monopoly price, then the monopoly price is strictly better than any price strictly higher than the monopoly price. If the opponent sets a price strictly below the monopoly price, the monopoly price is as good as any price strictly higher than the monopoly price. In fact, there is no belief about the opponent's prices for which a price strictly higher than the monopoly price is a best reply. Also in the first round, every price equal to at most c is weakly dominated by the price $c+1$. If the opponent sets a price weakly larger than $c+1$, then a price equal to $c+1$ is strictly better than a price equal to at most c. If the opponent sets a price strictly below $c+1$, then a price equal to $c+1$ is as good as a price equal to c and strictly better than a price strictly below c. That is, c is a best reply to the belief that the opponent sets a price of at most c but it is not the unique best reply. c can never be a best reply to a full support belief. Every other price p is a strict best response to $p+1$, so no other price is weakly dominated. To see this, note that for any $p \geq c+1$ it is strictly better to obtain all the demand at the price p than to obtain half of the demand at the price $p+1$. That is, consider any $p \geq c+2$. We need to show that $\frac{1}{2}(p-c)(\alpha-p) < (p-1-c)(\alpha-p+1) = (p-1-c)(\alpha-p)+p-1-c$. We have $\frac{1}{2}(p-c) \leq p-1-c$ and $p-1-c > 0$ because $p \geq c+2$. By the same argument, at each subsequent round of iterative elimination, the highest remaining price is weakly dominated by the next highest price. The pair of prices that remains is $(c+1, c+1)$.

Let p_ℓ be the highest price that remains after the ℓ-th round of elimination of weakly dominated prices. Proposition 3 implies that at each level $\ell = 0, 1, ...,$ the set of admissible prices coincide with the set of comprehensive rationalizable prices since every price in $\{c+1, ..., p_\ell\}$ is a strict best reply to a full support belief over prices remaining from the previous round. E.g., $p \in \{c+1, ..., p_\ell\}$ is a strict best reply to the belief that is concentrated on $p+1$. $\qquad \square$

Proof of Proposition 7

For each bidder $i \in I$, any bid strictly above $\max X$ and strictly below x^i is weakly dominated. Thus, by Remark 2 $R_0^i = \{b^i : \text{For all } x^i \in X, x^i \leq b(x^i) \leq \max X\}$.

At the second level, suppose a bidder bids strictly above his signal. If his bid is below the highest bid, then he does not obtain the object. Bidding his signal would be

a lexicographic best reply in this case as he receives nothing and pays nothing. If the second highest bid is below his signal, then he still receives the object. Bidding his signal would be a lexicographic best reply in this case since he would still obtain the object and his payment would remain unchanged.

Consider now the case in which the second highest bid is between his bid and his signal. By his second level full-support lexicographic beliefs he assumes that all bidders $j \neq i$ select a bid in $\{x^j, \max X\}$. Thus, in this case $x_{\max} > x_i$. This means he would pay more than the value of the object. A lexicographic best reply is to bid his signal instead, in which case he pays nothing and obtains nothing. In fact, it is the only lexicographic best reply since opponents may have drawn any signals from X. Thus we have shown that at the second level, bidding the signal remains as the only comprehensive rationalizable bid function. □

Proof of Proposition 8

Abreu and Matsushima (1994) show that any social choice function is exactly implementable in iterated admissible actions. In their proof, they show that in fact it is exactly implementable with one round elimination of weakly dominated actions followed by many rounds of elimination of strictly dominated actions. In their mechanism, the maximal reductions of both procedures coincide. Thus the result follows from Proposition 5 above. □

References

[1] Abreu, D. and H. Matsushima (1994). Exact implementation, Journal of Economic Theory 64, 1–19.

[2] Aliprantis, C. and K. Border (2007). Infinite dimensional analysis: A Hitchhiker's guide, 3rd edition, Springer-Verlag

[3] Asheim, G. (2001). Proper rationalizability in lexicographic beliefs, International Journal of Game Theory 30, 453–478.

[4] Asheim, G. and M. Dufwenberg (2003). Admissibility and common belief, Games and Economic Behavior 42, 208–234.

[5] Barelli, P. and S. Galanis (2013). Admissibility and event rationality, Games and Economic Behavior 77, 21–40.

[6] Battigalli, P. (1996). Strategic rationality orderings and the best rationalization principle, Games and Economic Behavior 13, 178–200.

[7] Ben Porath, E. (1997). Rationality, Nash equilibrium, and backward induction in perfect information games, Review of Economic Studies 64, 23–46.

[8] Bernheim, B.D. (1984). Rationalizable strategic behavior, Econometrica 52, 1007–1028.

[9] Blume, L., Brandenburger, A., and E. Dekel (1991a). Lexicographic probabilities and choice under uncertainty, Econometrica 59, 61–79.

[10] Blume, L., Brandenburger, A., and E. Dekel (1991b). Lexicographic probabilities and equilibrium refinements, Econometrica 59, 81–98.

[11] Börgers, T. (1994). Weak dominance and approximate common knowledge, Journal of Economic Theory 64, 265–276.

[12] Börgers, T. and L. Samuelson (1992). Cautious utility maximization and iterated weak dominance, International Journal of Game Theory 21, 13–25.

[13] Brandenburger, A. (1992). Lexicographic probabilities and iterated admissibility, in: Dasgupta, P., Gale, D., Hart, O., and E. Maskin (eds.), Economic analysis of markets and games, Cambridge, MA: MIT Press, 282–290.

[14] Brandenburger, A., Friedenberg, A. and J. Keisler (2008). Admissibility in games, Econometrica 76, 307–352.

[15] Brams, S.J. (1975). Game theory and politics, New York: Free Press.

[16] Brams, S.J., Kilgour, D.M., and M.D. Davis (1993). Unraveling in games of sharing and exchange, in: Binmore, K.G., Kirman, A., and P. Tani (eds.), Frontiers of Game Theory, MIT Press, 195–212.

[17] Catonini, E. and N. De Vito (2018a). Cautious belief and iterated admissibility, mimeo.

[18] Catonini, E. and N. De Vito (2018b). Weak belief and permissibility, mimeo.

[19] Dekel, E., Friedenberg, A. and M. Siniscalchi (2016). Lexicographic beliefs and assumption, Journal of Economic Theory 163, 955–985.

[20] Dekel, E. and D. Fudenberg (1990). Rational behavior with payoff uncertainty, Journal of Economic Theory 52, 243–267.

[21] Ewerhart, C. (2002). Ex-ante justifiable behavior, common knowledge, and iterated admissibility, mimeo.

[22] Farquharson, R. (1969). Theory of Voting, New Haven: Yale University Press.

[23] Halpern, J. and R. Pass (2009). A logical characterization of iterated admissibility, in: Proceedings of Twelfth Conference on Theoretical Aspects of Rationality and Knowledge, 146–155.

[24] Harstad, R. and D. Levin (1985). A class of dominance-solvable common value auctions, Review of Economic Studies 52, 525–528.

[25] Keisler, J. and Lee B.S. (2015). Common assumption of rationality, mimeo.

[26] Kohlberg, E. and J.-F. Mertens (1986). On the strategic stability of equilibria, Econometrica 54, 1003–1037.

[27] Lee B.S. (2015). Generalizing type spaces, mimeo.

[28] Lee B.S. (2016). Admissibility and assumption, Journal of Economic Theory 163, 42–72.

[29] Lo K.C. (1999). Nash equilibrium without mutual knowledge of rationality, Economic Theory 14, 621–633.

[30] Moulin, H. (1986). Game theory for the social sciences, 2nd edition, New York University Press.

[31] Moulin, H. (1979). Dominance solvable voting schemes, Econometrica 47, 1337–1351.

[32] Osborne, M. (2004). An introduction to game theory, Oxford University Press.

[33] Pearce, D. (1984). Rationalizable strategic behavior and the problem of perfection, Econometrica 52, 1029–1050.

[34] Perea, A. (2012). Epistemic game theory: Reasoning and choice, Cambridge University Press.

[35] Samuelson, L. (1992). Dominated strategies and common knowledge, Games and Economic Behavior 4, 284–313.

[36] Siniscalchi, M. (2016). Structural rationality in dynamic games, mimeo.

[37] Spohn, W. (1982). How to make sense of game theory, in: Stegmüller, W., Balzer, W., and W. Spohn (eds.), Philosophy of economics, Springer-Verlag, 239–270.

[38] Stahl, D.O. (1995). Lexicographic rationalizability and iterated admissibility, Economics Letters 47, 155–159.

[39] Yang C.-C. (2015). Weak assumption and iterative admissibility, Journal of Economic Theory 158, 87–101.