MEMORANDUM

No 38/2000

LEXICOGRAPHIC PROBABILITIES AND RATIONALIZABILITY IN EXTENSIVE GAMES

By
Geir B. Asheim and Andrés Perea

ISSN: 0801-1117

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.oekonomi.uio.no/
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no/
e-mail: frisch@frisch.uio.no

List of the last 10 Memoranda:

No 37	By Geir B. Asheim and Wolfgang Buchholz: THE MALLEABILITY OF UNDISCOUNTED UTILITARIANISM AS A CRITERION OF INTERGENERATIONAL JUSTICE. 22 p.
No 36	By Olav Bjerkholt: A turning point in the development of Norwegian economics - the establishment of the University Institute of Economics in 1932. 60 p.
No 35	By Jon Strand: Tax distortions, household production and black-market work. 33 p.
No 34	By Snorre Kverndokk and Knut Einar Rosendahl: CO_2 mitigation costs and ancillary benefits in the Nordic countries, the UK and Ireland: A survey. 53 p.
No 33	By Jon Strand: Competitive effort and employment determination with team production. 25 p.
No 32	By Øyvind Eitrheim, Eilev S. Jansen and Ragnar Nymoen: Progress from forecast failure - The Norwegian consumption function. 27 p.
No 31	By Michael Hoel and Tor Iversen: Genetic testing when there is a mix of public and private health insurance. 27 p.
No 30	By Geir Høidal Bjønnes and Dagfinn Rime: Customer Trading and Information in Foreign Exchange Markets. 38 p.
No 29	By Geir Høidal Bjønnes and Dagfinn Rime: FX Trading… LIVE! Dealer Behavior and Trading Systems in Foreign Exchange Markets. 36 p.
No 28	By Gunn Elisabeth Birkelund and Johan Heldal: Educational Homogamy in Norway -trends and patterns. 17 p.

A complete list of this memo-series is available in a PDF® format at:
http://www.oekonomi.uio.no/memo/

LEXICOGRAPHIC PROBABILITIES AND RATIONALIZABILITY IN EXTENSIVE GAMES

GEIR B. ASHEIM AND ANDRÉS PEREA

Abstract. The concepts of sequential and quasi-perfect rationalizability are defined in an epistemic model by means of lexicographic probabilities. These are non-equilibrium analogs to sequential and quasi-perfect equilibrium, for which epistemic characterizations are provided. The defined rationalizability concepts are shown to imply backward induction in generic perfect information games, but they do not yield forward induction. The relationship between various concepts are shown and illustrated. *JEL* Classification Number: C72.

1. Introduction

The aim of this paper is to define rationalizability concepts for extensive games that can be viewed as non-equilibrium analogs of 'sequential' and 'quasi-perfect' equilibrium. These concepts will be referred to as *sequential* and *quasi-perfect rationalizability*. By the phrase "non-equilibrium analog of sequential equilibrium" we mean that the sequential rationalizability concept, defined through an epistemic model, should characterize sequential equilibrium when adding the requirement that each player is certain of the beliefs that the opponent has about the player's own strategy choice. Likewise for quasi-perfect equilibrium. As a consequence, we provide epistemic characterizations of the concepts of 'sequential' and 'quasi-perfect' equilibrium. We show by means of examples that established rationalizability concepts for extensive games are not such analogs.

To avoid the issue of whether (and if so, how) each player's belief about the strategy choices of other players are stochastically independent, all discussion and analysis will be limited to two-player games.

To motivate our definitions and results, it is instructive to look at the relationship between equilibrium concepts illustrated in Table 1. Here, 'weak sequential' equilibrium refers to the equilibrium concept that results when each player only optimizes at information sets that the player's own strategy does not preclude from being reached, while 'quasi-perfect' equilibrium is the concept defined by van Damme [41] and which differs from Selten's [39] extensive form perfect equilibrium by having each player ignore the possibility of his own future mistakes. The arrows indicate that any proper equilibrium corresponds to a quasi-perfect equilibrium and so forth.

Date: 6 November 2000

Corresponding author: Geir B. Asheim, Department of Economics, University of Oslo, P.O. Box 1095 Blindern, N-0317 Oslo, Norway. *Tel*: +47 2285 5498 *Fax*: +47 2285 5035 *E-mail*: g.b.asheim@econ.uio.no

Acknowledgments: We have benefited from extensive discussions with Drew Fudenberg and Joseph Greenberg as well as comments from Eddie Dekel and Andrew McLennan. Asheim gratefully acknowledges the hospitality of Harvard University and financial support from the Research Council of Norway. Perea gratefully acknowledges the hospitality of Maastricht University.

Table 1. Relationship between different equilibrium concepts.

				Proper equilibrium [32]
				↓
		Strategic form perfect equil. [39]	←	*Quasi-perfect equilibrium* [41]
		↓		↓
Nash equi-librium	←	*Weak sequential equilibrium*	←	*Sequential equilibrium* [26]

There exist non-equilibrium analogs to some of these equilibrium concepts. Of course, the notion of 'rationalizability' due to Bernheim [12] and Pearce [34] is a non-equilibrium analog to Nash-equilibrium.[1] It is epistemically characterized by there being common belief of each player choosing rationally. Likewise, the notion of 'permissibility' due to Börgers [16] and Brandenburger [17] corresponds to Selten's [39] strategic form perfect equilibrium.[2] It is epistemically characterized by there being common belief of each player being cautious (in the sense of taking into account all opponent strategies) and choosing rationally. Finally, Ben-Porath [11] gives an epistemic characterization of a non-equilibrium analog to weak sequential equilibrium by there being common belief of each player choosing rationally at all information sets that the player's own strategy does not preclude from being reached. The resulting concept is coined 'weak extensive form rationalizablity' by Battigalli & Bonanno [9].

In the concepts of sequential and quasi-perfect equilibrium it is assumed that a player chooses rationally at every information set, also at information sets that the player's own strategy precludes from being reached. Within an epistemic model such an assumption is problematic because there can be information sets that are not reachable if the player chooses rationally. In fact, this has caused considerable problems when it comes to giving an epistemic characterization of backward induction. One solution that is suggested in Asheim [2] is instead to assume that each player *believes that the opponent* chooses rationally at every information set. Upon reaching a surprising information set, the player may then update his beliefs about the opponent's beliefs about the player's own strategy choice, or – if that cannot serve as an explanation – that the opponent has not chosen rationally at earlier information sets. Still, the player upholds the belief that the opponent will choose rationally at future information sets. This is the assumption that is used in the present paper to define sequential and quasi-perfect rationalizability.

Note that the perspective of viewing a behavioral strategy of an opponent as an expression of the beliefs of a player (instead of as the choice of an opponent) is consistent with the interpretation provided by Rubinstein [37] of strategies that

[1] A strategy is rationalizable (in two-player games, which is the class we consider here) if and only if it survives iterated elimination of strongly dominated strategies.

[2] A strategy is permissible if and only if it survives one round of weak elimination followed by iterated strong elimination. This is the so-called Dekel-Fudenberg [19] procedure.

TABLE 2. Relationship between different rationalizability concepts.

Common cert. belief that each player ...	... believes the oppon. chooses rationally only in the whole game		... believes the oppon. chooses rationally at all reachable info. sets		... believes the oppon. chooses rationally at all info. sets
... is cautious and respects preferences	[n.a.]		[n.a.]		*Proper rationalizability* [38, 3]
					↓
... is cautious	[n.a.]		*Permissibility (the D-F proc.)* [16, 17, 19]	←	*Quasi-perfect rationalizability* (defined here)
			↓		↓
... is not necessarily cautious	*Rationalizability* [12, 34]	←	*Weak ext. form rationalizability* [11]	←	*Sequential rationalizability* (here; [20, 24])
	Does not imply backward ind.		Does not imply backward ind.		Implies backward ind.

specify actions also at information sets that the strategy precludes from being reached. In addition, it solves a troubling feature raised by Rubinstein—can the opponent choose the beliefs of the player?—since the opponent strategy is here derived from the beliefs of the player subject to the constraint that the strategy specifies rational choice for the opponent at all of her information sets.

Sequential rationalizability is defined by there being common certain belief of the event that each player *believes that the opponent chooses rationally at all information sets*.[3] *Quasi-perfect rationalizability* is defined by there being common certain belief of the event that each player *is cautious* and *believes that the opponent chooses rationally at all information sets*. Since these are non-equilibrium concepts, it is not assumed that each player is certain of the beliefs that the opponent has about the player's own strategy choice. Hence, a player may in general update his belief about the opponent's beliefs about the player's own strategy choice upon reaching a surprising information set. However, if we *assume* that each player is certain of the beliefs that the opponent has about the player's own strategy choice, then we obtain epistemic characterizations of the corresponding equilibrium concepts: sequential and quasi-perfect equilibrium. To the best of our knowledge, there have not been such characterizations available.[4] Also, the rationalizability concepts are different from those which are established in the literature for extensive games.

Our definitions and characterizations make use of the representation of a *lexicographic probability system* (LPS) due to Blume, Brandenburger & Dekel [13]. An LPS is a hierarchy of subjective probability distributions, which allows one opponent strategy to be deemed infinitely more likely than another, while still taking

[3]'Certain belief' of an event means that the complement of the event is deemed 'impossible' (or formally, Savage-null).

[4]See, however, McLennan [27] who characterizes sequential equilibrium by means of consistent conditional probability systems. In contrast to our analysis he does so within the agent strategic form and does not employ an epistemic model.

into account the latter. Blume et al. [14] use LPSs to characterize strategic form perfect equilibrium and proper equilibrium. Strategic form perfect equilibrium is characterized by each player (1) being certain of the beliefs of the opponent and (2) being cautious and believing that the opponent chooses rationally, while proper equilibrium is characterized by each player (1) being certain of the beliefs of the opponent and (2) being cautious and *respecting opponent preferences* (in the sense that if the opponent prefers one strategy to another, then then the former is deemed infinitely more likely than the latter). Asheim [3] has used the latter characterization to formulate an alternative and equivalent definition of Schuhmacher's [38] concept of *proper rationalizability*, by assuming that there is common certain belief of caution and respect of opponent preferences. Here we instead assume that there is common certain belief of the event that each player believes that the opponent chooses rationally at all information sets, an event that can in a straightforward manner be formulated by means of LPSs. This by itself gives sequential rationalizability, while the addition of caution gives quasi-perfect rationalizability.[5]

When a player believes that the opponent chooses rationally in the whole game and is certain that she is cautious, then he believes that the opponent chooses rationally at all information sets that her own strategy does not preclude from being reached. This means that the cell in Table 2 to the left of 'permissibility' is not applicable. Moreover, when a player respects the preferences of the opponent and is certain that she is cautious, then he believes that the opponent chooses rationally at all information sets. Hence, the two cells in Table 2 to the left of 'proper rationalizability' are not applicable. The latter observation also means that 'proper rationalizability' implies 'quasi-perfect rationalizability', which clearly (since it satisfies an additional requirement) implies 'sequential rationalizability'.

Already in the original contributions on non-equilibrium concepts by Bernheim [12] and Pearce [34] there are suggestions on how to define rationalizability concepts for extensive games. Bernheim [12] defines the concept of *subgame rationalizability* by requiring rationalizability in every subgame. This concept coincides with sequential rationalizability for multi-stage games (games with almost perfect information), but no epistemic characterization is offered. Bernheim ([12], p. 1022) claims that it is possible to define a concept of sequential rationalizability, but does not indicate how this can be done. Dekel, Fudenberg & Levine [20] and Greenberg [24] intend to define a concept of sequential rationalizability that coincide with the present one in two-player games; Dekel et al. also use the term 'sequential rationalizability', while Greenberg calls his concept 'null mutually acceptable course of action'. However, their definitions rely on the notion of a 'mixture' or a 'convex hull of a set' of behavioral strategies. This is left ambiguous by Greenberg, while Dekel et al. formalize this notion (in their Def. 2.2 and footnote 11) by – in effect – not putting restrictions on behavior at non-reachable information sets, meaning that their concept of sequential rationalizability as defined does not imply backward induction in e.g. Γ_2 of the present Sect. 5.[6]

[5] For defining sequential rationalizability we could instead have used conditional probability systems (cf. e.g. [9, 10]). However, LPSs are more suited for imposing caution and respect of opponent preferences.

[6] Drew Fudenberg has informed us that he and his co-authors in [20] intended to define a concept of sequential rationalizability that implies backward induction in generic perfect information games. In a corrigendum [21] they formalize a 'mixture' of behavioral strategies by an approach that is equivalent to Approach 1 of Sect. 6, so that their definition of sequential rationalizability

Pearce [34] defines the concept of *extensive form rationalizability* (EFR), which has during the last few years been subject to a fair amount of attention (cf. Battigalli [8] and Battigalli & Siniscalchi [10]). To understand this concept in the context of Table 2, one should stay in the middle column as EFR only requires players to choose rationally at reachable information sets. Moreover, since caution is not imposed, the appropriate row is the lower one. However, EFR is very different from 'weak extensive form rationalizability'. It corresponds to an iterative procedure where, at any information set, any deleted strategy is deemed infinitely less likely than some remaining strategy (provided that there exists at least some remaining strategy under which the information set is reachable). This leads to a rationality ordering for each player (cf. Battigalli [7]). In the epistemic characterization by Battigalli & Siniscalchi [10], they invoke a non-monotonic epistemic operator and must employ a belief-complete epistemic model. Neither of these features play any role in the present analysis.[7]

EFR yields *forward induction* in common examples like the 'Battle-of-the-Sexes-with-Outside-Option' and 'Burning Money' games. In contrast, it follows that *none* of the concepts in Table 2 implies forward induction since not even the concept of proper *equilibrium* promotes forward induction in these games. EFR also leads to the *backward induction outcome*. As indicated in Table 2, the three concepts in the right column imply backward induction. However, they yield not only the backward induction outcome; the concepts of proper, quasi-perfect and sequential rationalibility even support the *backward induction procedure*.

Our paper is organized as follows. In Sect. 2 we present the epistemic model that will be used throughout the paper. In Sect. 3 we present our epistemic characterization of quasi-perfect equilibrium and define the concept of quasi-perfect rationalizability. The relationship to proper rationalizability is also investigated. In Sect. 4 we then do the corresponding analysis for sequential equilibrium/rationalizability, and show that sequential (and thus quasi-perfect and proper) rationalizability implies backward induction. In Sect. 5 we show by means of examples that the inclusions indicated by the arrows in Table 2 are strict. We also show that there is in general no inclusion between sequential and quasi-perfect rationalizability on the one hand and EFR on the other (at least in strategy space). Finally, in Sect. 6 we discuss whether there are algorithms for these rationalizability concepts. We also briefly relate our analysis to some relevant literature.

2. States, Types, Preferences, and Belief

The purpose of this section is to present a framework for extensive games where each player is modeled as a decision-maker under uncertainty. The decision-theoretic analysis builds on Blume, Brandenburger & Dekel [13]. The framework is summarized by the concept of a *belief system* (cf. Def. 1). Appendix A contains a presentation of the decision-theoretic terminology, notation and results that will be utilized.

---

works as intended. Their corrigendum was written independently of the material on 'mixtures' in Sect. 6.

[7]The difference between 'weak extensive form rationalizability' and EFR can be seen to be analogous to the difference between 'permissibility' and the procedure of iterated (maximal) elimination of weakly dominated strategies.

2.1. **An Extensive Game Form.** A finite extensive game form (without chance moves) includes a set of players $I = \{1, 2\}$ and a set of terminal nodes Z. For each player i, there is a finite collection of information sets H_i, with a finite set of actions $A(h)$ being associated with each $h \in H_i$. A *pure strategy* for player i is a function that to any $h \in H_i$ assigns an action in $A(h)$. Let S_i denote player i's finite set of pure strategies, and let $S = S_1 \times S_2$. Let $z : S \to Z$ map strategy vectors into terminal nodes (or outcomes), and refer to $((S_i)_{i \in I}, z)$ as the associated strategic game form. For any $h \in H_1 \cup H_2$ and any node $x \in h$, let $S(x) = S_1(x) \times S_2(x)$ denote the set of pure strategy vectors for which x is reached, and write $S(h) := \bigcup_{x \in h} S(x)$. Since perfect recall is assumed, $S(h) = S_1(h) \times S_2(h)$. For any h, $h' \in H_i$, h (weakly) precedes h' if and only if $S(h) \supseteq S(h')$. For any $h \in H_i$ and $a \in A(h)$, write $S_i(h, a) := \{s_i \in S_i(h) | s_i(h) = a\}$.

A *behavioral strategy* for player i, $\sigma_i = (\sigma_i(h))_{h \in H_i}$, is a function that to any $h \in H_i$ assigns a probability distribution in $\Delta(A(h))$. If $h \in H_i$, let $\sigma_i|_h$ denote the behavioral strategy with the following properties: (1) at player i information sets preceding h, $\sigma_i|_h$ determines with probability one the unique action leading to h, and (2) at all other player i information sets, $\sigma_i|_h$ coincides with σ_i. Say that σ_i is *outcome-equivalent* to a mixed strategy p_i ($\in \Delta(S_i)$) if, for any $s_j \in S_j$, σ_i and p_i induce the same probability distribution over terminal nodes. For any $h \in H_i$, $\sigma_i|_h$ is outcome-equivalent to some $p_i \in \Delta(S_i(h))$.

2.2. **States and Types.** When a strategic game form is turned into a decision problem for each player (see Tan & Werlang [40]), the uncertainty faced by a player concerns the strategy choice of his opponent, the belief of his opponent about the player's own strategy choice, and so on. A type of a player corresponds to a vNM utility function and a belief about the strategy choice of his opponent, a belief about the belief of his opponent about his own strategy choice, and so on. Models of such infinite hierarchies of beliefs (Böge & Eisele [15], Mertens & Zamir [30], Brandenburger & Dekel [18], Epstein & Wang [23]) yield $S \times T$ as the complete state space, where $T = T_1 \times T_2$ is the set of all feasible type vectors. Furthermore, for each i, there is a homeomorphism between T_i and the set of beliefs on $S \times T_j$, where j denotes i's opponent.

For each type of any player i, the type's decision problem is to choose one of i's strategies. For the modeling of this problem, the type's belief about his own strategy choice is not relevant and can be ignored. Hence, in the setting of a strategic game form the beliefs can be restricted to the set of opponent strategy-type pairs, $S_j \times T_j$. Combined with a vNM utility function, the set of beliefs on $S_j \times T_j$ corresponds to a set of "regular" binary relations on the set of acts on $S_j \times T_j$, where an *act* on $S_j \times T_j$ is a function that to any element of $S_j \times T_j$ assigns an objective randomization on Z.

In conformity with the literature on infinite hierarchies of beliefs, let

- the set of *states of the world* (or simply *states*) be $\Omega := S \times T$,
- each *type* t_i of any player i correspond to a binary relation $\succeq^{t_i}$ on the set of acts on $S_j \times T_j$.

However, as the above results on infinite hierarchies of beliefs are not applicable in the present setting, we instead consider an implicit model – with a finite type set T_i for each player i – from which infinite hierarchies of beliefs can be constructed. Moreover, since continuity is not imposed, the "regularity" conditions on $\succeq^{t_i}$ consist of *completeness*, *transitivity*, *objective independence*, *nontriviality*, *conditional* *continuity* and *non-null state independence*, meaning that $\succeq^{t_i}$ is represented by a vNM utility function $\upsilon_i^{t_i} : Z \to \mathbb{R}$ that assigns a payoff to any outcome and a lexicographic probability system (LPS) $\lambda^{t_i} = (\mu_1^{t_i}, \ldots, \mu_{L^{t_i}}^{t_i}) \in \Delta(S_j \times T_j)$ (cf. Blume et al. [13] and Appendix A). Being a vNM utility function, $\upsilon_i^{t_i}$ can be extended to objective randomizations on Z. If $E_j \subseteq S_j \times T_j$ and $\succeq_{E_j}^{t_i}$ is nontrivial, let $\ell^{t_i}(E_j) := \min\{\ell | \mu_\ell^{t_i}(E_j) > 0\}$.

The construction is summarized by the following definition.

Definition 1. A *belief system* for a game form $((S_i)_{i \in I}, z)$ consists of

-
- for each player i, a finite set of types T_i,
- for each type t_i of any player i, a binary relation $\succeq^{t_i}$ (t_i's *preferences*) on the set of acts on $S_j \times T_j$, where $\succeq^{t_i}$ is represented by a vNM utility function $\upsilon_i^{t_i}$ on the set of objective randomizations on Z and an LPS λ^{t_i} on $S_j \times T_j$.

2.3. **Certain Belief.** For each player i, i's certain belief can be derived from the belief system. To state this epistemic operator, let, for each player i and each state $\omega \in \Omega$, $t_i(\omega)$ denote the projection of ω on T_i, and let, for any $E \subseteq \Omega$, $E_j^{t_i} := \{(s_j, t_j) \in S_j \times T_j | \exists s_i \text{ s.t. } (s_i, s_j, t_i, t_j) \in E\}$ denote the set of opponent strategy-type pairs that are consistent with $\omega \in E$ and $t_i(\omega) = t_i$. Associate 'certain belief' of an event with the property that no element of the complement of the event is assigned positive probability by some probability distribution in λ^{t_i}:

$$K_i E := \{\omega \in \Omega | \kappa_j^{t_i(\omega)} \subseteq E_j^{t_i(\omega)}\},$$

where $\kappa_j^{t_i} := \text{supp}\lambda^{t_i}$ ($\subseteq S_j \times T_j$) denotes the set of opponent strategy-type pairs that t_i does not deem *Savage-null*.[8] Say that at ω, i certainly believes the event $E \subseteq \Omega$ if $\omega \in K_i E$ (or equivalently, $\kappa_j^{t_i(\omega)} \subseteq E_j^{t_i(\omega)}$). Write $KE := K_1 E \cap K_2 E$. Say that there is mutual certain belief of $E \subseteq \Omega$ at ω if $\omega \in KE$. Write $CKE := KE \cap KKE \cap KKKE \cap \ldots$. Say that there is common certain belief of $E \subseteq \Omega$ at ω if $\omega \in CKE$.

2.4. **Preferences Over Strategies.** Let $\succeq_{S_j}^{t_i}$ denote the *marginal* of $\succeq^{t_i}$ on S_j. A pure strategy $s_i \in S_i$ can be viewed as an act $\mathbf{x}_{S_j}$ on S_j that assigns $z(s_i, s_j)$ to any $s_j \in S_j$. Hence, $\succeq_{S_j}^{t_i}$ is a binary relation also on the subset of acts on S_j that correspond to i's pure strategies. Thus, $\succeq_{S_j}^{t_i}$ can be referred to as t_i's *preferences over i's pure strategies*. Since $\succeq^{t_i}$ is represented by a vNM utility function and an LPS, $\succeq_{S_j}^{t_i}$ shares these properties. Let

$$C_i^{t_i} := \{s_i \in S_i | \, s_i \succeq_{S_j}^{t_i} s_i' \text{ for all } s_i' \in S_i\}$$

denote t_i's set of most preferred strategies (i.e. t_i's *choice set*).

2.5. **Quasi-Perfect Best Response.** For each $h \in H_i$, let $\succeq_{S_j(h)}^{t_i}$ denote t_i's preferences over i's pure strategies conditional on $S_j(h)$. Let

$$C_i^{t_i}(h) := \{s_i \in S_i(h) | \, s_i \succeq_{S_j(h)}^{t_i} s_i' \text{ for all } s_i' \in S_i(h)\}$$

denote t_i's set of most preferred strategies conditional on $h \in H_i$. By the following lemma, if s_i is a most preferred strategy conditional on $h \in H_i$, then s_i is most preferred conditional on any $h' \in H_i$ that appears (weakly) after h and is reachable when i plays s_i.

[8] The notion 'certain belief' corresponds to what Morris [31] calls 'Savage-belief'.

Lemma 1. *If $s_i \in C_i^{t_i}(h)$, then $s_i \in C_i^{t_i}(h')$ for any $h' \in H_i$ with $s_i \in S_i(h') \subseteq S_i(h)$.*

Proof. Suppose that $s_i \in S_i(h') \backslash C_i^{t_i}(h')$. Then there exists $s_i' \in S_i(h')$ such that $s_i' \succ_{S_j(h')}^{t_i} s_i$. It follows from Mailath, Samuelson & Swinkels ([28], Defs. 2 and 3 and the 'if' part of Theorem 1) that $S(h')$ is a *strategic independence* for i. Hence, s_i' can be chosen such that $z(s_i', s_j) = z(s_i, s_j)$ for all $s_j \in S_j \backslash S_j(h')$. This implies that $s_i' \succ_{S_j(h)}^{t_i} s_i$, which means that $s_i \in S_i(h) \backslash C_i^{t_i}(h)$. □

If follows from Lemma 1 that

$$C_i^{t_i}(h) = \left\{ s_i \in S_i(h) \middle| \begin{array}{l} s_i \succeq_{S_j(h')}^{t_i} s_i' \text{ for all } s_i' \in S_i(h') \text{ whenever} \\ h' \in H_i \text{ satisfies } s_i \in S_i(h') \subseteq S_i(h) \end{array} \right\} .$$

Note that $C_i^{t_i}(h)$ is well-defined even if $\succeq_{S_j(h')}^{t_i}$ is not nontrivial for some h' satisfying $s_i \in S_i(h') \subseteq S_i(h)$. Say that the behavioral strategy σ_i is a *quasi-perfect best response* to $\succeq^{t_i}$ if, for each $h \in H_i$, $\sigma_i|_h$ is outcome-equivalent to some mixed strategy in $\Delta(C_i^{t_i}(h))$.

2.6. **Induced Behavioral Strategy.** If $h \in H_j$ satisfies that $\succeq_{S_j(h) \times \{t_j\}}^{t_i}$ — the conditional of $\succeq^{t_i}$ on $S_j(h) \times \{t_j\}$ — is nontrivial, write for any $s_j \in S_j(h)$,

$$\mu_{S_j(h)}^{(t_i|t_j)}(s_j) := \frac{\mu_\ell^{t_i}(s_j, t_j)}{\mu_\ell^{t_i}(S_j(h) \times \{t_j\})} ,$$

with $\ell = \ell^{t_i}(S_j(h) \times \{t_j\})$, where $\mu_{S_j(h)}^{(t_i|t_j)}$ is t_i's (first-order) belief over t_j's strategies conditional on $h \in H_j$ being reached. Write for any $a \in A(h)$,

$$\sigma_j^{(t_i|t_j)}(h)(a) := \mu_{S_j(h)}^{(t_i|t_j)}(S_j(h, a)) ,$$

implying that $\sigma_j^{(t_i|t_j)}(h) \in \Delta(A(h))$. Say that the behavioral strategy $\sigma_j^{(t_i|t_j)} = (\sigma_j^{(t_i|t_j)}(h))_{h \in H_j}$ *is induced for t_j by λ^{t_i}* if $\succeq_{S_j(h) \times \{t_j\}}^{t_i}$ is nontrivial for all $h \in H_j$.

2.7. **An Extensive Game.** Consider an extensive game form (cf. Sect. 2.1), and let, for each i, $\upsilon_i : Z \to \mathbb{R}$ be a vNM utility function that assigns a payoff to any outcome. Then the pair of the extensive game form and the vNM utility functions $(\upsilon_i)_{i \in I}$ is a finite *extensive game*, Γ. Let $G = (S_i, u_i)_{i \in I}$ be the corresponding finite *strategic game*, where for each i, the vNM utility function $u_i : S \to \mathbb{R}$ is defined by $u_i = \upsilon_i \circ z$ (i.e., $u_i(s) = \upsilon_i(z(s))$ for any $s = (s_1, s_2) \in S$). Assume that, for each i, there exist s, $s' \in S$ such that $u_i(s) > u_i(s')$.

The event that i plays the game G is given by

$$[u_i] := \{\omega \in \Omega | \upsilon_i^{t_i(\omega)} \circ z \text{ is a positive affine transformation of } u_i\} ,$$

while $[u_1] \cap [u_2]$ is the event that both players play G.

3. Consistency of Preferences

Usually requirements in deductive game theory are imposed on choice. E.g. rationality is a requirement on a pair (s_i, t_i), where s_i is said to be a 'rational choice' by t_i if $s_i \in C_i^{t_i}$. See e.g. Epstein ([22], Sect. 6) for a presentation of this approach in a general context.

The present paper follows Asheim & Dufwenberg [4] by imposing requirements on t_i only. Since t_i corresponds to the preferences $\succeq^{t_i}$, such requirements will be imposed on $\succeq^{t_i}$. Here we will focus on showing how this 'consistent preferences' approach to deductive game theory can be used to

- characterize sequential (Kreps & Wilson's [26]) and quasi-perfect (van Damme [41]) equilibrium, and
- define *sequential* and *quasi-perfect rationalizability* as non-equilibrium analogues to the concepts of Kreps & Wilson and van Damme in two-player extensive games[9]

through the notions of *sequential* and *quasi-perfect consistency*. Moreover, the relationship to the concept of *proper rationalizability* will be investigated (cf. Schumacher [38] and Asheim [3]).

3.1. **Quasi-Perfect Consistency.** Quasi-perfect consistency will be based on three requirements: The first of these ensures that each player i plays the game G, the second requirement ensures that each player i is *cautious*, implying that i is of a type t_i that satisfies, for any $h \in H_1 \cup H_2$, that $\succeq^{t_i}_{S_j(h)\times\{t_j\}}$ is non-trivial whenever the opponent type t_j is taken into account, while the third requirement entails that the preferences of each player i *i*nduces a behavioral strategy that is a quasi-*p*erfect *b*est *r*esponse for any opponent type t_j that is taken into account.

To impose these requirements, consider the following events

$$[cau_i] := \{\omega \in \Omega | \kappa_j^{t_i(\omega)} = S_j \times T_j^{t_i(\omega)}\}$$
$$[ipbr_i] := \{\omega \in \Omega | \text{ if } h \in H_j,\ t_j \in T_j^{t_i(\omega)} \text{ and } s'_j \in S_j(h) \backslash C_j^{t_j}(h), \text{ then } \exists s_j \in S_j(h) \text{ s.t. } (s_j, t_j) \gg (s'_j, t_j) \text{ acc. to } \succeq^{t_i(\omega)}\},$$

where $T_j^{t_i} := \text{proj}_{T_j}\kappa_j^{t_i}$ denotes the set of opponent types that t_i takes into account (i.e. does not deem Savage-null), and where $\gg$ means 'infinitely more likely' (cf. Appendix A). The interpretation of $[ipbr_i]$ as inducement of a quasi-perfect best response for any opponent type that is taken into account follows from the following lemma.

Lemma 2. *If $\omega \in [cau_i]$, then $\omega \in [ipbr_i]$ is equivalent to $\sigma_j^{(t_i(\omega)|t_j)}$ being a quasi-perfect best response to $\succeq^{t_j}$ whenever $t_j \in T_j^{t_i(\omega)}$.*

Proof. If $\omega \in [cau_i]$, then, $\forall t_j \in T_j^{t_i(\omega)}$, $\mu_{S_j(h)}^{(t_i(\omega)|t_j)}$ is well-defined. Furthermore, given $\omega \in [cau_i]$, $\omega \in [ipbr_i]$ is equivalent to $\forall t_j \in T_j^{t_i(\omega)}$, $\forall h \in H_j$,

$$\text{supp}\mu_{S_j(h)}^{(t_i(\omega)|t_j)} \subseteq C_j^{t_j}(h),$$

which in turn is equivalent to $\forall t_j \in T_j^{t_i(\omega)}$, $\forall h \in H_j$,

$$\sigma_j^{(t_i(\omega)|t_j)}|_h \text{ is outcome-equivalent to some } p_i \in \Delta(C_j^{t_j}(h)).$$

This means that $\forall t_j \in T_j^{t_i(\omega)}$, $\sigma_j^{(t_i(\omega)|t_j)}$ — the behavioral strategy induced for t_j by $\lambda^{t_i(\omega)}$ — is a quasi-perfect best response to $\succeq^{t_j}$. □

Say that at ω, i is *quasi-perfectly consistent* (with the game G and the preferences of his opponent) if $\omega \in A_i^{q\text{-}p}$, where

$$A_i^{q\text{-}p} := [u_i] \cap [cau_i] \cap [ipbr_i].$$

Refer to $A^{q\text{-}p} := A_1^{q\text{-}p} \cap A_2^{q\text{-}p}$ as the event of *quasi-perfect consistency*.

[9]As mentioned in the introduction, an extension to games with more than two players raises the issue of independence, which will not be addressed here. As has also been mentioned there, the concept of a quasi-perfect equilibrium differs from Selten's [39] extensive form perfect equilibrium by the property that, at each information set, the player taking an action ignores the possibility of his own future mistakes.

We can now *characterize* the concept of a quasi-perfect equilibrium as vectors of induced behavioral strategies in states where there is quasi-perfect consistency and mutual certain belief of the type vector.[10]

Proposition 1. *Consider a finite extensive two-player game* Γ*. A vector of behavioral strategies* $\sigma = (\sigma_1, \sigma_2)$ *is a quasi-perfect equilibrium if and only if there exists a belief system and* $\omega \in A^{q\text{-}p}$ *such that (1) there is mutual certain belief of* $(t_1(\omega), t_2(\omega))$ *at* ω*, and (2) for each* $i \in I$*,* σ_i *is induced for* $t_i(\omega)$ *by* $\lambda^{t_j(\omega)}$*.*

Proof. See Appendix B. □

We can next *define* the concept of quasi-perfectly rationalizable behavioral strategies as quasi-perfect best responses in states where there is common certain belief of quasi-perfect consistency.

Definition 2. A behavioral strategy σ_i for i is *quasi-perfectly rationalizable* in a finite extensive two-player game Γ if there exists a belief system with σ_i being a quasi-perfect best response to $\succeq^{t_i(\omega)}$ for some $\omega \in CKA^{q\text{-}p}$.

It turns out that a behavioral strategy is quasi-perfectly rationalizable if it is part of a quasi-perfect equilibrium.

Proposition 2. *If the vector of behavioral strategies* $\sigma = (\sigma_1, \sigma_2)$ *is a quasi-perfect equilibrium in a finite extensive two-player game* Γ*, then, for each* i*,* σ_i *is quasi-perfect rationalizable.*

Proof. This is a straightforward consequence of Prop. 1 and Lemma 2. □

Since a quasi-perfect equilibrium always exists, we obtain the following corollary.

Corollary 1. *In any finite extensive two-player game* Γ*, there exists a belief system with* $CKA^{q\text{-}p} \neq \emptyset$*, implying that there exists, for each* i*, a nonempty set of quasi-perfectly rationalizable strategies.*

In the next subsections we follow Asheim [3] by introducing the concepts of proper consistency and proper rationalizable strategies and show that proper consistency implies quasi-perfect consistency.

3.2. **Proper Consistency.** Proper consistency is obtained from quasi-perfect consistency by additionally imposing that each player deems one opponent strategy to be infinitely more likely than another if the opponent prefers the one to the other (*respects preferences*). Hence, consider

$$[resp_i] := \{\omega \in \Omega |\ \text{if } t_j \in T_j^{t_i(\omega)} \text{ and } s_j \succ_{S_i}^{t_j} s'_j, \text{ then } (s_j, t_j) \gg (s'_j, t_j) \text{ acc. to } \succeq^{t_i(\omega)}\}\,.$$

Say that at ω, i is *properly consistent* (with the game G and the preferences of his opponent) if $\omega \in A_i^{pr}$, where

$$A_i^{pr} := [u_i] \cap [cau_i] \cap [resp_i].$$

Refer to $A^{pr} := A_1^{pr} \cap A_2^{pr}$ as the event of *proper consistency*.

The following result follows from Blume, Brandenburger and Dekel's ([14], Prop. 5) characterization of Myerson's [32] proper equilibrium in two-player games.

[10] The definition of a quasi-perfect equilibrium is given in Appendix B. There is mutual certain belief of the type vector $(t_1(\omega), t_2(\omega))$ given ω if and only if, for each i, $T_j^{t_i(\omega)} = \{t_j(\omega)\}$.

Proposition 3. *Consider a finite strategic two-player game G. A vector of mixed strategies $p = (p_1, p_2) \in \Delta(S_1) \times \Delta(S_2)$ is a proper equilibrium if and only if there exists a belief system and $\omega \in A^{pr}$ such that (1) there is mutual certain belief of $(t_1(\omega), t_2(\omega))$ at ω, and (2) for each $i \in I$, and for any $s_i \in S_i$, $p_i(s_i) = \mu_1^{t_j(\omega)}(s_i, t_i(\omega))$.*

Proof. Cf. the proof of Prop. 1 in Asheim [3]. □

Furthermore, as proposed in Asheim [3] the concept of properly rationalizable strategies can be defined as most preferred strategies in states where there is common certain belief of proper consistency.[11]

Definition 3. A pure strategy s_i for i is *properly rationalizable* in a finite strategic two-player game G if there exists a belief system with $s_i \in C_i^{t_i(\omega)}$ for some $\omega \in CKA^{pr}$.

Any strategy used with positive probability in a proper equilibrium is properly rationalizable.

Proposition 4. *If $p = (p_1, p_2) \in \Delta(S_1) \times \Delta(S_2)$ is a proper equilibrium in a finite strategic two-player game G, then, for each i, any $s_i \in suppp_i$ is properly rationalizable.*

Proof. Cf. the proof of Prop. 2 in Asheim [3]. □

Since a proper equilibrium always exists, we obtain the following corollary.

Corollary 2. *In any finite strategic two-player game G, there exists a belief system with $CKA^{pr} \neq \emptyset$, implying that there exists, for each i, a nonempty set of properly rationalizable strategies.*

3.3. **The Relation Between Quasi-Perfect and Proper Consistency.** As was established by van Damme [41], any proper equilibrium in the strategic form corresponds to a quasi-perfect equilibrium in the extensive form. Using Blume et al.'s [14] characterization of proper equilibrium, this is accomplished by inducing a vector of behavioral strategies, cf. subsect. 2.6 above. In this sense, the concept of proper equilibrium is stronger than the concept of quasi-perfect equilibrium.

It is desirable to show that the corresponding relationships hold between the concepts of proper rationalizability and quasi-perfect rationalizability. This will be established below by showing, for any belief system and for each player i, that i respecting opponent preferences implies that the preferences of player i induces a behavioral strategy that is a quasi-perfect best response for any opponent type that is taken into account. This, by Props. 1 and 3, shows the relationship between the equilibrium concepts and, by Defs. 2 and 3, shows the relationship between the rationalizability concepts.

Proposition 5. *For any belief system and for each player i, $[resp_i] \subseteq [ipbr_i]$.*

To prove Prop. 5 we need the following lemma.

Lemma 3. *For any $h \in H_j$, if $s'_j \in S_j(h) \backslash C_j^{t_j}(h)$, then $\exists s_j$ s.t. $s_j \succ_{S_i}^{t_j} s'_j$.*

[11] It is shown in Asheim ([3], Prop. 3) that this definition of proper rationalizability corresponds to the one originally proposed by Schuhmacher [38].

Proof. Suppose that $s'_j \in S_j(h) \backslash C_j^{t_j}(h)$. Then there exists $s_j \in S_j(h)$ such that $s_j \succ_{S_i(h)}^{t_j} s'_j$. Since $S(h)$ is a strategic independence for j (cf. proof of Lemma 1), it follows that s_j can be chosen such that $z(s_j, s_i) = z(s'_j, s_i)$ for all $s_i \in S_i \backslash S_i(h)$. This implies that $s_j \succ_{S_i}^{t_j} s'_j$. □

Proof of Prop. 5. Consider any belief system with $\omega \in [resp_i]$. Assume that $h \in H_j$, $s'_j \in S_j(h) \backslash C_j^{t_j}(h)$, and $t_j \in T_j^{t_i(\omega)}$. By Lemma 3, $\exists s_j \in S_j(h)$ s.t. $s_j \succ_{S_i}^{t_j} s'_j$. Since $\omega \in [resp_i]$ and $t_j \in T_j^{t_i(\omega)}$, it follows that $\exists s_j \in S_j(h)$ s.t. $(s_j, t_j) \gg (s'_j, t_j)$ acc. to $\succeq^{t_i(\omega)}$. This shows that $\omega \in [ipbr_i]$ whenever $\omega \in [resp_i]$. □

Following the proof of Prop. 1 in Mailath, Samuelson & Swinkels [29] one can show that quasi-perfect rationalizability in every extensive form corresponding to a given strategic game coincides with proper rationalizability in that game.

Remark 1. Substitute the event

$$B_i[rat_j] := \{\omega \in \Omega |\ \text{if } t_j \in T_j^{t_i(\omega)} \text{ and } s'_j \in S_j \backslash C_j^{t_j}, \text{ then}$$
$$\exists s_j \in S_j \text{ s.t. } (s_j, t_j) \gg (s'_j, t_j) \text{ acc. to } \succeq^{t_i(\omega)}\}.$$

for $[ipbr_i]$. Write $A_i := [u_i] \cap [cau_i] \cap B_i[rat_j]$ and $A := A_1 \cap A_2$. Then a permissible strategy s_i can be characterized by the property that there exists a belief system with $s_i \in C_i^{t_i(\omega)}$ for some $\omega \in CKA$ (cf. [4], Prop. 5.1). Since $[ipbr_i] \subseteq B_i[rat_j]$, it follows from Def. 2 that quasi-perfect rationalizability refines permissibility as indicated in Table 2.

4. Sequentiality and Backward Induction

Given that we consider a set-up where there is mutual or certain belief of the event that players are cautious, it seems natural to consider players with preferences that induces a quasi-perfect best response for any opponent type that is taken into account. Informally, this entails that any player believes not only that the opponent chooses rationally at any information set, but also that she is cautious.

As we show below, we can still characterize sequential equilibrium and define sequential rationalizability within our framework by having players believe that their opponent considers only the induced first-order probability distribution at any information set. We thereby model players with preferences that induces a sequential best response for any opponent type, and where the choice of a sequential best response entails rational choice at any information set, but does *not* imply cautious behavior. This captures – as indicated in Table 2 – players that are not necessarily cautious, but who believe that the opponent chooses rationally at all information sets. The assumption of caution (implying that the marginal of a type's LPS on the opponent's strategy set has full support) is kept, but only as a technical requirement that enables the opponent's behavioral strategy and the player's own beliefs to be induced.

To present this analysis, we must first define the notion of a sequential best response, and then introduce the concept of sequential consistency as a weaker alternative to quasi-perfect consistency.

4.1. **Consistent Assessment.** The *beliefs* of player i, $\beta_i = (\beta_i(h))_{h \in H_i}$, is a function that to any $h \in H_i$ assigns a probability distribution over the nodes in h. An *assessment* $(\sigma, \beta) = ((\sigma_1, \sigma_2), (\beta_1, \beta_2))$ is *consistent* if there is a sequence $(\sigma^n, \beta^n)_{n \in \mathbb{N}}$

of assessments converging to (σ, β) such that for every n, σ^n is completely mixed and β^n is induced by σ^n using Bayes' rule.

4.2. **Induced Beliefs.** If $h \in H_i$ satisfies that $\succeq^{t_i}_{S_j(h)}$ — t_i's preferences over i's pure strategies conditional on $S_j(h)$ (cf. subsect. 2.5) — is nontrivial, write for any $s_j \in S_j(h)$,

$$\mu^{t_i}_{S_j(h)}(s_j) := \frac{\mu^{t_i}_\ell(\{s_j\} \times T_j)}{\mu^{t_i}_\ell(S_j(h) \times T_j)} \ ,$$

with $\ell = \ell^{t_i}(S_j(h) \times T_j)$, where $\mu^{t_i}_{S_j(h)}$ is t_i's (first-order) belief over opponent strategies conditional on $h \in H_i$ being reached. Write for any node $x \in h$,

$$\beta^{t_i}_i(h)(x) := \mu^{t_i}_{S_j(h)}(S_j(x)) \ ,$$

where $\beta^{t_i}_i(h)$ is t_i's (first-order) belief over nodes in h. Say that the beliefs $\beta^{t_i}_i = (\beta^{t_i}_i(h))_{h \in H_i}$ *are induced by* λ^{t_i} if $\succeq^{t_i}_{S_j(h)}$ is nontrivial for all $h \in H_i$.

The following result has independent interest and is used below for characterizing the concept of sequential equilibrium (cf. Prop. 7).[12]

Proposition 6. *Consider a finite extensive two-player game form. An assessment* $(\sigma, \beta) = ((\sigma_1, \sigma_2), (\beta_1, \beta_2))$ *is consistent if and only if there exists a belief system and* $\omega \in [cau_1] \cap [cau_2]$ *such that (1) there is mutual certain belief of* $(t_1(\omega), t_2(\omega))$ *at* ω*, (2) for each* $i \in I$, σ_i *is induced for* $t_i(\omega)$ *by* $\lambda^{t_j(\omega)}$*, and (3) for each* $i \in I$, β_i *is induced by* $\lambda^{t_i(\omega)}$.

Proof. The proof of this proposition is given in Appendix B. □

4.3. **Sequential Best Response.** Given some belief $\mu \in \Delta(S_j)$, write $u^{t_i}_i(s_i, \mu) := \sum_{s_j \in S_j} \mu(s_j) \upsilon^{t_i}_i(z(s_i, s_j))$ for expected payoff. If $h \in H_i$, let

$$\tilde{C}^{t_i}_i(h) = \left\{ s_i \in S_i(h) \,\middle|\, \begin{array}{l} u^{t_i}_i(s_i, \mu^{t_i}_{S_j(h')}) \geq u^{t_i}_i(s'_i, \mu^{t_i}_{S_j(h')}) \text{ for all } s'_i \\ \in S_i(h') \text{ whenever } h' \in H_i \text{ satisfies } s_i \in \\ S_i(h') \subseteq S_i(h) \text{ and } \succeq^{t_i}_{S_j(h')} \text{ is nontrivial} \end{array} \right\}$$

denote t_i's set of strategies that maximizes expected payoff conditional on any reachable information set for i that appears (weakly) after h. Note that $\tilde{C}^{t_i}_i(h)$ is well-defined even if $\succeq^{t_i}_{S_j(h')}$ is not nontrivial for some h' satisfying $s_i \in S_i(h') \subseteq S_i(h)$. Say that the behavioral strategy σ_i is a *sequential best response* to $\succeq^{t_i}$ if, for each $h \in H_i$, $\sigma_i|_h$ is outcome-equivalent to some mixed strategy in $\Delta(\tilde{C}^{t_i}_i(h))$.

[12]Several non-epistemic characterizations of consistent assessments have been presented. E.g. Battigalli [6] explores the relationship between consistency of assessments and strategic independence of conditional probability systems. It is shown that an assessment induced by a strategically independent conditional probability system is always consistent, and that the former is also a necessary condition for consistency if the game is with "observable deviators". Kolhlberg & Reny [25] characterize consistent assessments by means of relative probability systems, showing that weak independence of random variables and coordinate-wise exchangeability of collections of random vectors plays a key role. In Perea y Monsuwé, Jansen & Peters [35] an algebraic characterization of consistent assessments is given, in which the event that one action may be infinitely more likely than another is quantified by an additional parameter assigned to every action in the game. Such a parametrization thus yields a refinement of conditional and relative probability systems.

4.4. **Sequential Consistency.** Sequential consistency weakens quasi-perfect consistency by requiring that the preferences of each player i *i*nduces a behavioral strategy that is only a *s*equential (not quasi-perfect) *b*est *r*esponse for any opponent type t_j that is taken into account. Hence, consider

$$[isbr_i] := \{\omega \in \Omega | \text{ if } h \in H_j,\, t_j \in T_j^{t_i(\omega)} \text{ and } s'_j \in S_j(h)\backslash \tilde{C}_j^{t_j}(h), \text{ then } \exists s_j \in S_j(h) \text{ s.t. } (s_j, t_j) \gg (s'_j, t_j) \text{ acc. to } \succeq^{t_i(\omega)}\},$$

The interpretation of $[isbr_i]$ as inducement of a sequential best response for any opponent type that is taken into account follows from the following lemma.

Lemma 4. *If* $\omega \in [cau_i]$*, then* $\omega \in [isbr_i]$ *is equivalent to* $\sigma_j^{(t_i(\omega)|t_j)}$ *being a sequential best response to* $\succeq^{t_j}$ *whenever* $t_j \in T_j^{t_i(\omega)}$*.*

Proof. The proof is analogous to the proof of Lemma 2. □

Say that at ω, i is *sequentially consistent* (with the game G and the preferences of his opponent) if $\omega \in A_i^{seq}$, where

$$A_i^{seq} := [u_i] \cap [cau_i] \cap [isbr_i].$$

Refer to $A^{seq} := A_1^{seq} \cap A_2^{seq}$ as the event of *sequential consistency.*

We can now *characterize* the concept of a sequential equilibrium as vectors of induced behavioral strategies in states where there is sequential consistency and mutual certain belief of the type vector.[13]

Proposition 7. *Consider a finite extensive two-player game* Γ*. A vector of behavioral strategies* $\sigma = (\sigma_1, \sigma_2)$ *can be extended to a sequential equilibrium if and only if there exists a belief system and* $\omega \in A^{seq}$ *such that (1) there is mutual certain belief of* $(t_1(\omega), t_2(\omega))$ *at* ω*, and (2) for each* $i \in I$*,* σ_i *is induced for* $t_i(\omega)$ *by* $\lambda^{t_j(\omega)}$*.*

Proof. The proof of this proposition is given in Appendix B. □

We can next *define* the concept of sequentially rationalizable behavioral strategies as sequential best responses in states where there is common certain belief of sequential consistency.

Definition 4. A behavioral strategy σ_i for i is *sequentially rationalizable* in a finite extensive two-player game Γ if there exists a belief system with σ_i being a sequential best response to $\succeq^{t_i(\omega)}$ for some $\omega \in CKA^{seq}$.

It turns out that a behavior strategy is sequentially rationalizable if it is part of a vector of behavioral strategies that can be extended to a sequential equilibrium.

Proposition 8. *If* $\sigma = (\sigma_1, \sigma_2)$ *is a vector of behavioral strategies that can be extended to a sequential equilibrium in a finite extensive two-player game* Γ*, then, for each* i*,* σ_i *is sequentially rationalizable.*

Proof. This is a straightforward consequence of Prop. 7 and Lemma 4. □

Since a sequential equilibrium always exists, we obtain the following corollary.

Corollary 3. *In any finite extensive two-player game* Γ*, there exists a belief system with* $CKA^{seq} \neq \emptyset$*, implying that there exists, for each* i*, a nonempty set of sequentially rationalizable strategies.*

[13]The definition of a sequential equilibrium is given in Appendix B.

It is straightforward to establish that quasi-perfect consistency is stronger than sequential consistency.

Proposition 9. *For any belief system and for each player i, $[ipbr_i] \subseteq [isbr_i]$.*

To prove Prop. 9 we need the following lemma.

Lemma 5. *For any $h \in H_j$, $C_j^{t_j}(h) \subseteq \tilde{C}_j^{t_j}(h)$.*

Proof. Let $h \in H_j$. Then, $\forall h' \in H_j$ with $s_j \in S_j(h') \subseteq S_j(h)$ and $\succeq^{t_j}_{S_i h')}$ being nontrivial,

$$\sum_{s_i \in S_i} \mu^{t_j}_{S_i(h')}(s_i) \upsilon_j^{t_j}(z(s_j, s_i)) \geq \sum_{s_i \in S_i} \mu^{t_j}_{S_i(h')}(s_i) \upsilon_j^{t_i}(z(s'_j, s_i))$$

whenever $s_j \succeq^{t_j}_{S_i(h')} s'_j$. Hence, by Lemma 1, $C_j^{t_j}(h) \subseteq \tilde{C}_j^{t_j}(h)$. □

Proof of Prop. 9. The result follows from the definitions of $[ipbr_i]$ and $[isbr_i]$ since by Lemma 5, for any $h \in H_j$, $S_j(h) \backslash \tilde{C}_j^{t_j}(h) \subseteq S_j(h) \backslash C_j^{t_j}(h)$. □

Prop. 9 in conjunction with Props. 1 and 7 implies the well-known result that every quasi-perfect equilibrium is a sequential equilibrium, while Prop. 9 and Lemma 5 in conjunction with Defs. 2 and 4 imply that every quasi-perfectly rationalizable strategy is a sequentially rationalizable strategy.

4.5. **Backward Induction in Generic Perfect Information Games.** We end this section by showing how sequential (and thus, by Props. 5 & 9 and Lemma 5, quasi-perfect and proper) rationalizability implies backward induction in perfect information games.

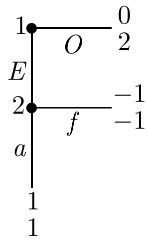
A finite extensive game Γ is of *perfect information* if, at any information set $h \in H_1 \cup H_2$, $h = \{x\}$; i.e. h contains only one node. It is *generic* if, for each i, $\upsilon_i(z) \neq \upsilon_i(z')$ whenever z and z' are different outcomes. Generic extensive games of perfect information have a unique subgame-perfect equilibrium. Moreover, in such games the procedure of backward induction yields in any subgame the unique subgame-perfect equilibrium outcome.

In a perfect information game, the action $a \in A(h)$ taken at the information set h determines the immediate succeeding information set, which can thus be denoted (h, a). Furthermore, any information set $h \in H_1 \cup H_2$ determines a subgame, implying that for each $i \in I$, and for any $h', h'' \in H_1 \cup H_2$ satisfying $S_i(h') \cap S_i(h'') \neq \emptyset$, it holds that $S_i(h') \subseteq S_i(h'')$ or $S_i(h') \supseteq S_i(h'')$. This means that $\tilde{C}_i^{t_i}(h)$ — t_i's set of strategies that maximizes expected payoff conditional on any reachable information set for i that appears (weakly) after h (cf. subsect. 4.2) — can be defined in an obvious way for all $h \in H_1 \cup H_2$. If $h \in H_j$, with $j \neq i$, it follows that

$$(1) \qquad \tilde{C}_i^{t_i}(h) = \bigcap\nolimits_{a \in A(h)} \tilde{C}_i^{t_i}(h, a) \text{ and, } \forall a \in A(h),\ S_i(h) = S_i(h, a)\ .$$

In the following proposition it is established that, for each subgame $h \in H_1 \cup H_2$, any vector of sequentially rationalizable strategies leads to the backward induction outcome in the subgame.

Proposition 10. *Consider a finite generic two-player extensive game of perfect information Γ with corresponding strategic game G. If, for some belief system, $\omega \in CKA^{seq}$, then, for each $h \in H_1 \cup H_2$, any mixed strategy vector $p = (p_1, p_2)$ where $p_i \in \Delta(\tilde{C}_i^{t(\omega)}(h))$ for each $i \in I$ leads to the backward induction outcome in the subgame h.*



	a	f
O	0, 2	0, 2
E	1, 1	-1,-1

FIGURE 1. Γ_1 and its strategic form.

Proof. The proof of this proposition is given in Appendix B. □

It is straightforward consequence of Props. 5 & 9 and Lemma 5 that, for each subgame $h \in H_1 \cup H_2$, any vector of quasi-perfect or properly rationalizable strategies leads to the backward induction outcome in the subgame.

Remark 2. Write $\tilde{C}_i^{t_i}$ for $\tilde{C}_i^{t_i}(h)$ if h is the initial node ($\tilde{C}_i^{t_i}$ is well-defined also for game forms without perfect information), and substitute the event

$$[iwbr_i] := \{\omega \in \Omega | \text{ if } t_j \in T_j^{t_i(\omega)} \text{ and } s'_j \in S_j \backslash \tilde{C}_j^{t_j}, \text{ then } \exists s_j \in S_j \text{ s.t. } (s_j, t_j) \gg (s'_j, t_j) \text{ acc. to } \succeq^{t_i(\omega)}\}.$$

for $[isbr_i]$. Write $A_i^{w-s} := [u_i] \cap [cau_i] \cap [iwbr_i]$ and $A^{w-s} := A_1^{w-s} \cap A_2^{w-s}$. Then a weakly extensive form rationalizable strategy s_i can be characterized by the property that there exists a belief system with $s_i \in C_i^{t_i(\omega)}$ for some $\omega \in CKA^{w-s}$. Since $[isbr_i] \subseteq [iwbr_i]$, it follows from Def. 4 that sequential rationalizability refines weak extensive form rationalizability as indicated in Table 2.

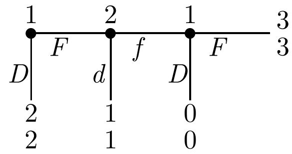
5. Examples

In this section we will offer examples that show that the inclusions in Table 2 are strict. We will also illustrate how sequential and quasi-perfect rationalizability differ from extensive form rationalizability (EFR) (cf. [34, 8, 10]).

The first example (Γ_1) is the well-known entry game where the entrant (player 1) can enter (E) or stay out (O), and the incumbent (player 2) can accommodate (a) or fight (f). Here rationalizability has no bite what-so-ever, while only E for 1 and a for 2 are rationalizable according to any of the other concepts in Table 2. This shows that the left arrow at the left of the table indicates a strict inclusion.

To show that the left arrows at the right indicate strict inclusions, consider Γ_2, which was introduced by Reny ([36], Fig. 1), and which has appeared in many contributions. The sets of permissible strategies equal $\{DD, DF, FF\}$ for 1 and $\{d, f\}$ for 2. This can easily be establish by applying the Dekel-Fudenberg procedure (cf. footnote 2). Since the game is generic, these are also the sets of weakly extensive form rationalizable strategies. However, only FF and f are implied by backward induction. Hence by Prop. 10, these are the only strategies that are quasi-perfectly/sequentially rationalizable.

Since, in simultaneous move games, quasi-perfect rationalizability coincides with permissibility and sequential rationalizability coincides with (ordinary) rationalizability, one example is sufficient to establish that the downward-pointing arrows in

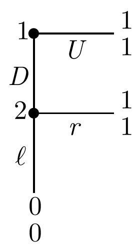


	d	f
DD	2, 2	2, 2
DF	2, 2	2, 2
FD	1, 1	0, 0
FF	1, 1	3, 3

FIGURE 2. Γ_2 and its strategic form.

	ℓ	c	r
U	1, 1	1, 1	1, 0
M	1, 1	2, 2	2, 2
D	0, 1	2, 2	3, 3

FIGURE 3. G_3



	ℓ	r
U	1, 1	1, 1
D	0, 0	1, 1

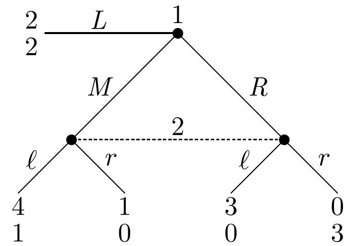
FIGURE 4. Γ_4 and its strategic form.

Table 2 indicate strict inclusions. The strategic game G_3 is due to Blume, Brandenburger & Dekel ([14], Fig. 1). Since there is no strong dominance, all strategies are rationalizable for both players, while the Dekel-Fudenberg procedure can be used to show that $\{M, D\}$ and $\{c, r\}$ are the set of permissible strategies. However, only D and r are properly rationalizable (cf. [3], Sect. 2).

It is instructive, though, to use games with a dynamic structure to illustrate that the downward-pointing arrows indicate strict inclusion. For the two downward-pointing arrows on the lower part of Table 2, consider Γ_4. Here the sets of weakly extensive form and sequentially rationalizable strategies are $\{U, D\}$ for 1 and $\{r\}$ for 2. However, if 1 takes into account the possibility that 2 may choose ℓ, then he strictly prefers U to D, implying that only U is permissible and quasi-perfectly rationalizable.

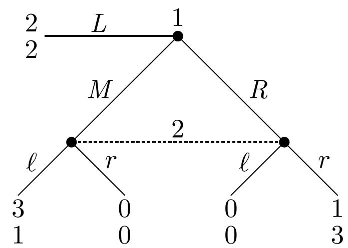
To show that the downward-pointing arrow on the upper part of Table 2, consider Γ_5. Here the sets of permissible and quasi-perfectly rationalizable strategies are $\{L, M\}$ for 1 and $\{\ell, r\}$ for 2. However, since M strongly dominates R, it follows that 2 prefers ℓ to r if she respects 1's preferences. Hence, only ℓ is properly rationalizable for 2, which in turn implies that only M is properly rationalizable for 1.

A slight variation of Γ_5 yields Γ_6, which is the 'Battle-of-the-Sexes-with-Outside-Option' game. For this game all rationalizability concepts of Table 2 yield the same conclusion: $\{L, M\}$ for 1 and $\{\ell, r\}$ for 2 are rationalizable independently of what



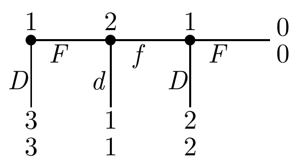
	ℓ	r
L	2, 2	2, 2
M	4, 1	1, 0
R	3, 0	0, 3

FIGURE 5. Γ_5 and its strategic form.



	ℓ	r
L	2, 2	2, 2
M	3, 1	0, 0
R	0, 0	1, 3

FIGURE 6. Γ_6 (= BoSwOO) and its strategic form.



	d	f
DD	3, 3	3, 3
DF	3, 3	3, 3
FD	1, 1	2, 2
FF	1, 1	0, 0

FIGURE 7. Γ_7 and its strategic form.

concept is being applied. The reason is that both (L, r) and (M, ℓ) are proper equilibria, implying by Prop. 4 that $\{L, M\}$ for 1 and $\{\ell, r\}$ for 2 are properly rationalizable, while – of course – the strongly dominated strategy R cannot be rationalizable according to any of the concepts. A forward induction argument would entail that 2 should deem M more likely than the strongly dominated strategy R, implying that she prefers ℓ to r. This in turn would imply that M would be the unique most preferred strategy for 1. EFR rationalizability yields this conclusion. Hence, Γ_6 demonstrates that the concepts defined here – sequential and quasi-perfect rationalizability – differs from EFR.

In Γ_6, any strategy that is rationalizable according to EFR is also sequentially and quasi-perfectly rationalizable, while the converse does not hold. Consider, however, Γ_7, which shows that the inclusion between the concepts may also go in the other direction. Since both sequential and quasi-perfect rationalizability imply the backward induction procedure (cf. Prop. 10), it follows that only DD and f are sequentially/quasi-perfectly rationalizable. In constrast, the sets of strategies rationalizable according to EFR are $\{DD, DF\}$ for 1 and $\{d, f\}$. In any case, the backward induction *outcome* is implied. It is still of interest to note that there are examples where any sequentially and quasi-perfectly rationalizable strategy is also rationalizable according to EFR, while the converse does not hold. Hence, at least in strategy space, it is not the case that EFR implies sequential/quasi-perfect rationalizability. Whether this can be established also in outcome space appears to be an open question.

6. Discussion

In this section we investigate the possibilities for constructing algorithms that lead to the set of sequential and quasi-perfect rationalizable behavioral strategies. We also briefly relate our analysis to some relevant literature.

The concepts of rationalizable and permissible strategies correspond to algorithms that iteratively eliminate strategies. I.e. rationalizability corresponds to the iterative elimination of strongly dominated strategies, or equivalently, strategies that do not constitute a best response to a mixture with support included in the set of opponent strategies that have not yet been eliminated. It would seem natural to conjecture that sequential and quasi-perfect rationalizability can be seen to correspond to algorithms where *behavioral strategies* are iteratively eliminated. However, such procedures are not straightforward as it is not necessarily clear what is meant by having a 'mixture' over behavioral strategies for the opponent.[14]

There seems to be two ways to define such a 'mixture' that will both lead to sequential and quasi-perfect rationalizability as defined here:

1. Each type of any player is endowed with a sequence of probability distributions on $S_j \times T_j$, where each distribution has the same support, and such that if some (s_j, t_j) pair is in the support of the distributions, so is (s'_j, t_j) for any $s'_j \neq s_j$. In the limit this sequence induces a behavioral strategy for each opponent type that is in the support of the distributions, and the limit of the marginals of the distributions on S_j induces the behavioral strategy that is the 'mixture'.
2. Each type of any player is endowed with an LPS on $S_j \times T_j$, such that if some (s_j, t_j) pair is in the support of the LPS, so is (s'_j, t_j) for any $s'_j \neq s_j$. Such an LPS induces a behavioral strategy for each opponent type that is in the support of the LPS, and the marginal of this LPS on S_j induces the behavioral strategy that is the 'mixture'.

Approach 1 has the advantage of being more "standard" (although perhaps more cumbersome). Furthermore, it facilitates the analysis of games with more than two players if one wants to impose that each player's belief about the strategy choices of other players are stochastically independent. Partly for this reason, such a "trembles" approach has independently been proposed and adopted by Dekel, Fudenberg & Levine [21] to correct the problem with their formalization of a 'mixture' of behavioral strategies in [20] (cf. Sect. 1). Approach 2, which is the approach chosen here, has the advantage of being simpler to use in a model of interactive epistemology as well as facilitating the comparison to proper rationalizability. Furthermore, it highlights that the behavioral strategy induced for a type of player j by the LPS of a type of player i, is not a chosen strategy by player j, but describes what i beliefs about j conditional on j being of this specific type.

[14] E.g. if a player believes that the opponent uses one of two behavioral strategies and these are held "to be equally likely", what does the player believe about the opponent's behavior at information sets that can only be reached if none of these strategies are being played?

Thus, the present analysis suggests an algorithm in terms of preferences that are presented by LPSs. In the epistemic model, types correspond to such LPSs. Furthermore, *common* certain belief of quasi-perfect consistency is the limit when the order of *mutual* certain belief of quasi-perfect consistency goes towards infinity. For a given order of mutual certain belief of quasi-perfect consistency, an LPS for a player on the set of opponent strategy-type vectors survives if (1) the opponent types that are taken into account correspond to LPSs that have not yet been eliminated at a lower order, and (2) for each such opponent type, the player's LPS induces a quasi-perfect best response given that type's LPS.

However, as the definition of quasi-perfect consistency makes clear (cf. subsect. 3.1), it is only the sets of most preferred strategies at the opponent's information sets that matter. Hence, an algorithm can be constructed in terms of functions from information sets to sets of most preferred strategies. Since the game is finite, the collection of this kind of functions is also finite, implying that such an algorithm converges in a finite number of rounds.

Indeed, for all three rationalizability concepts in the right column of Table 2, it appears that algorithms must work in terms of preferences. These can either be in terms of the LPSs that represent these preferences, or directly in terms of the binary relation over the player's own strategies that these preferences gives rise to. In any such binary relation for player i, preference corresponds to a subset of $S_i \times S_i$, where (s_i, s_i') being an element of this set means that s_i is preferred to s_i'. Each such subset in turn determines a unique function from information sets to sets of most preferred strategies. Again, the collection of this kind of subsets is finite, implying that an algorithm in terms of such sets converges in a finite number of rounds. It is, however, outside the scope of the present paper to investigate the properties of the different types of algorithms in more detail.

Here we have investigated polar cases, equilibrium concepts on the one hand, where it is assumed that each player is certain of the beliefs of his opponents about the player's own strategy choice, and rationalizability concepts on the other hand, for which no such assumption is being made. One can also consider intermediate concepts where e.g. each player is certain of the beliefs of his opponent about the player's own actions along the path that will be followed. Such possibilities, which have been subject to fruitful investigation by Dekel, Fudenberg & Levine [20] (cf. their concept of a 'sequential rationalizable self-confirming equilibrium') and Greenberg [24] (cf. his concept of a 'path mutually acceptable course of action'), also fall outside the set of issues that are treated in the present paper.

Appendix A. The Decision-Theoretic Framework

The purpose of this appendix is to present the decision-theoretic terminology, notation and results utilized and referred to in the main text.

Consider a decision-maker under uncertainty. Let F be a finite set of states, where the decision-maker is uncertain about what state in F will be realized. Let Z be a finite set of outcomes. In the tradition of Anscombe & Aumann [1], the decision-maker is endowed with a binary relation over all functions that to each element of F assigns an objective randomization on Z. Any such function $\mathbf{x}_F : F \to \Delta(Z)$ is called an *act* on F. Write $\mathbf{x}_F$ and $\mathbf{y}_F$ for acts on F. A *complete* and *transitive* binary relation on the set of acts on F is denoted by $\succeq_F$, where $\mathbf{x}_F \succeq_F \mathbf{y}_F$ means that $\mathbf{x}_F$ is *preferred* or *indifferent* to $\mathbf{y}_F$. As usual, let $\succ_F$ (*preferred to*) and $\sim_F$ (*indifferent to*) denote the asymmetric and symmetric parts of $\succeq_F$. A binary relation $\succeq_F$ on the set of acts on F is said to satisfy

- *objective independence* if $\mathbf{x}'_F \succ_F$ (respectively $\sim_F$) $\mathbf{x}''_F$ iff $\gamma\mathbf{x}'_F + (1-\gamma)\mathbf{y}_F \succ_F$ (respectively $\sim_F$) $\gamma\mathbf{x}''_F + (1-\gamma)\mathbf{y}_F$, whenever $0 < \gamma < 1$ and $\mathbf{y}_F$ is arbitrary.
- *nontriviality* if there exist $\mathbf{x}_F$ and $\mathbf{y}_F$ such that $\mathbf{x}_F \succ_F \mathbf{y}_F$.

If $E \subseteq F$, let $\mathbf{x}_E$ denote the restriction of $\mathbf{x}_F$ to E. Define the *conditional* binary relation $\succeq_E$ by $\mathbf{x}'_F \succeq_E \mathbf{x}''_F$ if, for arbitrary $\mathbf{y}_F$, $(\mathbf{x}'_E, \mathbf{y}_{-E}) \succeq_F (\mathbf{x}''_E, \mathbf{y}_{-E})$, where $-E$ denotes $F \backslash E$. Say that the state $f \in F$ is *Savage-null* if $\mathbf{x}_F \sim_{\{f\}} \mathbf{y}_F$ for all acts $\mathbf{x}_F$ and $\mathbf{y}_F$ on F. A binary relation $\succeq_F$ is said to satisfy

- *conditional continuity* if, $\forall f \in F$, there exist $0 < \gamma < \delta < 1$ such that $\delta\mathbf{x}'_F + (1-\delta)\mathbf{x}''_F \succ_{\{f\}} \mathbf{y}_F \succ_{\{f\}} \gamma\mathbf{x}'_F + (1-\gamma)\mathbf{x}''_F$ whenever $\mathbf{x}'_F \succ_{\{f\}} \mathbf{y}_F \succ_{\{f\}} \mathbf{x}''_F$.
- *non-null state independence* if $\mathbf{x}_F \succ_{\{e\}} \mathbf{y}_F$ iff $\mathbf{x}_F \succ_{\{f\}} \mathbf{y}_F$ whenever e and f are not Savage-null and $\mathbf{x}_F$ and $\mathbf{y}_F$ satisfy $\mathbf{x}_F(e) = \mathbf{x}_F(f)$ and $\mathbf{y}_F(e) = \mathbf{y}_F(f)$.

If e, $f \in F$, then e is deemed *infinitely more likely* than f ($e \gg f$) if e is not Savage-null and $\mathbf{x}_F \succ_{\{e\}} \mathbf{y}_F$ implies $(\mathbf{x}_{-\{f\}}, \mathbf{x}'_{\{f\}}) \succ_{\{e,f\}} (\mathbf{y}_{-\{f\}}, \mathbf{y}'_{\{f\}})$ for all $\mathbf{x}'_F$, $\mathbf{y}'_F$. According to this definition, f may, but need not, be Savage-null if $e \gg f$.

If $\upsilon : Z \to \mathbb{R}$ is a vNM utility function, abuse notation slightly by writing $\upsilon(p) = \sum_{z \in Z} p(z)\upsilon(z)$ whenever $p \in \Delta(Z)$ is an objective randomization. Say that $\mathbf{x}_E$ *strongly dominates* $\mathbf{y}_E$ w.r.t. υ if, $\forall f \in E$, $\upsilon(\mathbf{x}_E(f)) > \upsilon(\mathbf{y}_E(f))$. Say that $\mathbf{x}_E$ *weakly dominates* $\mathbf{y}_E$ w.r.t. υ if, $\forall f \in E$, $\upsilon(\mathbf{x}_E(f)) \geq \upsilon(\mathbf{y}_E(f))$, with strict inequality for some $e \in E$. Say that $\succeq_F$ is *admissible* on E ($\neq \emptyset$) if $\mathbf{x}_F \succ_F \mathbf{y}_F$ whenever $\mathbf{x}_E$ weakly dominates $\mathbf{y}_E$.

The following representation result due to Blume, Brandenburger & Dekel ([13], Theorem 3.1) can now be stated. It requires the notion of a *lexicographic probability system* (LPS) which consists of L levels of subjective probability distributions: If $L \geq 1$ and, $\forall \ell \in \{1, \ldots, L\}$, $\mu_\ell \in \Delta(F)$, then $\lambda = (\mu_1, ..., \mu_L)$ is an LPS on F. Let $\Delta(F)$ denote the set of LPSs on F, and let, for two utility vectors v and w, $v \geq_L w$ denote that, whenever $w_\ell > v_\ell$, there exists $\ell' < \ell$ such that $v_{\ell'} > w_{\ell'}$.

Proposition A1. *If $\succeq_F$ is complete and transitive, and satisfies objective independence, nontriviality, conditional continuity, and non-null state independence, then there exists a vNM utility function $\upsilon : Z \to \mathbb{R}$ and an LPS $\lambda = (\mu_1, ..., \mu_L) \in \Delta(F)$ such that ${}_F \succeq_F {}_F$ iff*

$$\left(\sum\nolimits_{f \in F} \mu_\ell(f)\upsilon({}_F(f))\right)_{\ell=1}^{L} \geq_L \left(\sum\nolimits_{f \in F} \mu_\ell(f)\upsilon({}_F(f))\right)_{\ell=1}^{L} .$$

If $F = F_1 \times F_2$ and $\succeq_F$ is a binary relation on the set of acts on F, then say that $\succeq_{F_1}$ is the *marginal* of $\succeq_F$ on F_1 if, $\mathbf{x}_{F_1} \succeq_{F_1} \mathbf{y}_{F_1}$ iff $\mathbf{x}_F \succeq_F \mathbf{y}_F$ whenever $\mathbf{x}_{F_1}(f_1) = \mathbf{x}_F(f_1, f_2)$ and $\mathbf{y}_{F_1}(f_1) = \mathbf{y}_F(f_1, f_2)$ for all (f_1, f_2).

Appendix B. Proofs of Propositions 1, 6, 7, and 10

Propositions 1, 6, and 7 are concerned with a state ω where there is mutual certain belief of the type vector $(t_1(\omega), t_2(\omega))$. This means that for each player i, $\text{supp}\lambda^{t_i(\omega)} \subseteq S_j \times \{t_j(\omega)\}$, where $\lambda^{t_i(\omega)} = (\mu_1^{t_i(\omega)}, \ldots, \mu_{L^{t_i}}^{t_i(\omega)}) \in \mathbf{L}\Delta(S_j \times T_j)$ represents $\succeq^{t_i(\omega)}$. For the proofs it is convenient to write $\lambda^i = (\mu_1^i, \ldots, \mu_{L^i}^i) \in \mathbf{L}\Delta(S_j)$, where $L^i = L^{t_i}$ and, $\forall \ell \in \{1, \ldots, L^i\}$, $\mu_\ell^i(s_j) = \mu_\ell^{t_i(\omega)}(s_j, t_j(\omega))$ for all $s_j \in S_j$. It follows that λ^i represents $\succeq_{S_j}^{t_i(\omega)}$. If $S'_j \subseteq S_j$ and $\succeq_{S'_j}^{t_i(\omega)}$ is nontrivial, let $\ell^i(S'_j) := \min\{\ell \mid \mu_\ell^i(S_j) > 0\}$. If for all $h \in H_j$, $\succeq_{S_j(h) \times \{t_j(\omega)\}}^{t_i(\omega)}$ and thus $\succeq_{S_j(h)}^{t_i(\omega)}$ are nontrivial, then say that λ^i *induces* σ_j if, $\forall h \in H_j$, $\forall a \in A(h)$,

$$\sigma_j(h)(a) = \frac{\mu_\ell^i(S_j(h,a))}{\mu_\ell^i(S_j(h))} ,$$

where $\ell = \ell^i(S_j(h))$. These definitions entail that λ^i inducing σ_j is equivalent to σ_j being induced for $t_j(\omega)$ by $\lambda^{t_i(\omega)}$. If for all $h \in H_i$, $\succeq_{S_j(h)}^{t_i(\omega)}$ are nontrivial, then say that λ^i

induces the beliefs β_i if, $\forall h \in H_i$, $\forall x \in h$,

$$\beta_i(h)(x) = \frac{\mu^i_\ell(S_j(x))}{\mu^i_\ell(S_j(h))} ,$$

where $\ell = \ell^i(S_j(h))$. These definitions entail that λ^i inducing β_i is equivalent to β_i being induced by $\lambda^{t_i(\omega)}$.

Define the concepts of a *behavioral representation* of a mixed strategy and the *mixed representation* of a behavioral strategy in the standard way (cf. e.g. [33], p. 159). If σ_j and p_j are both completely mixed, and σ_j is a behavorial representation of p_j or p_j is the mixed representation of σ_j, then, $\forall h \in H_j$, $\forall a \in A(h)$,

$$\sigma_j(h)(a) = \frac{p_j(S_j(h,a))}{p_j(S_j(h))} .$$

If p_j is a completely mixed mixed strategy and $h \in H_i$, let $p_j|_h$ be defined by

$$p_j|_h(s_j) = \begin{cases} \frac{p_j(s_j)}{p_j(S_j(h))} & \text{if } s_j \in S_j(h) \\ 0 & \text{otherwise} . \end{cases}$$

If σ_i is any behavioral strategy for i and σ_j is a *completely mixed* behavioral strategy for j, then abuse notation slightly by writing, for each $h \in H_i$,

$$u_i(\sigma_i, \sigma_j)|_h := u_i(p_i, p_j|_h) , \tag{2}$$

where p_i is outcome-equivalent to $\sigma_i|_h$ and p_j is the mixed representation of σ_j.

Definition B1. A behavioral strategy profile $\sigma = (\sigma_1, \sigma_2)$ is *quasi-perfect equilibrium* if there is a sequence $(\sigma^n)_{n\in\mathbb{N}}$ of completely mixed behavioral strategy profiles converging to σ such that for every $n \in \mathbb{N}$, $i \in I$ and $h \in H_i$,

$$u_i(\sigma_i, \sigma^n_j)|_h = \max_{\sigma'_i} u_i(\sigma'_i, \sigma^n_j)|_h . \tag{3}$$

If σ_i and σ_j are any behavioral strategies for i and j, and β_i is the beliefs of i, then let, for each $h \in H_i$, $u_i(\sigma_i, \sigma_j; \beta_i)|_h$ denote i's expected payoff conditional on h, given the belief $\beta_i(h)$ at h, and given that future behavior is determined by σ_i and σ_j.

Definition B2. An assessment $(\sigma, \beta) = ((\sigma_1, \sigma_2), (\beta_1, \beta_2))$ is a *sequential equilibrium* if it is consistent and it satisfies that for every $i \in I$ and $h \in H_i$,

$$u_i(\sigma_i, \sigma_j; \beta_i)|_h = \max_{\sigma'_i} u_i(\sigma'_i, \sigma_j; \beta_i)|_h .$$

For the proof of Prop. 1 we use two results from Blume, Brandenburger & Dekel ([14]; henceforth referred to as BBD). To state these results, we introduce the following notation. Let $\lambda = (\mu_1, ..., \mu_L)$ be an LPS on a finite set F and let $r = (r_1, ..., r_{L-1}) \in (0,1)^{L-1}$. Then, $r\Box\lambda$ denotes the probability distribution on F given by the nested convex combination

$$(1-r_1)\mu_1 + r_1\left[(1-r_2)\mu_2 + r_2\left[(1-r_3)\mu_3 + r_3\left[...\right]...\right]\right] .$$

Lemma B1 (Prop. 2 in BBD)**.** *Let $(p^n)_{n\in\mathbb{N}}$ be a sequence of probability distributions on a finite set F. Then, there exists a subsequence p^m of $(p^n)_{n\in\mathbb{N}}$, an LPS $\lambda = (\mu_1, ..., \mu_L)$ and a sequence r^m of vectors in $(0,1)^{L-1}$ converging to zero such that $p^m = r^m\Box\lambda$ for all m.*

The following lemma is a variant of Prop. 1 in BBD.

Lemma B2. *Consider a type t_i of player i whose preferences over acts on S_j are represented by υ_i (recalling from subsect. 2.7 that $u_i = \upsilon_i \circ z$) and $\lambda^i = (\mu^i_1, \dots, \mu^i_{L_i}) \in \mathbf{L}\Delta(S_j)$. Then, (a) $s_i \succ^{t_i}_{S_j} s'_i$ if and only if for every sequence $(r^n)_{n\in\mathbb{N}}$ in $(0,1)^{L_i-1}$ converging to zero there is a subsequence r^m such that*

$$\sum_{s_j} (r^m\Box\lambda^i)(s_j)u_i(s_i, s_j) > \sum_{s_j} (r^m\Box\lambda^i)(s_j)u_i(s'_i, s_j)$$

for all m, and (b) the same result would hold if the phrase "for every sequence.." is replaced by "for some sequence...".

Proof. (a) Suppose that $s_i \succ^{t_i}_{S_j} s'_i$. Then, there is some $k \in \{1, ..., L_i\}$ such that

$$\sum_{s_j} \mu^i_\ell(s_j) u_i(s_i, s_j) = \sum_{s_j} \mu^i_\ell(s_j) u_i(s'_i, s_j) \tag{4}$$

for all $\ell < k$ and

$$\sum_{s_j} \mu^i_k(s_j) u_i(s_i, s_j) > \sum_{s_j} \mu^i_k(s_j) u_i(s'_i, s_j). \tag{5}$$

Let $(r^n)_{n\in\mathbb{N}}$ be a sequence in $(0,1)^{L_i-1}$ converging to zero. By (4) and (5),

$$\sum_{s_j} (r^n \square \lambda^i)(s_j) u_i(s_i, s_j) > \sum_{s_j} (r^n \square \lambda^i)(s_j) u_i(s'_i, s_j)$$

if n is large enough. The other direction follows directly from the proof of Prop. 1 in BBD. The proof of part (b) follows from the proof of Prop. 1 in BBD. □

Proof of Proposition 1. (Only if.) Let (σ_1, σ_2) be a quasi-perfect equilibrium. By definition, there is a sequence $(\sigma^n)_{n\in\mathbb{N}}$ of completely mixed behavioral strategy profiles converging to σ such that for every $n \in \mathbb{N}$, $i \in I$ and $h \in H_i$,

$$u_i(\sigma_i, \sigma^n_j)|_h = \max_{\sigma'_i} u_i(\sigma'_i, \sigma^n_j)|_h \ .$$

For every n and $j \in I$, let p^n_j be the mixed representation of σ^n_j. By Lemma B1, the sequence $(p^n_j)_{n\in\mathbb{N}}$ of probability distributions on S_j contains a subsequence p^m_j such that we can find an LPS $\lambda^i = (\mu^i_1, \dots, \mu^i_{L^i})$ with full support on S_j and a sequence of vectors $r^m \in (0,1)^{L_i-1}$ converging to zero with

$$p^m_j = r^m \square \lambda^i$$

for all m. W.l.o.g., we assume that $p^n_j = r^n \square \lambda^i$ for all $n \in \mathbb{N}$.

First, we show that λ^i induces the behavioral strategy σ_j. Let $\tilde{\sigma}_j$ be the behavioral strategy induced by λ^i. By definition, $\forall h \in H_j$, $\forall a \in A(h)$,

$$\begin{aligned} \tilde{\sigma}_j(h)(a) &= \frac{\mu^i_\ell(S_j(h,a))}{\mu^i_\ell(S_j(h))} = \lim_{n\to\infty} \frac{(r^n \square \lambda^i)(S_j(h,a))}{(r^n \square \lambda^i)(S_j(h))} \\ &= \lim_{n\to\infty} \frac{p^n_j(S_j(h,a))}{p^n_j(S_j(h))} = \lim_{n\to\infty} \sigma^n_j(h)(a) = \sigma_j(h)(a) \ , \end{aligned}$$

where $\ell = \ell^i(S_j(h))$. For the fourth equation we used the fact that p^n_j is the mixed representation of σ^n_j. Hence, for each $i \in I$, λ^i induces σ_j.

Construct a belief system having a state space with a single type vector, $\Omega = S \times \{t_1\} \times \{t_2\}$, and where, for each $i \in I$, $\succeq^{t_i}$ is represented by υ_i and $\lambda^{t_i} = (\mu^{t_i}_1, \dots, \mu^{t_i}_{L^i}) \in \mathbf{L}\Delta(S_j \times \{t_j\})$, recalling from subsect. 2.7 that $u_i = \upsilon_i \circ z$, and letting λ^{t_i} be determined by, for each $\ell \in \{1, \dots, L^i\}$, $\mu^{t_i}_\ell(s_j, t_j) = \mu^i_\ell(s_j)$ for all $s_j \in S_j$. By construction, $\Omega = [u_1] \cap [u_2] \cap [cau_1] \cap [cau_2]$. Furthermore, for all $\omega \in \Omega$, there is mutual certain belief of the type vector (t_1, t_2) at ω.

By Lemma 2 it remains to be shown that σ_i is a quasi-perfect best response to $\succeq^{t_i}$ for each $i \in I$. Fix a player i and let $h \in H_i$ be given. Let p_i $(\in \Delta(S_i(h)))$ be outcome-equivalent to $\sigma_i|_h$ and let p^n_j be the mixed representation of σ^n_j. Then, since (σ_1, σ_2) is a quasi-perfect equilibrium, it follows from (2) that

$$u_i(p_i, p^n_j|_h) = \max_{p'_i \in \Delta(S_i(h))} u_i(p'_i, p^n_j|_h)$$

for all n. Hence, $p_i(s_i) > 0$ implies that

$$\sum_{s_j \in S_j(h)} p_j^n|_h(s_j) u_i(s_i, s_j) = \max_{s_i' \in S_i(h)} \sum_{s_j \in S_j(h)} p_j^n|_h(s_j) u_i(s_i', s_j) \tag{6}$$

for all n. Let $\lambda^i|_h$ be the conditional LPS on $S_j(h)$ induced by λ^i. By construction of the belief system, λ^i represents $\succeq_{S_j}^{t_i}$, and hence, it follows that $\lambda^i|_h$ represents $\succeq_{S_j(h)}^{t_i}$. Since $p_j^n = r^n \Box \lambda^i$ for all n there exist vectors $r^n|_h$ converging to zero such that $p_j^n|_h = r^n|_h \Box \lambda^i|_h$ for all n. Together with equation (6) we obtain that $p_i(s_i) > 0$ implies

$$\sum_{s_j \in S_j(h)} (r^n|_h \Box \lambda^i|_h)(s_j) u_i(s_i, s_j) = \max_{s_i' \in S_i(h)} \sum_{s_j \in S_j(h)} (r^n|_h \Box \lambda^i|_h)(s_j) u_i(s_i', s_j) \, . \tag{7}$$

We show that $p_i(s_i) > 0$ implies $s_i \in C_i^{t_i}(h)$. Suppose that $s_i \in S_i(h) \backslash C_i^{t_i}(h)$. Then, there is some $s_i' \in S_i(h)$ with $s_i' \succ_{S_j(h)}^{t_i} s_i$. From Lemma B2(a) it follows that $r^n|_h$ has a subsequence $r^m|_h$ for which

$$\sum_{s_j} (r^m|_h \Box \lambda^i|_h)(s_j) u_i(s_i', s_j) > \sum_{s_j} (r^m|_h \Box \lambda^i|_h)(s_j) u_i(s_i, s_j)$$

for all m, which is a contradiction to (7). Hence, $s_i \in C_i^{t_i}(h)$ whenever $p_i(s_i) > 0$, which implies that $p_i \in \Delta(C_i^{t_i}(h))$. Hence, $\sigma_i|_h$ is outcome equivalent to some $p_i \in \Delta(C_i^{t_i}(h))$. This holds for every $h \in H_i$, and hence, $\sigma_i|_h$ is a quasi-perfect best response to $\succeq^{t_i}$. This completes the first part of the proof.

(If) Suppose there exists a belief system and $\omega \in A^{q\text{-}p}$ such that (1) there is mutual certain belief of $(t_1(\omega), t_2(\omega))$ at ω, and (2) for each $i \in I$, σ_i is induced for $t_i(\omega)$ by $\lambda^{t_j(\omega)}$. We show that (σ_1, σ_2) is a quasi-perfect equilibrium.

For each i, derive $\lambda^i = (\mu_1^i, \dots, \mu_{L^i}^i) \in \mathbf{L}\Delta(S_j)$ from $\lambda^{t_i(\omega)} = (\mu_1^{t_i(\omega)}, \dots, \mu_{L^{t_i}}^{t_i(\omega)}) \in \mathbf{L}\Delta(S_j \times T_j)$ by letting $L^i = L^{t_i}$ and, $\forall \ell \in \{1, \dots, L^i\}$, $\mu_\ell^i(s_j) = \mu_\ell^{t_i(\omega)}(s_j, t_j(\omega))$ for all $s_j \in S_j$. Since $\omega \in [u_i]$, it follows that $\succeq_{S_j}^{t_i(\omega)}$ is represented by υ_i (recalling from subsect. 2.7 that $u_i = \upsilon_i \circ z$) and λ^i. Furthermore, λ^i induces σ_j. Choose sequences $(r^n)_{n \in \mathbb{N}}$ in $(0,1)^{L_i - 1}$ converging to zero and let the sequences $(p_j^n)_{n \in \mathbb{N}}$ of mixed strategies be given by $p_j^n = r^n \Box \lambda^i$ for all n. Since $\omega \in [cau_i]$, λ^i has full support and for every n, p_j^n is completely mixed. For every n, let σ_j^n be a behavioral representation of p_j^n. Since λ^i induces σ_j, it follows that $(\sigma_j^n)_{n \in \mathbb{N}}$ converges to σ_j; this is shown explicitly under the 'if' part of Prop. 6.

Fix a player i and an information set $h \in H_i$. We must show that

$$u_i(\sigma_i, \sigma_j^n)|_h = \max_{\sigma_i'} u_i(\sigma_i', \sigma_j^n)|_h \tag{8}$$

for all n, which implies that (σ_1, σ_2) is a quasi-perfect equilibrium.

Let p_i $(\in \Delta(S_i(h)))$ be outcome-equivalent to $\sigma_i|_h$. It follows from (2) that equation (8) is equivalent to

$$u_i(p_i, p_j^n|_h) = \max_{p_i' \in \Delta(S_i(h))} u_i(p_i', p_j^n|_h)$$

for all n. Hence, we must show that $p_i(s_i) > 0$ implies that

$$\sum_{s_j \in S_j(h)} p_j^n|_h(s_j) u_i(s_i, s_j) = \max_{s_i' \in S_i(h)} \sum_{s_j \in S_j(h)} p_j^n|_h(s_j) u_i(s_i', s_j) \tag{9}$$

for all n. In fact, it suffices to show this equation for infinitely many n, since in this case we can choose a subsequence for which the above equation holds, and this would be sufficient to show that (σ_1, σ_2) is a quasi-perfect equilibrium.

Since, by assumption, σ_i is a quasi-perfect best response to $\succeq^{t_i(\omega)}$, $\sigma_i|_h$ is outcome equivalent to some mixed strategy in $\Delta(C_i^{t_i(\omega)}(h))$. Hence, $p_i \in \Delta(C_i^{t_i(\omega)}(h))$. Let $p_i(s_i) >$

0. By construction, $s_i \in C_i^{t_i(\omega)}(h)$. Suppose that s_i would not satisfy (9) for infinitely many n. Then, there exists some $s_i' \in S_i(h)$ such that

$$\sum_{s_j \in S_j(h)} p_j^n|_h(s_j)u_i(s_i,s_j) < \sum_{s_j \in S_j(h)} p_j^n|_h(s_j)u_i(s_i',s_j)$$

for infinitely many n. Assume, w.l.o.g., that it is true for all n. Let $\lambda^i|_h$ be the conditional LPS on $S_j(h)$ induced by λ^i. Since λ^i represents $\succeq_{S_j}^{t_i(\omega)}$, it follows that $\lambda^i|_h$ represents $\succeq_{S_j(h)}^{t_i(\omega)}$. Since $p_j^n = r^n \Box \lambda^i$ for all n there exist vectors $r^n|_h$ converging to zero such that $p_j^n|_h = r^n|_h \Box \lambda^i|_h$ for all n. This implies that

$$\sum_{s_j} (r^n|_h \Box \lambda^i|_h)(s_j)u_i(s_i,s_j) < \sum_{s_j} (r^n|_h \Box \lambda^i|_h)(s_j)u_i(s_i',s_j)$$

for all n. By Lemma B2(b), it follows that $s_i' \succ_{S_j(h)}^{t_i(\omega)} s_i$, which is a contradiction to the fact that $s_i \in C_i^{t_i(\omega)}(h)$. Hence, $p_i(s_i) > 0$ implies (9) for infinitely many n, and as a consequence, (σ_1, σ_2) is a quasi-perfect equilibrium. □

Proof of Proposition 6. (Only if.) Let (σ, β) be consistent. Then, by definition, there is a sequence $(\sigma^n)_{n\in\mathbb{N}}$ of completely mixed behavioral strategy profiles converging to σ such that the sequence $(\beta^n)_{n\in\mathbb{N}}$ of induced belief systems converges to β. For each i and all n, let p_i^n be the mixed representation of σ_i^n. By Lemma B1, there exists for each i, an LPS $\lambda^i = (\mu_1^i, ..., \mu_{L^i}^i)$ with full support on S_j and a sequence $(r^n)_{n\in\mathbb{N}}$ in $(0,1)^{L^i-1}$ converging to zero such that $p_j^n = r^n \Box \lambda^i$ for all n.

We first show that λ^i induces the behavioral strategy σ_j. Let $\tilde{\sigma}_j$ be the behavioral strategy induced by λ^i. By definition, $\forall h \in H_j$, $\forall a \in A(h)$,

$$\begin{aligned} \tilde{\sigma}_j(h)(a) &= \frac{\mu_\ell^i(S_j(h,a))}{\mu_\ell^i(S_j(h))} = \lim_{n\to\infty} \frac{(r^n \Box \lambda^i)(S_j(h,a))}{(r^n \Box \lambda^i)(S_j(h))} \\ &= \lim_{n\to\infty} \frac{p_j^n(S_j(h,a))}{p_j^n(S_j(h))} = \lim_{n\to\infty} \sigma_j^n(h)(a) = \sigma_j(h)(a) \ , \end{aligned}$$

where $\ell = \ell^i(S_j(h))$. For the fourth equation we used the fact that p_j^n is the mixed representation of σ_j^n. Hence, for each $i \in I$, λ^i induces σ_j.

We then show that λ^i induces the beliefs β_i. Let $\tilde{\beta}_i$ be the player i beliefs induced by λ^i. By definition, $\forall h \in H_i$, $\forall x \in h$,

$$\begin{aligned} \tilde{\beta}_i(h)(x) &= \frac{\mu_\ell^i(S_j(x))}{\mu_\ell^i(S_j(h))} = \lim_{n\to\infty} \frac{r^n \Box \lambda^i(S_j(x))}{r^n \Box \lambda^i(S_j(h))} \\ &= \lim_{n\to\infty} \frac{p_j^n(S_j(x))}{p_j^n(S_j(h))} = \lim_{n\to\infty} \beta_i^n(h)(x) = \beta_i(h)(x), \end{aligned}$$

where $\ell = \ell^i(S_j(h))$. For the fourth equality we used the facts that p_j^n is the mixed representation of σ_j^n and β_i^n is induced by σ_i^n. Hence, for each $i \in I$, λ^i induces β_i.

Construct a belief system having a state space with a single type vector, $\Omega = S \times \{t_1\} \times \{t_2\}$, and where, for each $i \in I$, $\succeq^{t_i}$ is represented by $\lambda^{t_i} = (\mu_1^{t_i}, \ldots, \mu_{L^i}^{t_i}) \in \mathbf{L}\Delta(S_j \times \{t_j\})$, letting λ^{t_i} be determined by, for each $\ell \in \{1, \ldots, L^i\}$, $\mu_\ell^{t_i}(s_j, t_j) = \mu_\ell^i(s_j)$ for all $s_j \in S_j$. By construction, $\Omega = [cau_1] \cap [cau_2]$. Furthermore, for all $\omega \in \Omega$, there is mutual certain belief of the type vector (t_1, t_2) at ω. Hence, the 'only if' part of the proof is established since for each $i \in I$, λ^i induces σ_j and β_i.

(If) Suppose there exists a belief system and $\omega \in [cau_1] \cap [cau_2]$ such that (1) there is mutual certain belief of $(t_1(\omega), t_2(\omega))$ at ω, (2) for each $i \in I$, σ_i is induced for $t_i(\omega)$ by $\lambda^{t_j(\omega)}$, and (3) for each $i \in I$, β_i is induced by $\lambda^{t_i(\omega)}$. We show that (σ, β) is a consistent assessment.

For each i, derive $\lambda^i = (\mu_1^i, \ldots, \mu_{L^i}^i) \in \mathbf{L}\Delta(S_j)$ from $\lambda^{t_i(\omega)} = (\mu_1^{t_i(\omega)}, \ldots, \mu_{L^{t_i}}^{t_i(\omega)}) \in \mathbf{L}\Delta(S_j \times T_j)$ by letting $L^i = L^{t_i}$ and, $\forall \ell \in \{1, \ldots, L^i\}$, $\mu_\ell^i(s_j) = \mu_\ell^{t_i(\omega)}(s_j, t_j(\omega))$ for all $s_j \in S_j$. It follows that $\succeq_{S_j}^{t_i(\omega)}$ is represented by λ^i. Furthermore, λ^i induces σ_j and β_i.

Choose sequences $(r^n)_{n\in\mathbb{N}}$ in $(0,1)^{L_i-1}$ converging to zero and let the sequences $(p_j^n)_{n\in\mathbb{N}}$ of mixed strategies be given by $p_j^n = r^n\Box\lambda^i$ for all n. Since $\omega\in[cau_i]$, λ^i has full support and for every n, p_j^n is completely mixed. For every n, let σ_j^n be a behavioral representation of p_j^n and let β_i^n be the player i beliefs induced by σ_j^n. We show that $(\sigma_j^n)_{n\in\mathbb{N}}$ converges to σ_j and that $(\beta_i^n)_{n\in\mathbb{N}}$ converges to β_i, which imply consistency of (σ,β).

Since σ_j^n is a behavioral representation of p_j^n, we have, $\forall h\in H_j$, $\forall a\in A(h)$,

$$\sigma_j^n(h)(a) = \frac{p_i^n(S_j(h,a))}{p_j^n(S_j(h))}$$

for all n. Hence, since by assumption σ_j is induced by λ^i,

$$\begin{aligned}\lim_{n\to\infty}\sigma_j^n(h)(a) &= \lim_{n\to\infty}\frac{p_j^n(S_j(h,a))}{p_j^n(S_j(h))} = \lim_{n\to\infty}\frac{r^n\Box\lambda^i(S_j(h,a))}{r^n\Box\lambda^i(S_j(h))}\\ &= \frac{\mu_\ell^i(S_j(h,a))}{\mu_\ell^i(S_j(h))} = \sigma_j(h)(a),\end{aligned}$$

where $\ell = \ell^i(S_j(h))$. Hence, $(\sigma_j^n)_{n\in\mathbb{N}}$ converges to σ_j.

Since β_i^n is induced by σ_j^n and σ_j^n is a behavioral representation of p_j^n, we have, $\forall h\in H_i$, $\forall x\in h$,

$$\beta_i^n(h)(x) = \frac{p_j^n(S_j(x))}{p_j^n(S_j(h))}$$

for all n. Hence, since by assumption β_i is induced by λ^i,

$$\begin{aligned}\lim_{n\to\infty}\beta_i^n(h)(x) &= \lim_{n\to\infty}\frac{p_j^n(S_j(x))}{p_j^n(S_j(h))} = \lim_{n\to\infty}\frac{r^n\Box\lambda^i(S_j(x))}{r^n\Box\lambda^i(S_j(h))}\\ &= \frac{\mu_\ell^i(S_j(x))}{\mu_\ell^i(S_j(h))} = \beta_i(h)(x),\end{aligned}$$

where $\ell = \ell^i(S_j(h))$. Hence, $(\beta_i^n)_{n\in\mathbb{N}}$ converges to β_i, which completes the proof of the proposition. □

Proof of Proposition 7. (Only if.) Let (σ,β) be a sequential equilibrium. Since (σ,β) is consistent, we can follow the 'only if' part of the proof of Prop. 6 in constructing for each $i\in I$, an LPS $\lambda^i = (\mu_1^i,...,\mu_{L^i}^i)$ with full support on S_j. This in turn leads to a belief system having a state space with a single type vector, $\Omega = S\times\{t_1\}\times\{t_2\}$, and where, for each $i\in I$, $\succeq^{t_i}$ is represented by υ_i and $\lambda^{t_i} = (\mu_1^{t_i},\dots,\mu_{L^i}^{t_i})\in \mathbf{L}\Delta(S_j\times\{t_j\})$, recalling from subsect. 2.7 that $u_i = \upsilon_i\circ z$, and letting λ^{t_i} be determined by, for each $\ell\in\{1,\dots,L^i\}$, $\mu_\ell^{t_i}(s_j,t_j) = \mu_\ell^i(s_j)$ for all $s_j\in S_j$. By construction, $\Omega = [u_1]\cap[u_2]\cap[cau_1]\cap[cau_2]$. Furthermore, for all $\omega\in\Omega$, there is mutual certain belief of the type vector (t_1,t_2) given ω, and for each $i\in I$, σ_i is induced for t_i by λ^{t_j}.

By Lemma 4 it remains to be shown that σ_i is a sequential best response to $\succeq^{t_i}$ for each $i\in I$. Suppose that σ_i is not a sequential best response to $\succeq^{t_i}$. Then there is some information set $h\in H_i$ and some mixed strategy $p_i\in\Delta(S_i(h))$ that is outcome-equivalent to $\sigma_i|_h$ such that there exist $s_i\in S_i(h)$ with $p_i(s_i)>0$ and $s_i'\in S_i(h)$ having the property that

$$u_i(s_i,\mu_{S_j(h)}^i) < u_i(s_i',\mu_{S_j(h)}^i)\ ,$$

where, given some belief $\mu\in\Delta(S_j)$, $u_i(s_i,\mu) := \sum_{s_j\in S_j}\mu(s_j)\upsilon_i(z(s_i,s_j))$, and where for any $s_j\in S_j(h)$,

$$\mu_{S_j(h)}^i(s_j) := \frac{\mu_\ell^i(s_j)}{\mu_\ell^i(S_j(h))}$$

with $\ell = \ell^i(S_j(h))$. Since the beliefs β_i and the behavior strategy σ_j are induced by λ^i, it follows that $u_i(s_i, \mu^i_{S_j(h)}) = u_i(s_i, \sigma_j; \beta_i)|_h$ and $u_i(s'_i, \mu^i_{S_j(h)})|_h = u_i(s'_i, \sigma_j; \beta_i)|_h$ and hence

$$u_i(s_i, \sigma_j; \beta_i)|_h < u_i(s'_i, \sigma_j; \beta_i)|_h,$$

which is a contradiction to the fact that (σ, β) is sequentially rational.

(If) Suppose there exists a belief system and $\omega \in A^{seq}$ such that (1) there is mutual certain belief of $(t_1(\omega), t_2(\omega))$ at ω, and (2) for each $i \in I$, σ_i is induced for $t_i(\omega)$ by $\lambda^{t_j(\omega)}$.

For each i, derive $\lambda^i = (\mu^i_1, \ldots, \mu^i_{L^i}) \in \mathbf{L}\Delta(S_j)$ from $\lambda^{t_i(\omega)} = (\mu^{t_i(\omega)}_1, \ldots, \mu^{t_i(\omega)}_{L^{t_i}}) \in \mathbf{L}\Delta(S_j \times T_j)$ by letting $L^i = L^{t_i}$ and, $\forall \ell \in \{1, \ldots, L^i\}$, $\mu^i_\ell(s_j) = \mu^{t_i(\omega)}_\ell(s_j, t_j(\omega))$ for all $s_j \in S_j$. Since $\omega \in [u_i]$, it follows that $\succeq^{t_i(\omega)}_{S_j}$ is represented by υ_i (recalling from subsect. 2.7 that $u_i = \upsilon_i \circ z$) and λ^i. Furthermore, λ^i induces σ_j and β_i. Since $\omega \in [cau_1] \cap [cau_2]$ it follows from the 'if' part of Prop. 6 that (σ, β) is consistent.

Fix a player i and an information set $h \in H_i$. We must show that

$$u_i(\sigma_i, \sigma_j; \beta_i)|_h = \max_{\sigma'_i} u_i(\sigma'_i, \sigma_j; \beta_i)|_h \ , \tag{10}$$

which implies that $((\sigma_1, \sigma_2), (\beta_1, \beta_2))$ is a sequential equilibrium.

Suppose that $u_i(\sigma_i, \sigma_j; \beta_i)|_h < u_i(\sigma'_i, \sigma_j; \beta_i)|_h$ for some σ'_i. Let $p_i \in \Delta(S_i(h))$ be outcome-equivalent to $\sigma_i|_h$. Then, there is some $s_i \in S_i(h)$ with $p_i(s_i) > 0$ and some $s'_i \in S_i(h)$ such that

$$u_i(s_i, \sigma_j; \beta_i)|_h < u_i(s'_i, \sigma_j; \beta_i)|_h.$$

Since the beliefs β_i and the behavior strategy σ_j are induced by λ^i, it follows that $u_i(s_i, \sigma_j; \beta_i)|_h = u_i(s_i, \mu^i_{S_j(h)})$ and $u_i(s'_i, \sigma_j; \beta_i)|_h = u_i(s'_i, \mu^i_{S_j(h)})|_h$ (where we use the notation that has been introduced in the 'only if' part of this proof) and hence

$$u_i(s_i, \mu^i_{S_j(h)}) < u_i(s'_i, \mu^i_{S_j(h)}),$$

which contradicts the fact that σ_i is a sequential best response to $\succeq^{t_i}$. □

For the proof of Prop. 10 we must derive some properties of the certain belief operator (cf. subsect. 2.3). It is easy to check that $K_i\Omega = \Omega$ and $K_i\emptyset = \emptyset$, and, for any events E and F, $K_iE \cap K_iF = K_i(E \cap F)$, $K_iE \subseteq K_iK_iE$, and $\neg K_iE \subseteq K_i(\neg K_iE)$, implying that, for any event E, $K_iE = K_iK_iE$. Write $K^0E := E$ and, for each $g \geq 1$, $K^gE := KK^{g-1}E$. Since $K_i(E \cap F) = K_iE \cap K_iF$ and $K_iK_iE = K_iE$, it follows $\forall g \geq 2$, $K^gE = K_1K^{g-1}E \cap K_2K^{g-1}E \subseteq K_1K_1K^{g-2}E \cap K_2K_2K^{g-2}E = K_1K^{g-2}E \cap K_2K^{g-2}E = K^{g-1}E$. The truth axiom ($K_iE \subseteq E$) is not satisfied, since an event can be certainly believed even though the true state is an element of the complement of the event. However, since $A^{seq} = A^{seq}_1 \cap A^{seq}_2$ is an event that concerns the type vector, mutual certain belief of A^{seq} implies that A^{seq} is true: $KA^{seq} = K_1A^{seq} \cap K_2A^{seq} \subseteq K_1A^{seq}_1 \cap K_2A^{seq}_2 = A^{seq}_1 \cap A^{seq}_2 = A^{seq}$ since, for each i, $K_iA^{seq}_i = A^{seq}_i$. Hence, it follows that *(i)* $\forall g \geq 1$, $K^gA^{seq} \subseteq K^{g-1}A^{seq}$, and *(ii)* $\exists g' \geq 0$ such that $K^gA^{seq} = CKA^{seq}$ for $g \geq g'$ since Ω is finite.

For the proof of Prop. 10 we also need to establish more structure for perfect information games. Set $H^{-1} = Z$ (i.e. the set of terminal nodes) and determine, $\forall g \geq 0$, H^g as follows: $h \in H^g$ if and only if h satisfies that

$$g = 1 + \max\{g' | \exists h' \in H^{g'} \text{ and } a \in A(h) \text{ such that } h' = (h, a)\} \ .$$

In words, $h \in H^g$ if and only if g is the maximal number of decision nodes between h and a terminal node in the subgame determined by h.

Finally, let $S^{BI}(h) = S^{BI}_1(h) \times S^{BI}_2(h)$ ($\subseteq S(h)$) denote the set of pure strategy vectors that is consistent with the backward induction outcome in the subgame h.

Proof of Prop. 10. In view of properties of the certain belief operator (cf. the paragraph above), it is sufficient to show for any $g = 0, \ldots, \max\{g' | H^{g'} \neq \emptyset\}$ that if there exists a belief system with $\omega \in K^gA^{seq}$, then $\tilde{C}^{t(\omega)}_1(h) \times \tilde{C}^{t(\omega)}_2(h) \subseteq S^{BI}(h)$ for any $h \in H^g$. This is established by induction.

($g=0$) Let $h\in H^0$ and assume w.l.o.g. that $h\in H_i$. Then trivially $\tilde{C}_j^{t_j}(h)=S_j(h)=S_j^{BI}(h)$. Let $t_i=t_i(\omega)$ for some $\omega\in K^0A^{seq}=A^{seq}$. Then it follows that $\tilde{C}_i^{t_i}(h)=S_i^{BI}(h)$ since Γ is generic and $\omega\in A^{seq}\subseteq[u_i]\cap[cau_i]$.

($g=1,\dots,\max\{g'|H^{g'}\neq\emptyset\}$) Suppose that it has been established for $g'=0,\dots,g-1$ that if there exists a belief system with $\omega\in K^{g'}A^{seq}$, then $\tilde{C}^{t(\omega)}(h')\subseteq S^{BI}(h')$ for any $h'\in H^{g'}$. Let $h\in H^g$ and assume w.l.o.g. that $h\in H_i$. Let $t_j=t_j(\omega)$ for some $\omega\in K^{g-1}A^{seq}$. Then by (1) and the premise it follows that $S_j(h)=S_j(h,a)$ and $\tilde{C}_j^{t_j}(h)\subseteq\tilde{C}_j^{t_j}(h,a)\subseteq S_j^{BI}(h,a)$ if $a\in A(h)$. This implies that $\tilde{C}_j^{t_j}(h)\subseteq\bigcap_{a\in A(h)}S_j^{BI}(h,a)\subseteq S_j^{BI}(h)$. Hence, any $s_j\in\tilde{C}_j^{t_j}(h)$ is consistent with the backward induction outcome in any subgame (h,a) immediately succeeding h.

Now, consider i. Let $t_i=t_i(\omega)$ for some $\omega\in K^gA^{seq}$. The preceding argument implies that $\tilde{C}_j^{t_j}(h)\subseteq\bigcap_{a\in A(h)}S_j^{BI}(h,a)$ whenever $t_j\in T_j^{t_i}$ since $\omega\in K^gA^{seq}\subseteq K_iK^{g-1}A^{seq}$. Let $s_i^{BI}\in S_i(h)$ determine play in accordance with backward induction at any h' appearing (weakly) after h (i.e. at all h' satisfying $S_i(h)\supseteq S_i(h')$), and let $s_i'\in S_i(h)$ be a strategy that differs from s_i^{BI} by assigning a different action only at h. As any pure strategy in S_i can be viewed as an act on S_j (cf. subsect. 2.4), write $\mathbf{x}_{S_j}$ for the act on S_j that s_i^{BI} can be viewed as (i.e. $\mathbf{x}_{S_j}$ assigns $z(s_i^{BI},s_j)$ to any $s_j\in S_j$), and write $\mathbf{y}_{S_j}$ for the act on S_j that s_i' can be viewed as (i.e. $\mathbf{y}_{S_j}$ assigns $z(s_i',s_j)$ to any $s_j\in S_j$). Let $\mathbf{x}$ and $\mathbf{y}$ be the acts on $S_j\times T_j$ that satisfy $\mathbf{x}(s_j,t_j)=\mathbf{x}_{S_j}(s_j)$ and $\mathbf{y}(s_j,t_j)=\mathbf{y}_{S_j}(s_j)$ for all (s_j,t_j). Then,

$$\mathbf{x}_{\bigcap_{a\in A(h)}S_j^{BI}(h,a)\times T_j}\quad\text{strongly dominates}\quad\mathbf{y}_{\bigcap_{a\in A(h)}S_j^{BI}(h,a)\times T_j}$$

by backward induction since Γ is generic and $\omega\in K^gA^{seq}\subseteq[u_i]$. Since $\tilde{C}_j^{t_j}(h)\subseteq\bigcap_{a\in A(h)}S_j^{BI}(h,a)$ whenever $t_j\in T_j^{t_i}$, it follows that, $\forall t_j\in T_j^{t_i}$,

$$\mathbf{x}_{\tilde{C}_j^{t_j}(h)\times\{t_j\}}\quad\text{strongly dominates}\quad\mathbf{y}_{\tilde{C}_j^{t_j}(h)\times\{t_j\}}\,. \tag{11}$$

Since $\omega\in K^gA^{seq}\subseteq[isbr_i]$, it follows that there exists $s_j\in\tilde{C}_j^{t_j}(h)$ such that $(s_j,t_j)\gg(s_j',t_j)$ according to $\succeq^{t_i}$ whenever $t_j\in T_j^{t_i}$ and $s_j'\in S_j(h)\backslash\tilde{C}_j^{t_j}(h)$, which by (11) implies that $\mathbf{x}\succ^{t_i}_{S_j(h)\times\{t_j\}}\mathbf{y}$. Since this holds for all $t_j\in T_j^{t_i}$, it follows that $\mathbf{x}\succ^{t_i}_{S_j(h)\times T_j}\mathbf{y}$ and $\mathbf{x}_{S_j}\succ^{t_i}_{S_j(h)}\mathbf{y}_{S_j}$.

It has thereby been established that $s_i'\in S_i(h)\backslash\tilde{C}_i^{t_i}(h)$ if s_i' differs from backward induction only by the action taken at h. However, since $s_i\in\tilde{C}_i^{t_i}(h)$ implies $s_i\in\tilde{C}_i^{t_i}(h,s_i(h))$ (by definition) and, $\forall a\in A(h)$, $\tilde{C}_i^{t_i}(h,a)\subseteq S_i^{BI}(h,a)$ (by the premise), it follows that any $s_i\in\tilde{C}_i^{t_i}(h)$ is consistent with the backward induction outcome in the subgame $(h,s_i(h))$ immediately succeeding h. Hence, $\tilde{C}_i^{t_i}(h)\subseteq S_i^{BI}(h)$. □

References

1. Anscombe, F.J., Aumann, R.: A definition of subjective probability. Annals of Mathematical Statistics **34** (1963) 199–205.
2. Asheim, G.B.: On the epistemic foundation for backward induction. Memorandum No. 30, Dept. of Economics, University of Oslo. Revised (2000).
3. Asheim, G.B.: Proper consistency. Memorandum No. 31, Dept. of Economics, University of Oslo. Revised (2000).
4. Asheim, G.B., Dufwenberg, M.: Admissibility and common belief. Memorandum No. 7, Dept. of Economics, University of Oslo. Revised (2000).
5. Aumann, R.: Backward induction and common knowledge of rationality. Games Econ. Beh. **8** (1995) 6–19.
6. Battigalli, P.: Strategic independence and perfect Bayesian equilibria. J. Econ. Theory **70** (1996) 201–234.
7. Battigalli, P.: Strategic rationality orderings and the best rationalization principle. Games Econ. Beh. **13** (1996) 178–200.
8. Battigalli, P.: On rationalizability in extensive games. J. Econ. Theory **74** (1997) 40–61.

9. Battigalli, P., Bonanno, G.: Recent results on belief, knowledge and the epistemic foundations of game theory. Research in Econ. **53** (1999) 149–225.
10. Battigalli, P., Siniscalchi, M.: Interactive beliefs and forward induction. Princeton University (2000).
11. Ben-Porath, E.: Rationality, Nash equilibrium, and backwards induction in perfect information games. Rev. Econ. Stud. **64** (1997) 23–46.
12. Bernheim, D.: Rationalizable strategic behavior. Econometrica **52** (1984) 1007–1028.
13. Blume, L., Brandenburger, A., Dekel, E.: Lexicographic probabilities and choice under uncertainty. Econometrica **59** (1991) 61–79.
14. Blume, L., Brandenburger, A., Dekel, E.: Lexicographic probabilities and equilibrium refinements. Econometrica **59** (1991) 81–98.
15. Böge, W., Eisele, T.: On the solutions of Bayesian games. Int. J. Game Theory **8** (1979) 193–215.
16. Börgers, T.: Weak dominance and approximate common knowledge. J. Econ. Theory **64** (1994) 265–276.
17. Brandenburger, A.: Lexicographic probabilities and iterated admissibility. In "Economic Analysis of Markets and Games"
18. Brandenburger, A., Dekel, E.: Hierarchies of beliefs and common knowledge. J. Econ. Theory **59** (1993) 189–198.
19. Dekel, E., Fudenberg, D.: Rational behavior with payoff uncertainty. J. Econ. Theory **52** (1990) 243–267.
20. Dekel, E., Fudenberg, D., Levine, D.: Payoff information and self-confirming equilibrium. J. Econ. Theory **89** (1999) 165–185.
21. Dekel, E., Fudenberg, D., Levine, D.: Subjective uncertainty over behavior strategies: A correction. Harvard University (2000).
22. Epstein, L.G.: Preference, rationalizability and equilibrium. J. Econ. Theory **73** (1997) 1–29.
23. Epstein, L.G., Wang, T.: "Beliefs about beliefs" without probabilities. Econometrica **64** (1996) 1343–1373.
24. Greenberg, J.: Towering over Babel: Worlds apart – but acting together. McGill University (1996).
25. Kohlberg, E., Reny, P.J.: Independence on relative probability spaces and consistent assessments in game trees. J. Econ. Theory **75** (1997) 280–313.
26. Kreps, D.M., Wilson, R.: Sequential equilibrium. Econometrica **50** (1982) 863–894.
27. McLennan, A.: Consistent conditional systems in noncooperative game theory. Int. J. Game Theory **18** (1989) 141–174.
28. Mailath, G., Samuelson, L., Swinkels, J.: Extensive form reasoning in normal form games. Econometrica **61** (1993) 273–302.
29. Mailath, G., Samuelson, L., Swinkels, J.: How proper is sequential equilibrium? Games Econ. Beh. **18** (1997) 193–218.
30. Mertens, J.-M., Zamir, S.: Formulation of Bayesian analysis for games of incomplete information. Int. J. Game Theory **14** (1985) 1–29.
31. Morris, S.: Alternative notions of belief. In "Epistemic Logic and the Theory of Games and Decisions" (Bacharach, Gérard-Varet, Mongin, Shin, Eds.), Kluwer, Dordrecht (1997).
32. Myerson, R.: Refinement of the Nash equilibrium concept. Int. J. Game Theory **7** (1978) 73–80.
33. Myerson, R.: "Game Theory", Harvard University Press, Cambridge, MA (1991).
34. Pearce, D.G.: Rationalizable strategic behavior and the problem of perfection. Econometrica **52** (1984) 1029–1050.
35. Perea y Monsuwé, A., Jansen, M., Peters, H.: Characterization of consistent assessments in extensive form games. Games Econ. Beh. **21** (1997) 238–252.
36. Reny, P.J.: Backward induction, normal form perfection and explicable equilibria. Econometrica **60** (1992) 627–649.
37. Rubinstein, A.: Comments on the interpretation of game theory. Econometrica **59** (1991) 909–924.
38. Schuhmacher, F.: Proper rationalizability and backward induction. Int. J. Game Theory **28** (1999) 599–615.
39. Selten, R.: Reexamination of the perfectness concept for equilibrium points in extensive games. Int. J. Game Theory **4** (1975) 25–55.

40. Tan, T., Werlang, S.R.C.: The Bayesian foundations of solution concepts of games. J. Econ. Theory **45** (1988) 370–391.
41. van Damme, E.: A relation between perfect equilibria in extensive form games and proper equilibria in normal form games. Int. J. Game Theory **13** (1984) 1–13.

University of Oslo, Universidad Carlos III de Madrid
E-mail address: `g.b.asheim@econ.uio.no, perea@eco.uc3m.es`