MEMORANDUM

No 30/99

On the Epistemic Foundation for Backward Induction

By
Geir B. Asheim

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.sv.uio.no/sosoek/
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 29	By Jon Strand: Labor unions versus individualized bargaining with heterogeneous labor. 36 p.
No 28	By Jon Vislie: The role of foreign ownership in domestic environmental regulation under Asymmetric. 19 p.
No 27	By Erik Biørn: Random Coefficients in Regression Equation Systems: The Case with unbalanced Panel Data. 19 p.
No 26	By Gunnar Bårdsen, Eilev S. Jansen and Ragnar Nymoen: Econometric inflation targeting. 38 p.
No 25	By Cheetal K Chand and Karl O. Moene: Rent Grabbing and Russia's Economic Collapse. 35 p.
No 24	By Rolf Golombek and Espen R. Moen: Do Voluntary Agreements Lead to Cost Efficiency? 45 p.
No 23	By Atle Seierstad: Necessary conditions involving generalized directional derivatives for optimal control of nonsmooth retarded Volterra integral equations in Banach space. 68 p.
No 22	By Karl Ove Moene and Michael Wallerstein: Inequality, Social Insurance and Redistribution. 40 p.
No 21	By Kai Leitemo: Inflation Targeting Strategies in small open Economies. 44 p.
No 20	By Erik Biørn: Estimating Regression Systems from unbalanced Panel Data: A Stepwise Maximum Likelihood Procedure. 18 p.

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

ON THE EPISTEMIC FOUNDATION FOR BACKWARD INDUCTION

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 address: g.b.asheim@econ.uio.no

Proposed running head:

BACKWARD INDUCTION

Abstract:

I characterize backward induction in an epistemic model of perfect information games where players have common certain belief of the consistency of preferences rather than the rationality of choice. In this approach, backward induction corresponds to common certain belief of 'belief *in each subgame* of opponent rationality'. At an interpretative level this result resembles the one established by Aumann [6]. By instead imposing common certain belief of 'belief (*only in the whole game*) of opponent rationality', I interpret Ben-Porath's [14] support of the Dekel-Fudenberg procedure. *JEL* Classification Number: C72.

Date: 28 December 1999.

Acknowledgements: This paper builds in part on joint work with Martin Dufwenberg, who has contributed with helpful suggestions. I also thank Adam Brandenburger, Thorsten Clausing, Stephen Morris and Ylva Søvik for useful comments. A preliminary and abbreviated version of this paper is included in "Logic, Game Theory, and Social Choice – Proceedings of the International Conference, LGS'99" (de Swart, Ed.), Tilburg University Press (1999).

1. Introduction

In recent years, two widely cited and influential contributions on backward induction in finite perfect information games have appeared, namely Aumann [6] and Ben-Porath [14]. These contributions — both of which consider *generic* perfect information games (where all payoffs are different) — reach opposite conclusions: While Aumann establishes that *Common Knowledge of Rationality* implies that the backward induction outcome is reached, Ben-Porath shows that the backward induction outcome is *not* the only outcome that is consistent with *Common Certainty of Rationality*. The models of Aumann and Ben-Porath are different. One such difference is that Aumann makes use of 'knowledge' in the sense of 'true knowledge', while Ben-Porath's analysis is based on 'certainty' in the sense of 'belief with probability one'. Another is that the term 'rationality' is used in different senses.

The present paper shows how the conclusions of Aumann and Ben-Porath can be captured by imposing requirements on the players within the same general framework. Furthermore, the interpretations of the present analysis correspond closely to the intuitions that Aumann and Ben-Porath convey in their discussions. Hence, the present contribution may increase our understanding of the differences between the analyses of Aumann and Ben-Porath, and thereby enhance our understanding of the epistemic conditions underlying backward induction. For ease of presentation, the analysis will be limited to 2-player games. This is purely a matter of convenience as everything can directly be generalized to n-player games (with $n > 2$).

Among the large literature on backward induction during the last couple of decades,[1] Reny's [40] impossibiliy result is of special importance. Reny associates a player's 'rationality' in an extensive game with what can be called 'reachable subgame rationality'; i.e. that a player chooses rationally in all subgames that are not precluded from being reached by the player's own strategy. He shows that there exist perfect information games where the event that both players satisfy *reachable subgame rationality cannot be commonly believed in all subgames*. E.g. in the centipede game that is illustrated in Fig. 1 of Sect. 4, common belief of reachable subgame rationality cannot be held in the subgame defined by 2's decision node. The reason is that if 1 believes that 2 is rational in the subgame, and if 1 believes that 2 believes that 1 will be rational in the subgame defined by 1's second decision node, then 1 believes that 2 will choose d, implying that only D is a best response for 1. Then the fact that the subgame defined by 2's decision node has been reached, contradicts 2's belief that 1 is rational in the whole game.

[1] Among contributions that are not otherwise referred to here are Basu [10], Bicchieri [15], Binmore [16, 17], Bonanno [21], Gul [31], Kaneko [32], Rabinowicz [39], Rosenthal [41] and Vilks [47], as well as Asheim [4] and Schuhmacher [44].

As a response, Ben-Porath [14] imposes that *common belief of reachable subgame rationality* is held *in the whole game only*. However, reachable subgame rationality being commonly believed in the whole game only, does not imply backward induction. In the centipede game of Fig. 1, the strategies D and FD for player 1 and d and f for player 2 are consistent with such common belief, while backward induction implies that down is played at any decision node.

In order to obtain an epistemic characterization of backward induction, Aumann [6] considers 'all subgame rationality' in the sense that a player chooses *rationally in all subgames*. However, the event that players satisfy all subgame rationality is somewhat problematic. If — in the centipede game of Fig. 1 — 1 believes or knows that 2 chooses d, then only by choosing the strategy DD will 1 satisfy all subgame rationality. However, what does it mean that 1 chooses DD in the counterfactual event that player 2's decision node were reached? It is perhaps more natural — as suggested by Stalnaker ([45], Sect. 5) — to consider 2's belief about 1's subsequent action if 2's decision node were reached. Since Aumann [6] assumes knowledge of rational choice in an S5 partition structure, such a question of belief revision cannot be asked within Aumann's model.

By imposing that each player *takes all opponent strategies into account* ('caution'), the present paper ensures that each player takes the possibility of reaching any subgame of the extensive form into account. This means that a rational choice in the whole game implies a rational choice in all subgames that are not precluded from being reached by the player's own strategy. Hence, by imposing the strategic form restriction of 'caution', one may consider 'rationality' instead of 'reachable subgame rationality' (as established by Lemma 1 and the subsequent text).

The main distinguishing feature of the present analysis is, however, to consider the event that a player *believes in* opponent rationality rather than the event that the player himself *is* rational. Asheim & Dufwenberg [5] (AD) show the following result (reproduced as Prop. 1 of Sect. 3.1): Strategies surviving the Dekel-Fudenberg [27] procedure, where one round of weak elimination is followed by iterated strong elimination, can be characterized as maximal strategies when there is common certain belief that each player satisfies 'caution' and *believes in the whole game that the opponent chooses rationally* ('belief of opponent rationality'). For generic perfect information games, Ben-Porath shows that the set of outcomes consistent with common belief of reachable subgame rationality corresponds to the set of outcomes that survives the Dekel-Fudenberg procedure. Hence, maximal strategies when there is common certain belief of 'caution' and 'belief of opponent rationality' correspond to outcomes that are promoted by Ben-Porath's analysis.

An extensive game offers choice situations, not only in the whole game, but also in proper subgames. In perfect information games (and, more generally, in multi-stage games) the subgames constitute an exhaustive set of such choice situations. Hence, in perfect information games one can argue that 'belief of opponent rationality' should be replaced by 'belief in each subgame of opponent rationality': Each player *believes in each subgame that his opponent chooses rationally in the subgame*. The main results of the present paper (Props. 2 and 3 of Sect. 5.2) show how, for generic perfect information games, common certain belief of 'caution' and 'belief in each subgame of opponent rationality' is possible and uniquely determines the backward induction outcome. Hence, by substituting 'belief in each subgame of opponent rationality' for 'belief of opponent rationality', the present analysis provides an alternative route to Aumann's conclusion, namely that common knowledge (or certain belief) of an appropriate form of (belief of) rationality implies the backward induction outcome.

This epistemic foundation for backward induction requires common *certain belief* of 'caution' and 'belief in each subgame of opponent rationality', where the term 'certain belief' is being used in the sense that an event is certainly believed if the complement is Savage-null (cf. Sect. 2.3). As shown by a counterexample in Remark 2 of Sect. 5, the characterization does not obtain if instead common *belief* (in a sense that generalizes belief with probability one) is considered. This, in turn, means that the event of which there is common certain belief — namely 'caution' and 'belief in each subgame of opponent rationality' — cannot be further restricted by considering the intersection with 'rationality'. The reason is that 'caution' is in general inconsistent with certain belief of opponent 'rationality', as the latter prevents a player from taking into account the possibility that the opponent does not choose rationally.

Note that 'caution' and 'belief (in each subgame) of opponent rationality' are requirements on the beliefs of players. Thus, the analysis follows AD by arguing that in deductive game theory, requirements should be imposed on the beliefs of players rather than their choice. Since the beliefs of players determine their preferences, this amounts to imposing requirements on preferences.[2] The present analysis allows, but do not require, subjective probabilities, which are arguably not part of the backward induction argument in generic perfect information games (cf. Aumann [6] and Brandenburger [24]). Hence, preferences need not be complete. By not requiring subjective probabilities, the analysis is related to the filter model of (conditional) belief presented by Brandenburger [24].

The paper is organized as follows. Section 2 presents the formal framework in which extensive games will be analyzed. AD's characterization of

[2]Instead of imposing rational choice in the sense that a driver *chooses* to drive on the right side of the road if he believes that his opponent chooses to drive on the right side of the road, AD suggest to impose consistent preferences in the sense that a driver *prefers* to drive on the right side of the road if he believes that his opponent prefers to drive on the right side of the road. This follows a tradition in equilibrium analysis where Nash (perfect/proper) equilibrium is defined as an equilibrium in conjectures (cf. Blume et al. [19]).

the Dekel-Fudenberg procedure is reviewed in Sect. 3. Section 4 applies this result to generic extensive games of perfect information and compares, by means of an example, the present analysis to that of Ben-Porath [14]. Section 5 introduces 'belief in each subgame of opponent rationality' as an alternative epistemic condition and establishes the paper's main results. Section 6 interprets the analysis in view of Aumann [6] as well as Battigalli's [11] concept of a 'rationality ordering'.

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 et al. [18]. For the analysis of extensive games, *continuity* must be relaxed to allow each player to take into account the possibility that any subgame can be reached. Moreover, by not imposing *completeness*, the analysis does not require subjective probabilities. The framework is summarized by the concept of a *belief system* (cf. Def. 1). The Appendix contains a presentation of the decision-theoretic terminology, notation and results that will be utilized.

2.1. **An Extensive Game Form.** Inspired by Osborne & Rubinstein ([38], Ch. 6), a finite extensive game form of almost perfect information with 2 players and $M-1$ stages can be described as follows. Both perfect information games and finitely repeated games yield game forms that fit this description. The sets of *histories* is determined inductively: The set of histories at the beginning of the first stage 1 is $H^1 = \{\emptyset\}$. Let H^m denote the set of histories at the beginning of stage m. At $h \in H^m$, let, for each player $i \in N := \{1, 2\}$, i's action set be denoted $A_i(h)$, where i is inactive at h if $A_i(h)$ is a singleton. Write $A(h) := A_1(h) \times A_2(h)$. Define the set of histories at the beginning of stage $m+1$ as follows: $H^{m+1} := \{(h, a) \,|\, h \in H^m \text{ and } a \in A(h)\}$. This concludes the induction. Let $H := \bigcup_{m=1}^{M-1} H^m$ denote the set of *subgames* and let $Z := H^M$ denote the set of *outcomes*.

A *pure strategy* for player i is a function s_i that assigns an action in $A_i(h)$ to any $h \in H$. Let S_i denote player i's finite set of pure strategies, and write $S := S_1 \times S_2$. Write p, r, and s ($\in S$) for pure strategy vectors. Let $z : S \to Z$ map strategy vectors into outcomes.[3] Then $((S_i)_{i \in N}, z)$ is a finite *strategic* two-player *game form*. For any $h \in H \cup Z$, let $S(h) = S_1(h) \times S_2(h)$ denote the set of strategy vectors that are *consistent* with h being reached.

[3] A pure strategy $s_i \in S_i$ can be viewed as an act on S_j that assigns $z(s_i, s_j) \in Z$ to any $s_j \in S_j$. The set of pure strategies S_i is partitioned into equivalent classes of acts since a pure strategy s_i also determines actions in subgames which s_i prevents from being reached. Each such equivalent class corresponds to a *plan of action* in the sense of Rubinstein [42]. As there is no need here to differentiate between identical acts, the concept of a plan of action would have sufficed. Since completeness is not imposed, a type of a player need not make any assessment concerning the relative likelihood of identical acts.

If h' is the predecessor of h, then $S(h') \supseteq S(h)$. If $s_i \in S_i$ and $h \in H$, let $s_i|_h$ denote the strategy in $S_i(h)$ satisfying $s_i|_h(h') = s_i(h')$ at any $h' \in H$ except at h' with $S(h') \supset S(h)$ where $s_i|_h(h')$ is dictated by $s_i|_h$ being consistent with h.

2.2. **States and Types.** When a strategic game form is turned into a decision problem for each player (see Tan & Werlang [46]), the uncertainty faced by a player concerns the strategy choice of his opponent, the belief of his opponent about his 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.

Given an assumption of coherency, models of such infinite hierarchies of beliefs (Armbruster & Böge [3], Böge & Eisele [20], Mertens & Zamir [35], Brandenburger & Dekel [25], Epstein & Wang [30]) yield $S \times T$ as the complete state space, where S is the underlying space of uncertainty and 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. Combined with a vNM utility function, the set of beliefs on $S \times T_j$ corresponds to the set of "regular" binary relations on the set of acts on $S \times T_j$, where an *act* on $S \times T_j$ is a function that to any element of $S \times T_j$ assigns an objective randomization on Z.

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 decision is not relevant and can be ignored. Hence, models of infinite hierarchies of beliefs — in the setting of a strategic game form — imply that each type of any player i corresponds to a "regular" binary relation on the set of acts on $S_j \times T_j$.

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, like AD, I do not construct a complete state space by explicitly modeling infinite hierarchies of beliefs. For tractability I instead directly consider an implicit model — with a finite type set T_i for each player i — from which infinite hierarchies of beliefs can be constructed.[4] Moreover, since completeness and continuity of preferences are not imposed, the "regularity" conditions on $\succeq^{t_i}$ consist of *reflexivity*, *transitivity*, *objective independence*, *nontriviality*, *conditional completeness*, *conditional continuity* and *non-null state independence*, meaning that $\succeq^{t_i}$ is conditionally represented

[4]This is not purely a matter of convenience as Brandenburger [24] and Brandenburger & Keisler [26] have shown that a complete state space may not exist if beliefs are not based on subjective probabilities. In contrast to Battigalli & Siniscalchi's [13] epistemic foundation for backward (and forward) induction, a complete state space is not needed for the present analysis.

by a vNM utility function $\upsilon_i^{t_i} : Z \to \mathbb{R}$ that assigns a payoff to any outcome (cf. the Appendix).[5] Being a vNM utility function, $\upsilon_i^{t_i}$ can be extended to objective randomizations on Z. Since $\succeq^{t_i}$ is conditionally represented, it follows that strong and weak dominance are well-defined. The construction is summarized by the following definition.

Definition 1. A *belief system* for a game form $((S_i)_{i\in N}, z)$ 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 conditionally represented by a vNM utility function $\upsilon_i^{t_i}$ on the set of objective randomizations on Z.

2.3. **Epistemic operators.** When preferences are not continuous, one can differentiate between belief and certain belief in a manner that will be explained below. Both 'belief' and 'certain belief' are subjective, as they are derived from preferences (following the approach of Morris [37]); hence, neither operator satisfies the truth axiom. To state these operators, 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 event $E \subseteq \Omega$, $E_j^{t_i} := \{(s_j, t_j) \in S_j \times T_j | \exists (s'_1, s'_2, t'_1, t'_2) \in E$ s.t. $(s'_j, t'_j) = (s_j, t_j)$ and $t'_i = t_i\}$ denote the set of opponent strategy-type pairs that are consistent with $\omega \in E$ and $t_i(\omega) = t_i$.

It is perhaps easier to introduce these concepts in the case when preferences are complete and, thus, representable in terms of an LPS $\lambda^{t_i} = (\mu_1^{t_i}, \ldots, \mu_L^{t_i}) \in \mathbf{L}\Delta(S_j \times T_j)$. Then an event is 'certainly believed' if no element of the complement 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 $\kappa_j^{t_i} := \mathrm{supp}\lambda^{t_i}$ $(\subseteq S_j \times T_j)$. On the other hand, an event is 'believed' if no element of the complement is assigned positive probability by $\mu_1^{t_i}$:[6]

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

where $\beta_j^{t_i} := \mathrm{supp}\mu_1^{t_i}$ $(\subseteq S_j \times T_j)$. It follows that $K_i E \subseteq B_i E$ (i.e. 'certain belief' implies 'belief') since $\beta_j^{t_i} = \mathrm{supp}\mu_1^{t_i} \subseteq \kappa_j^{t_i} = \mathrm{supp}\lambda^{t_i} := \cup_{\ell=1}^{L} \mathrm{supp}\mu_\ell^{t_i}$.

To generalize $\kappa_j^{t_i}$ (and thus $K_i E$) to incomplete preferences, let

$$\kappa_j^{t_i} := \{(s_j, t_j) \in S_j \times T_j | (s_j, t_j) \text{ is not Savage-null acc. to } \succeq^{t_i}\}$$

[5]If conditional completeness is strengthened to completeness, then it follows from Blume et al. [18] that $\succeq^{t_i}$ is represented by $\upsilon_i^{t_i}$ and a lexicographic probability system (LPS) $\lambda^{t_i} = (\mu_1^{t_i}, \ldots, \mu_L^{t_i}) \in \mathbf{L}\Delta(S_j \times T_j)$ (cf. Prop. A2 of the Appendix). If, in addition, conditional continuity is strengthened to continuity, then $\succeq^{t_i}$ is represented by $\upsilon_i^{t_i}$ and a subjective probability distribution $\mu^{t_i} \in \Delta(S_j \times T_j)$. Continuity is inconsistent with the present analysis due to the requirement of 'caution'. However, completeness, implying a subjective probability representation through an LPS, is consistent with – but not a necessary part of – the analysis.

[6]This notion of 'belief' in the case of complete preferences corresponds to Brandenburger's [23] 'first-order knowledge' and Ben-Porath's [14] 'certainty'.

denote the set of opponent strategy-type pairs that t_i does not deem Savage-null.[7] This generalizes the case of complete preferences, since in that case suppλ^{t_i} is the set of opponent strategy-type pairs that t_i does not deem Savage-null.

To generalize $\beta_j^{t_i}$ (and thus B_iE) to incomplete preferences, say that $\succeq^{t_i}$ *is admissible on* β_j, where $\emptyset \neq \beta_j \subseteq S_j \times T_j$, if $\mathbf{x} \succ^{t_i} \mathbf{y}$ whenever $\mathbf{x}_{\beta_j}$ weakly dominates $\mathbf{y}_{\beta_j}$. If $\succeq^{t_i}$ is admissible on β_j, then any $(s_j', t_j') \in \beta_j$ is deemed infinitely more likely than any $(s_j'', t_j'') \in S_j \times T_j \backslash \beta_j$. Since (s_j', t_j') being infinitely more likely than (s_j'', t_j'') implies that (s_j'', t_j'') is *not* infinitely more than (s_j', t_j'), it follows that $\beta_j' \subseteq \beta_j''$ or $\beta_j' \supseteq \beta_j''$ whenever $\succeq^{t_i}$ is admissible on both β_j' and β_j''. Since, in addition, $\succeq^{t_i}$ is admissible on $\kappa_j^{t_i}$, it follows that there exists a unique smallest (w.r.t. set inclusion) set on which $\succeq^{t_i}$ is admissible; let this set be denoted $\beta_j^{t_i}$:[8]

$$\succeq^{t_i} \text{ is admissible on } \beta_j^{t_i} \ \text{ and } \ \beta_j \supseteq \beta_j^{t_i} \text{ whenever } \succeq^{t_i} \text{ is admissible on } \beta_j \,.$$

This generalizes the case of complete preferences, since in that case supp$\mu_1^{t_i}$ is the unique smallest set of opponent strategy-type pairs on which $\succeq^{t_i}$ is admissible. Also with incomplete preferences it follows that $K_iE \subseteq B_iE$ since $\succeq^{t_i}$ is admissible on $\kappa_j^{t_i}$; i.e. $\beta_j^{t_i} \subseteq \kappa_j^{t_i}$. If $\beta_j^{t_i} \neq \kappa_j^{t_i}$, then t_i's preferences are not continuous.

In addition to $K_iE \subseteq B_iE$, it follows that the operators B_i and K_i satisfy

$$\begin{aligned} B_iE \cap B_iF &= B_i(E \cap F) & \qquad K_iE \cap K_iF &= K_i(E \cap F) \\ B_i\emptyset &= \emptyset & K_i\Omega &= \Omega \\ B_iE &\subseteq K_iB_iE & K_iE &\subseteq K_iK_iE \\ \neg B_iE &\subseteq K_i(\neg B_iE) & \neg K_iE &\subseteq K_i(\neg K_iE). \end{aligned}$$

Since $K_iE \subseteq B_iE$ implies that $K_i\emptyset = \emptyset$, $B_i\Omega = \Omega$, $B_iE \subseteq B_iB_iE$ and $\neg B_iE \subseteq B_i(\neg B_iE)$, both operators B_i and K_i correspond to KD45 systems. Since an event can be certainly believed even though the true state is an element of the complement of the event, it follows that neither operator satisfies the truth axiom (i.e. $K_iE \subseteq E$ and $B_iE \subseteq E$ need not hold).

Say that i believes the event $E \subseteq \Omega$ given ω if $\omega \in B_iE$ (or equivalently, $\beta_j^{t_i(\omega)} \subseteq E_j^{t_i(\omega)}$). Write $BE := B_1E \cap B_2E$. Say that there is mutual belief of $E \subseteq \Omega$ given ω if $\omega \in BE$. Write $CBE := BE \cap BBE \cap BBBE \cap \dots$. Say that there common belief of $E \subseteq \Omega$ given ω if $\omega \in CBE$. Say that i certainly believes the event $E \subseteq \Omega$ given ω if $\omega \in K_iE$ (or equivalently, $\kappa_j^{t_i(\omega)} \subseteq E_j^{t_i(\omega)}$). Write $KE := K_1E \cap K_2E$. Say that there is mutual certain belief of $E \subseteq \Omega$ given ω if $\omega \in KE$. Write $CKE := KE \cap KKE \cap KKKE \cap \dots$. Say that there is common certain belief of $E \subseteq \Omega$ given ω if $\omega \in CKE$.

[7] The term 'certain belief' for this notion is also used by Morris [37]. It is stronger than Ben-Porath's [14] 'certainty' as it does not allow the complement of a certainly believed event to be taken into account, but weaker than Aumann's [6] 'knowledge' as a certainly believed event need not be true.

[8] This notion of 'belief' is related to, but differs from, Morris' [37] 'strong belief'.

2.4. **Preferences over Strategies.** Let $\succeq^{t_i}_{S_j}$ 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 $s_j \in S_j$. A *mixed* strategy $x_i \in \Delta(S_i)$ corresponds to an act $\mathbf{x}_{S_j}$ on S_j that assigns $z(x_i, s_j)$ to any $s_j \in S_j$. Hence, $\succeq^{t_i}_{S_j}$ is a binary relation also on the subset of acts on S_j that correspond to i's mixed strategies. Thus, $\succeq^{t_i}_{S_j}$ can be referred to as t_i's *preferences over i's mixed strategies*. The set of mixed strategies $\Delta(S_i)$ is the set of acts that are at t_i's actual disposal.

Likewise, in any subgame h is $\succeq^{t_i}_{S_j(h)}$ t_i's conditional preferences over i's mixed strategies in h. Since $\succeq^{t_i}$ is reflexive and transitive and satisfies objective independence, $\succeq^{t_i}_{S_j(h)}$ shares these properties, and

$$C_i^{t_i}(h) := \{s_i \in S_i(h) | s_i \text{ is maximal w.r.t. } \succeq^{t_i}_{S_j(h)} \text{ in } \Delta(S_i(h))\}$$

is non-empty and supports any maximal mixed strategy. Refer to $C_i^{t_i}(h)$ as t_i's *choice set* in the subgame h, and refer to $C_i^{t_i} : H \to 2^{S_i(h)} \backslash \{\emptyset\}$ as t_i's *choice function*. Write $C_i^{t_i} := C_i^{t_i}(\emptyset)$, and write, for any $h \in H$, $C^t(h) := C_1^{t_1}(h) \times C_2^{t_2}(h)$.

By the following lemma, if s_i is maximal in a subgame h, then s_i is maximal in any later subgame that s_i is consistent with.

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

Proof. Suppose that s_i is not maximal w.r.t. $\succeq^{t_i}_{S_j(h')}$ in $\Delta(S_i(h'))$. Then there exists $\mathbf{x}_{S_j}$ such that $\mathbf{x}_{S_j} \succ^{t_i}_{S_j(h')} \mathbf{y}_{S_j}$, where $\mathbf{x}_{S_j}$ assigns $z(x_i, s_j)$ to any $s_j \in S_j$ with $x_i \in \Delta(S_i(h'))$, and where $\mathbf{y}_{S_j}$ assigns $z(s_i, s_j)$ to any $s_j \in S_j$. By Mailath et al. ([34], Defs. 2 and 3 and the if-part of Theorem 1), $S(h')$ is a *strategic independence* for i. Hence, $\mathbf{x}_{S_j}$ can be chosen such that $\mathbf{x}_{S_j}(s_j) = \mathbf{y}_{S_j}(s_j)$ for all $s_j \in S_j \backslash S_j(h')$. This implies that $\mathbf{x}_{S_j} \succ^{t_i}_{S_j(h)} \mathbf{y}_{S_j}$, which contradicts that s_i is maximal w.r.t. $\succeq^{t_i}_{S_j(h)}$ in $\Delta(S_i(h))$. □

Under the assumption that $\kappa_j^{t_i} \cap S_j(h) \times T_j \neq \emptyset$ (which is implied if t_i satisfies 'caution'; cf. Sect. 3.1), it follows that $\succeq^{t_i}_{S_j(h)}$ is nontrivial.

2.5. **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 payoff to any outcome. Then the pair of the extensive game form and the vNM utility functions $(\upsilon_i)_{i \in N}$ is a finite *extensive game* of almost perfect information, Γ. Let $G = (S_i, u_i)_{i \in N}$ 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 r, $s \in S$ such that $u_i(r) > u_i(s)$.

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 $s_i \in C_i^{t_i}$, and where the event that i is rational is defined as[9]

$$[rat_i] := \{(s_1, s_2, t_1, t_2) \in \Omega | s_i \in C_i^{t_i}\}.$$

The present paper follows AD by imposing requirements on t_i only. Since t_i corresponds to the preferences $\succeq^{t_i}$, such requirements will be imposed on $\succeq^{t_i}$. In support of this alternative approach — which will be referred to by the term 'consistent preferences' — one can note the following: The approach allows

- ... requirements to be imposed on types rather than strategy-type pairs.
- ... conventional concepts like 'rationalizable strategies' and strategies surviving the Dekel-Fudenberg procedure to be characterized under very weak and natural conditions (see e.g. Prop. 1 below).
- ... requirements like 'caution' and 'belief (in each subgame) of opponent rationality' to be imposed in a straightforward manner. Under 'rational choice' the notion of 'certain belief' must be weakened to accommodate caution (cf. Börgers ([22], pp. 266–267) and Epstein ([29], p. 3)). It is unclear how 'belief in each subgame of opponent rationality' can be imposed under the 'rational choice' approach.

Here I will focus on showing how 'consistent preferences' as an approach to deductive game-theoretic analysis can be used to shed light on the analyses of Aumann [6] and Ben-Porath [14], and thereby enhance our understanding of the epistemic conditions underlying backward induction. For this purpose, it is useful to reproduce AD's characterization of the Dekel-Fudenberg procedure.

3.1. Admissible Consistency.

The Dekel-Fudenberg procedure is made up of one round of elimination of weakly dominated strategies followed by iterated elimination of strongly dominated strategies. AD characterize this procedure by imposing three requirements: The first of these ensures that each player plays the game G, the second requirement ensures that each player takes all opponent strategies into account, thereby taking into account the possibility that any subgame in an extensive game be reached (*caution*), while the third requirement ensures that each player believes that the opponent chooses rationally (*belief of opponent rationality*).

To impose these requirements, consider the following events

$$[cau_i] := \{\omega \in \Omega | \kappa_j^{t_i(\omega)} = S_j \times T_j^{t_i(\omega)}\}$$

$$B_i[rat_j] = \{\omega \in \Omega | (r_j, t_j) \in \beta_j^{t_i(\omega)} \text{ implies } r_j \in C_j^{t_j}\},$$

[9] See e.g. Epstein ([29], Sect. 6) for a presentation of this approach in a general context.

where $T_j^{t_i} := \text{proj}_{T_j} \kappa_j^{t_i}$ denotes the set of opponent types that t_i does not deem Savage-null.[10]

- If $\omega \in [cau_i]$, then (s_j, t_j) is not deemed Savage-null acc. to $\succeq^{t_i(\omega)}$ whenever t_j is not deemed Savage-null. This means that, $\forall (s_j, t_j) \in S_j \times T_j^{t_i(\omega)}$, $\omega \notin K_i\{(s_1', s_2', t_1', t_2') \in \Omega | (s_j', t_j') \neq (s_j, t_j)\}$ (cf. Dekel & Gul's [28] definition of caution). It implies that the marginal of $\succeq^{t_i(\omega)}$ on S_j (i.e., $t_i(\omega)$'s preferences over S_i, $\succeq_{S_j}^{t_i(\omega)}$) is admissible on S_j.
- If $\omega \in B_i[rat_j]$, then i believes given ω that j is rational.

Say that i is *admissibly consistent* (with the game G and the preferences of his opponent) given ω if $\omega \in A_i$, where

$$A_i := [u_i] \cap [cau_i] \cap B_i[rat_j].$$

Refer to $A := A_1 \cap A_2$ as the event of *admissible consistency*. The Dekel-Fudenberg procedure can now be characterized as maximal strategies in states where there is common certain belief of admissible consistency.

Proposition 1. (Asheim & Dufwenberg [5]**)** *A pure strategy r_i for i survives the Dekel-Fudenberg procedure in a finite strategic game G if and only if there exists a belief system with $r_i \in C_i^{t_i(\omega)}$ for some $\omega \in CKA$.*

4. Generic Games of Perfect Information

A finite extensive game is

- ... of *perfect information* if, at any $h \in H$, there exists at most one player that has a non-singleton action set.
- ... *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. If p denotes the unique subgame-perfect equilibrium, then, for any subgame h, $z(p|_h)$ is the backward induction outcome in the subgame h, and $S(z(p|_h))$ is the set of strategy vectors consistent with the backward induction outcome in the subgame h.

Both Aumann [6] and Ben-Porath [14] analyze generic extensive games of perfect information. As already pointed out, while Aumann establishes

[10] If $\omega \in [cau_i] \cap B_i[rat_j]$ and $\succeq^{t_i(\omega)}$ is complete, then $\succeq^{t_i(\omega)}$ can be represented by $\upsilon_i^{t_i(\omega)}$ and an LPS $\lambda^{t_i(\omega)} = (\mu_1^{t_i(\omega)}, \ldots, \mu_L^{t_i(\omega)}) \in \mathbf{L}\Delta(S_j \times T_j)$ satisfying $\text{supp}\lambda^{t_i(\omega)} = S_j \times T_j^{t_i(\omega)}$ and $\mu_1^{t_i(\omega)}(r_j, t_j) > 0$ only if $r_j \in C_j^{t_j}$. Note that $\omega \in [cau_i] \cap B_i[rat_j]$ does not imply — but is consistent with — $\succeq^{t_i(\omega)}$ being complete, while $\omega \in [cau_i] \cap B_i[rat_j]$ is *not* consistent with $\succeq^{t_i(\omega)}$ being continuous.

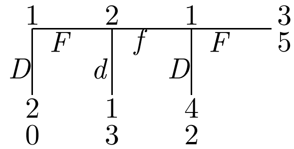

FIGURE 1. A centipede game

TABLE 1. A belief system for the game of Fig. 1.

t_1':	t_2'	t_2''
d	$\left(\frac{4}{5}, \frac{7}{10}\right)$	$\left(0, \frac{1}{10}\right)$
f	$\left(0, \frac{1}{10}\right)$	$\left(\frac{1}{5}, \frac{1}{10}\right)$

t_1'':	t_2'	t_2''
d	$\left(\frac{3}{5}, \frac{5}{10}\right)$	$\left(0, \frac{1}{10}\right)$
f	$\left(0, \frac{1}{10}\right)$	$\left(\frac{2}{5}, \frac{3}{10}\right)$

t_2':	t_1'	t_1''
D	$\left(\frac{1}{2}, \frac{1}{4}\right)$	$\left(0, \frac{1}{8}\right)$
FD	$\left(0, \frac{1}{8}\right)$	$\left(\frac{1}{2}, \frac{1}{4}\right)$
FF	$\left(0, \frac{1}{8}\right)$	$\left(0, \frac{1}{8}\right)$

t_2'':	t_1'	t_1''
D	$\left(1, \frac{1}{2}\right)$	$(0, 0)$
FD	$\left(0, \frac{1}{4}\right)$	$(0, 0)$
FF	$\left(0, \frac{1}{4}\right)$	$(0, 0)$

that common (true) knowledge of (all subgame) rationality[11] implies that the backward induction outcome is reached, Ben-Porath shows that the backward induction outcome is not the only outcome that is consistent with common belief (in the whole game) of (reachable subgame) rationality. The purpose of the present section is to interpret the analysis of Ben-Porath by applying Prop. 1 to the class of generic perfect information games.

Ben-Porath [14] establishes through his Theorem 1 that the set of outcomes consistent with common belief (in the whole game) of (reachable subgame) rationality corresponds to the set of outcomes that survive the Dekel-Fudenberg procedure. Hence, by Prop. 1, maximal strategies when there is common certain belief of admissible consistency correspond to the outcomes promoted by Ben-Porath's analysis.

4.1. **An Example.** To illustrate how common certain belief of admissible consistency is consistent with outcomes other than the unique backward induction outcome, consider the simple centipede game of Fig. 1 where backward induction implies that down is being played at any decision node. Let $T_1 = \{t_1', t_1''\}$ and $T_2 = \{t_2', t_2''\}$. Assume that the preferences of each type t_i of any player i are represented by a vNM utility function $v_i^{t_i}$ satisfying $v_i^{t_i} \circ z = u_i$ and a 2-level LPS on $S_j \times T_j$. In Table 1, the first numbers in the parentheses express primary probability distributions, while the second numbers express secondary probability distributions. The strategies DD and DF are merged as their relative likelihood does not matter; see footnote 3. Note that all types satisfy 'caution'. With these 2-level LPSs each type's

[11] Aumann's [6]) analysis is based on *substantive* rationality. See Aumann [6], pp. 14–16, and Aumann [7].

preferences over the player's own strategies are given by

$$t_1' : D \succ FD \succ FF$$
$$t_1'' : FD \succ D \succ FF$$
$$t_2' : d \succ f$$
$$t_2'' : f \succ d$$

It is easy to check that all types satisfy 'belief of opponent rationality' (e.g. both t_2' and t_2'' assign positive (primary) probability to an opponent strategy-type pair only if it is a maximal strategy for the opponent type, i.e. D in the case of t_1' and FD in the case of t_1''). Thus, with $\Omega = S \times T_1 \times T_2$, it follows that $\Omega = A = CKA$. Hence, preferences consistent with common certain belief of admissible consistency need not reflect backward induction since FD and f are maximal strategies.

Note that type t_2', conditional on his decision node being reached (i.e. 1 choosing FD or FF), updates his beliefs about the type of player 1 and assigns (primary) probability one to 1 being of type t_1''. Consequently, the conditional belief of type t_2' about 1's strategy choice assigns (primary) probability one to FD. Type t_2'', on the other hand, does not admit the possibility that 1 is of another type than t_1'. Since the choice of F at 1's first decision node is not rational for t_1', there is no restriction concerning the conditional belief of type t_2'' about the choice at 1's second decision node. In the terminology of Ben-Porath, a "surprise" has occurred. Subsequent to such a surprise, a type need not believe that the opponent type chooses rationally among his remaining strategies.

Remark 1. Ben-Porath uses the extensive form as a means of imposing that the possibility of reaching any subgame is taken into account. Here, this is ensured by the strategic form restriction that preferences be admissible (cf. the requirement of 'caution' in Sect. 3.1 as well as the discussion by Stalnaker ([45], Sect. 4)). Instead I will below explicitly use the extensive form for making restrictions on the beliefs that types hold in subgames concerning opponent rationality.

5. Belief in Each Subgame of Opponent Rationality

A simultaneous game offers only one choice situation. Hence, for a game in this class, it seems reasonable that belief of opponent rationality is held *in the whole game only*, as formalized by the requirement 'belief of opponent rationality'. An extensive game with a nontrivial dynamic structure, however, offers such choice situations, not only in the whole game, but also in proper subgames. Moreover, for extensive games of almost perfect information, the subgames constitute an exhaustive set of such choice situations. This motivates imposing *belief in each subgame of opponent rationality*.

To reach a subgame $h \in H$ is an objectively knowable event $S_1(h) \times S_2(h) \times T_1 \times T_2$. Conditional on any such objectively knowable event, a conditional belief operator for each player i, $B_i(h)$, can be defined. To state this operator for any subgame $h \in H$, write $\Omega_j(h) := S_j(h) \times T_j$, and let, for any event $E \subseteq \Omega$, $E_j^{t_i}(h) := E_j^{t_i} \cap \Omega_j(h)$.

In any subgame h is $\succeq_{\Omega_j(h)}^{t_i}$ the conditional preferences of t_i in h. Say that $\succeq_{\Omega_j(h)}^{t_i}$ *is admissible on* β_j, where $\emptyset \neq \beta_j \subseteq \Omega_j(h)$, if $\mathbf{x} \succ_{\Omega_j(h)}^{t_i} \mathbf{y}$ whenever $\mathbf{x}_{\beta_j}$ weakly dominates $\mathbf{y}_{\beta_j}$. If $\succeq_{\Omega_j(h)}^{t_i}$ is admissible on β_j, then any $(s_j', t_j') \in \beta_j$ is deemed infinitely more likely than any $(s_j'', t_j'') \in \Omega_j(h) \backslash \beta_j$. Since (s_j', t_j') being infinitely more likely than (s_j'', t_j'') implies that (s_j'', t_j'') is *not* infinitely more than (s_j', t_j'), it follows that $\beta_j' \subseteq \beta_j''$ or $\beta_j' \supseteq \beta_j''$ whenever $\succeq_{\Omega_j(h)}^{t_i}$ is admissible on both β_j' and β_j''. If the set of opponent strategy-type pairs that t_i does not deem Savage-null in h, $\kappa_j^{t_i} \cap \Omega_j(h)$, is non-empty — which is implied by t_i satisfying 'caution' — then $\succeq_{\Omega_j(h)}^{t_i}$ is admissible on $\kappa_j^{t_i} \cap \Omega_j(h)$. Hence, under this assumption, there exists a unique smallest (w.r.t. set inclusion) set on which $\succeq_{\Omega_j(h)}^{t_i}$ is admissible; let this set be denoted $\beta_j^{t_i}(h)$:

$$\succeq_{\Omega_j(h)}^{t_i} \text{ is adm. on } \beta_j^{t_i}(h) \text{ and } \beta_j \supseteq \beta_j^{t_i}(h) \text{ whenever } \succeq_{\Omega_j(h)}^{t_i} \text{ is adm. on } \beta_j .$$

Otherwise, set $\beta_j^{t_i}(h) = \emptyset$. Say that i believes an event E conditional on h given ω if $\beta_j^{t_i(\omega)}(h) \subseteq E_j^{t_i(\omega)}(h)$, or equivalently, $\omega \in B_i(h)E$, where

$$B_i(h)E := \{\omega \in \Omega | \beta_j^{t_i(\omega)}(h) \subseteq E_j^{t_i(\omega)}(h)\}.$$

If h' is a predecessor of h (i.e. $S(h') \supseteq S(h)$) and $\beta_j^{t_i}(h') \cap \Omega_j(h) \neq \emptyset$, then it follows from the above definitions that $\beta_j^{t_i}(h) = \beta_j^{t_i}(h') \cap \Omega_j(h)$. Furthermore, $K_i E \subseteq B_i(h)E$ (since $\beta_j^{t_i}(h) \subseteq \kappa_j^{t_i} \cap \Omega_j(h)$), and

$$\begin{aligned} B_i(h)E \cap B_i(h)F &= B_i(h)(E \cap F) \\ B_i(h)\emptyset &= \{\omega \in \Omega | \kappa_j^{t_i(\omega)} \cap \Omega_j(h) = \emptyset\} \\ B_i(h)E &\subseteq K_i B_i(h)E \\ \neg B_i(h)E &\subseteq K_i(\neg B_i(h)E). \end{aligned}$$

Since $K_i E \subseteq B_i(h)E$ implies that $B_i(h)\Omega = \Omega$, $B_i(h)E \subseteq B_i(h)B_i(h)E$ and $\neg B_i(h)E \subseteq B_i(h)(\neg B_i(h)E)$, the operator $B_i(h)$ corresponds to a K45 system.

Note that i's belief conditional on the subgame h is "well defined" (in the sense that $\beta_i^{t_i(\omega)} \neq \emptyset$) in any state ω where i deems it possible that h can be reached (i.e., $\kappa_j^{t_i(\omega)} \cap \Omega_j(h) \neq \emptyset$, meaning that $\succeq_{\Omega_j(h)}^{t_i}$ is nontrivial). Hence, a "well defined" conditional belief in h is implied by 'caution' alone; it does not require that h is actually being reached. This means that a requirement on i's belief conditional on h is a requirement on the type of player i only; it does not impose that i makes a strategy choice consistent with h.

Say that s_j is a 'rational choice' by t_j in h if $s_j \in C_j^{t_j}(h)$, and let the event that j is rational in h be defined as

$$[rat_j(h)] := \{(s_1, s_2, t_1, t_2) \in \Omega | s_j \in C_j^{t_j}(h)\} \; (\subseteq S_i \times S_j(h) \times T_i \times T_j).$$

Consider the event that i believes in h that j is rational in h:

$$B_i(h)[rat_j(h)] = \{\omega \in \Omega | (r_j, t_j) \in \beta_j^{t_i(\omega)}(h) \text{ implies } r_j \in C_j^{t_j}(h)\}.$$

If $\omega \in \bigcap_{h\in H} B_i(h)[rat_j(h)]$, then, conditional on any subgame h, i believes given ω that j is rational in h. In other words, $\bigcap_{h\in H} B_i(h)[rat_j(h)]$ is the event that player i believes in each subgame h that the opponent j is rational in h.[12]

Consider a finite extensive game Γ of almost perfect information with corresponding strategic game G. Say that i is *admissibly subgame consistent* (with Γ and the preferences of his opponent) given ω if $\omega \in A_i^*$, where

$$A_i^* := [u_i] \cap [cau_i] \cap \left(\bigcap_{h\in H} B_i(h)[rat_j(h)] \right).$$

Refer to $A^* := A_1^* \cap A_2^*$ as the event of *admissible subgame consistency*. This definition of admissible subgame consistency can be applied to any finite extensive game of almost perfect information. However, in order to relate to Aumann's [6] Theorems A and B, the following analysis is concerned with generic perfect information games.

5.1. **The Example Revisited.** In the belief system of Table 1, type t_2'' does not satisfy 'belief in each subgame of opponent rationality'. By 'belief in each subgame of opponent rationality', any type of player 2 must believe, conditional on the subgame defined by 2's decision node, that 1 chooses his maximal strategy, FD, in the subgame. This means that any type of player 2 prefers d to f, implying that no type of player 1 satisfying 'belief in each subgame of opponent rationality' can prefer FD to D. Thus, common certain belief of admissible subgame consistency entails that any types of players 1 and 2 have the preferences

$$D \succ FD \succ FF$$
$$d \succ f$$

respectively, meaning that if any type of a player chooses a maximal strategy in a subgame, then his choice is made in accordance with backward induction. Demonstrating that this conclusion holds in general for generic perfect information games is the main result of the present paper.

5.2. **Main results.** In analogy with Aumann's [6] Theorems A and B, it is established that

- ... any vector of maximal strategies in a subgame of a generic perfect information game, in a state where there is common certain belief of admissible subgame consistency, leads to the backward induction outcome in the subgame (Prop. 2). Hence, by substituting $\bigcap_{h\in H} B_i(h)[rat_j(h)]$ for $B_i[rat_j]$, the present analysis yields support to Aumann's conclusion, namely that if there is common knowledge (or certain belief) of an

[12] Note that the requirement of such 'belief in each subgame of opponent rationality' allows a player to update his belief about the type of his opponent. Hence, there is no assumption of 'epistemic independence' between different agents in the sense of Stalnaker [45]; cf. Remark 2 of Sect. 5.2. Still, the requirement can be considered a non-inductive analogue to 'forward knowledge of rationality' as defined by Balkenborg & Winter [9].

appropriate form of (belief of) rationality, then the backward induction outcome results.

- ... for any generic perfect information game, common certain belief of admissible subgame consistency is possible (Prop. 3). Hence, the result of Prop. 2 is not empty.

Proposition 2. *Consider a finite generic extensive game of perfect information Γ with corresponding strategic game G. If, for some belief system for G, $\omega \in CKA^*$, then, for each $h \in H$, $C^{t(\omega)}(h) \subseteq S(z(p|_h))$, where p denotes the unique subgame-perfect equilibrium.*

Proof. 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 $\forall g \geq 2$,

$$\begin{aligned} K^gE &= K_1K^{g-1}E \cap K_2K^{g-1}E \\ &\subseteq K_1K_1K^{g-2}E \cap K_2K_2K^{g-2}E \\ &= K_1K^{g-2}E \cap K_2K^{g-2}E = K^{g-1}E \,. \end{aligned}$$

Even though the truth axiom ($K_iE \subseteq E$) is not satisfied, it follows that mutual certain belief of A^* implies that A^* is true, since $A^* := A_1^* \cap A_2^*$ is an event that concerns the type vector: $KA^* = K_1A^* \cap K_2A^* \subseteq K_1A_1^* \cap K_2A_2^* = A_1^* \cap A_2^* = A^*$ since, for each i, $K_iA_i^* = A_i^*$. Hence, *(i)* $\forall g \geq 1$, $K^gA^* \subseteq K^{g-1}A^*$, and *(ii)* $\exists g' \geq 0$ such that $K^gA^* = CKA^*$ for $g \geq g'$ since Ω is finite.

In view of these properties, it is sufficient to show for any $g = 0, \dots, M-2$ that if there exists a belief system with $\omega \in K^gA^*$, then $C^{t(\omega)}(h) \subseteq S(z(p|_h))$ for any $h \in H^{M-1-g}$. This is established by induction.

($g = 0$) Let $h \in H^{M-1}$. First, consider j with a singleton action set at h. Then trivially $C_j^{t_j}(h) = S_j(h) = S_j(z(p|_h))$. Now, consider i with a non-singleton action set at h; since Γ has perfect information, there is at most one such i. Let $t_i = t_i(\omega)$ for some $\omega \in K^0A^* = A^*$. Then it follows that $C_i^{t_i}(h) = S_i(z(p|_h))$ since Γ is generic and $\omega \in A^* \subseteq [u_i] \cap [cau_i]$.

($g = 1, \dots, M-2$) Suppose that it has been established for $g' = 0, \dots, g-1$ that if there exists a belief system with $\omega \in K^{g'}A^*$, then $C^{t(\omega)}(h') \subseteq S(z(p|_{h'}))$ for any $h' \in H^{M-1-g'}$. Let $h \in H^{M-1-g}$. First, consider j with a singleton action set at h. Let $t_j = t_j(\omega)$ for some $\omega \in K^{g-1}A^*$. Then, by Lemma 1 and the premise, $S_j(h) = S_j(h,a)$ and

$$C_j^{t_j}(h) \subseteq C_j^{t_j}(h,a) \subseteq S_j(z(p|_{(h,a)}))$$

if a is a feasible action vector at h. This implies that

$$C_j^{t_j}(h) \subseteq \bigcap\nolimits_a S_j(z(p|_{(h,a)})) \subseteq S_j(z(p|_h))\,.$$

Now, consider i with a non-singleton action set at h; since Γ has perfect information, there is at most one such i. Let $t_i = t_i(\omega)$ for some $\omega \in K^gA^*$. The preceding argument implies that $C_j^{t_j}(h) \subseteq \bigcap_a S_j(z(p|_{(h,a)}))$ whenever $t_j \in T_j^{t_i}$ since $\omega \in K^gA^* \subseteq K_iK^{g-1}A^*$. Let $s_i \in S_i(h)$ be a strategy that differs from $p_i|_h$ by assigning a different action at h (i.e., $z(s_i, p_j|_h) \neq$

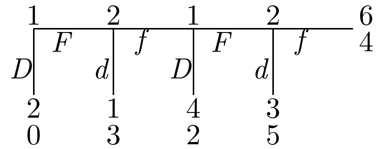

FIGURE 2. A four-legged centipede game

$z(p|_h)$ and $s_i(h') = p_i|_h(h')$ whenever $S_i(h) \supset S_i(h')$). Write $\mathbf{x}_{S_j}$ for the act on S_j that $p_i|_h$ can be viewed as, and write $\mathbf{y}_{S_j}$ for the act on S_j that s_i can be viewed as. 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 (recalling that $\Omega_j(\cdot) := S_j(\cdot) \times T_j$),

$$\mathbf{x}_{\cap_a \Omega_j(z(p|_{(h,a)}))} \quad \text{strongly dominates} \quad \mathbf{y}_{\cap_a \Omega_j(z(p|_{(h,a)}))}$$

by backward induction since Γ is generic and $\omega \in K^g A^* \subseteq [u_i]$. Since $C_j^{t_j}(h) \subseteq \bigcap_a S_j(z(p|_{(h,a)}))$ whenever $t_j \in T_j^{t_i}$, it follows that, $\forall t_j \in T_j^{t_i}$,

$$\mathbf{x}_{C_j^{t_j}(h) \times \{t_j\}} \quad \text{strongly dominates} \quad \mathbf{y}_{C_j^{t_j}(h) \times \{t_j\}},$$

and, thus, $\omega \in K^g A^* \subseteq B_i(h)[rat_j(h)]$ implies that

$$\mathbf{x} \succ^{t_i}_{S_j(h) \times T_j} \mathbf{y} \quad \text{and} \quad \mathbf{x}_{S_j} \succ^{t_i}_{S_j(h)} \mathbf{y}_{S_j}.$$

By Lemma 1 and the premise that $C_i^{t_i}(h, a) \subseteq S_i(z(p|_{(h,a)}))$ if a is a feasible action vector at h, it follows that $C_i^{t_i}(h) \subseteq S_i(z(p|_h))$. □

Remark 2. It follows from the proof of Prop. 2 that, for a generic perfect information game with $M-1$ stages, it is sufficient with $M-2$ order mutual certain belief of admissible subgame consistency in order to obtain backward induction. Hence, $K^{M-2}A^*$ can be substituted for CKA^*.

Backward induction will not be obtained, however, if CBA^* is substituted for CKA^*. This can be shown by considering a counter-example that builds on the four-legged centipede game of Fig. 2 and the belief system of Table 2. In the table the preferences of each type t_i of any player i are represented by a vNM utility function $v_i^{t_i}$ satisfying $v_i^{t_i} \circ z = u_i$ and a 1 or 3-level LPS on $S_j \times T_j$, where $T_1 = \{t_1', t_1'', t_1'''\}$ and $T_2 = \{t_2', t_2''\}$. Inspection shows that $A^* = S \times \{t_1', t_1''\} \times \{t_2', t_2''\}$, while type t_1''' of player 1 satisfies neither 'caution' nor 'belief in each subgame of opponent rationality'. Provided that i is of a type in $\{t_i', t_i''\}$, it follows that i *believes* given (s, t_1, t_2) that the opponent is of a type in $\{t_j', t_j''\}$. This implies that $CBA^* = A^*$. Since FD is the maximal strategy for t_1'' and fd is the maximal strategy for t_2'', it follows that preferences consistent with common *belief* of admissible subgame consistency need not reflect backward induction. However, 2 does not *certainly believe* given (s, t_1, t_2'') that the opponent is *not* of type t_1'''. Therefore, $KA^* = A^* = S \times \{t_1', t_1''\} \times \{t_2'\}$, while $KKA^* = \emptyset$. Hence, preferences that yield maximal

TABLE 2. A belief system for the game of Fig. 2.

t_1' :	t_2'	t_2''	t_1'' :	t_2'	t_2''	t_1''' :	t_2'	t_2''
d	$\left(\frac{4}{5},\frac{7}{10},\frac{7}{12}\right)$	$\left(0,\frac{1}{10},\frac{1}{12}\right)$	d	$\left(\frac{3}{5},\frac{5}{10},\frac{5}{12}\right)$	$\left(0,\frac{1}{10},\frac{1}{12}\right)$	d	(0)	(0)
fd	$\left(0,\frac{1}{10},\frac{1}{12}\right)$	$\left(\frac{1}{5},\frac{1}{10},\frac{1}{12}\right)$	fd	$\left(0,\frac{1}{10},\frac{1}{12}\right)$	$\left(\frac{2}{5},\frac{3}{10},\frac{3}{12}\right)$	fd	(0)	(0)
ff	$\left(0,0,\frac{1}{12}\right)$	$\left(0,0,\frac{1}{12}\right)$	ff	$\left(0,0,\frac{1}{12}\right)$	$\left(0,0,\frac{1}{12}\right)$	ff	$\left(\frac{1}{2}\right)$	$\left(\frac{1}{2}\right)$

t_2' :	t_1'	t_1''	t_1'''	t_2'' :	t_1'	t_1''	t_1'''
D	$\left(\frac{1}{2},\frac{1}{3},\frac{1}{4}\right)$	$\left(0,\frac{1}{6},\frac{1}{8}\right)$	$(0,0,0)$	D	$\left(1,\frac{1}{2},\frac{1}{3}\right)$	$(0,0,0)$	$\left(0,0,\frac{1}{12}\right)$
FD	$\left(0,\frac{1}{6},\frac{1}{8}\right)$	$\left(\frac{1}{2},\frac{1}{3},\frac{1}{4}\right)$	$(0,0,0)$	FD	$\left(0,\frac{1}{4},\frac{1}{6}\right)$	$(0,0,0)$	$\left(0,0,\frac{1}{12}\right)$
FF	$\left(0,0,\frac{1}{8}\right)$	$\left(0,0,\frac{1}{8}\right)$	$(0,0,0)$	FF	$\left(0,0,\frac{1}{6}\right)$	$(0,0,0)$	$\left(0,\frac{1}{4},\frac{1}{6}\right)$

strategies in contradiction with backward induction are not consistent with common *certain belief* of admissible subgame consistency.

The example shows that $\omega \in A^*$ is consistent with $t_i(\omega)$ updating his beliefs about the preferences of his opponent conditional on a subgame being reached. I.e., if 1 is of type t_1', then in the whole game 1 assigns (primary) probability $\frac{4}{5}$ to 2 being of type t_2' with preferences $d \succ fd \succ ff$, while in the subgame defined by 1's second decision node 1 assigns (primary) probability 1 to 2 being of type t_2'' with preferences $fd \succ d \sim ff$. This shows that Stalnaker's [45] assumption of 'epistemic independence' is not made; a player is in principle allowed to learn about the type of his opponent on the basis of previous play. However, in a belief system with $CKA^* \neq \emptyset$, $\omega \in CKA^*$ implies that 1 certainly believes given ω that 2 is of a type with preferences $d \succ fd \succ ff$. In other words, if there is common certain belief of admissible subgame consistency, there is essentially nothing to learn about the opponent.

Proposition 3. *For any finite generic extensive games of perfect information Γ with corresponding strategic game G, there exists a belief system for G with $CKA^* \neq \emptyset$.*

Proof. Construct a belief system with only one type of each player, and write $\Omega := S \times \{(t_1, t_2)\}$. Write, $\forall i \in N$, $\forall m \in \{1, \dots, M-1\}$, $P_j^m := \{p_j|_h | h \in H^m\}$ and, $P_j^M := S_j$. Let, $\forall i \in N$, $\lambda^{t_i} = (\mu_1^{t_i}, \dots, \mu_M^{t_i}) \in \mathbf{L}\Delta(S_j \times \{t_j\})$ satisfy the following requirement: $\forall m \in \{1, \dots, M\}$, $\text{supp}\mu_m^{t_i} = P_j^k \times \{t_j\}$. By letting $\succeq^{t_i}$ be represented by a vNM utility function $\upsilon_i^{t_i}$ satisfying $\upsilon_i^{t_i} \circ z = u_i$ and the LPS λ^{t_i}, then (1) $\forall i \in N$, $[u_i] \cap [cau_i] = \Omega$. Let $\lambda_{S_j}^{t_i}$ denote the marginal of λ^{t_i} on S_j, and let, $\forall h \in H$, $\lambda_{S_j(h)}^{t_i} = (\mu_1^{t_i}|_{S_j(h)}, \dots, \mu_L^{t_i}|_{S_j(h)})$ denote the conditional of $\lambda_{S_j}^{t_i}$ on $S_j(h)$ (see Blume et al. [18], Def. 4.2). By the properties of a subgame-perfect equilibrium, $\forall h \in H$, $\mu_1^{t_i}|_{S_j(h)}(p_j|_h) = 1$ and $p_i|_h \in C_i^{t_i}(h)$. Hence, since likewise $p_j|_h \in C_j^{t_j}(h)$, we have that (2) $\forall i \in N$, $\bigcap_{h \in H} B_i(h)[rat_j(h)] = \Omega$. By (1) and (2), it follows that $CKA^* = A^* = \Omega \neq \emptyset$. $\square$

Remark 3. The constructive proof of Prop. 3 can be used to show, for any game of almost perfect information and for any subgame-perfect equilibrium (in strategies that are not weakly dominated), that the subgame-perfect equilibrium outcome corresponds to a vector of maximal strategies in a state where there is common certain belief of admissible subgame consistency.

6. Discussion

Consider a generic perfect information game. Say that a type's preferences are *in accordance with backward induction* if, in any subgame, a strategy is maximal only if it is consistent with the backward induction outcome. Using this terminology, Prop. 2 can be restated as follows: *Under common certain belief of admissible subgame consistency in a generic perfect information game, players are of types with preferences that are in accordance with backward induction.* Furthermore, common certain belief of admissible subgame consistency implies that players cannot admit the possibility that the opponent is of a type with preferences *not* in accordance with backward induction. This reflects in spirit a conclusion that can be drawn from Aumann's analysis.

However, since admissible subgame consistency is imposed on preferences, reaching 2's decision node and 1's second decision node in the centipede game of Fig. 1 does not contradict common certain belief of admissible subgame consistency. Of course, these decision nodes will not be reached if players are rational (i.e. choose maximal strategies). But that players satisfy 'belief in each subgame of opponent rationality' does not imply that they will actually choose maximal strategies; rather, it means that they 'believe' (in a sense that generalizes belief with probability one) in any subgame that their opponent will be rational.

6.1. **Certain Belief vs. Belief.** The term 'certain belief' (cf. Sect. 2.3) signifies, in the present paper, that the complement of a certainly believed event is not taken into account. This notion of 'certain belief' is strong; in fact, it is the strongest form considered by Morris [37]. Each of the notions 'approximate knowledge' (Monderer & Samet [36], Börgers [22]), 'first-order knowledge' (Brandenburger [23]), and 'certainty' (Ben-Porath [14]) represents an effective weakening of 'certain belief' exactly because the complement of a believed event *is* allowed to be taken into account.[13]

The analysis in the present paper can use 'certain belief' in the sense that the complement of a certainly believed event is *not* taken into account, since certain belief of admissible (subgame) consistency concerns preferences and

[13] As the present notion of 'certain belief' is derived from preferences, it does not satisfy the truth axiom. Hence, an event can be certainly believed even though the true state is an element of the complement of the event. This is in line with the belief definitions of e.g. Brandenburger & Dekel [25], Morris [37] (even in the strongest form that he considers), and Epstein [29].

not choice. When there is common certain belief of admissible (subgame) consistency in the sense of Sect. 3 (5), then

- each player i certainly believes (in the sense of not taking into account the complement) that his opponent is of a type with preferences that satisfy 'caution' and 'belief (in each subgame) of opponent rationality',
- each player i certainly believes (in the sense of not taking into account the complement) that his opponent certainly believes (in the sense of not taking into account the complement) that he himself is of a type with preferences that satisfy 'caution' and 'belief (in each subgame) of opponent rationality',

and so on.

By 'caution' it is not the case that a player certainly believes that the opponent will not play any particular strategy. On the contrary, 'caution' imposes that a player takes into account all opponent strategies, implying that he deems possible that any subgame in the extensive game be reached.[14] By 'belief (in each subgame) of opponent rationality' the type of any player merely 'believes' (conditional on any subgame) that the opponent will choose a maximal strategy, in the sense that the type's preferences (in the subgame) is admissible on a subset of the opponent's set of maximal strategies (in the subgame).

6.2. **Rationality Orderings.** The constructive proof of Prop. 3 shows how common certain belief of admissible subgame rationality may lead a type t_i of player i to have preferences over i's strategies that are represented by a vNM utility function $v_i^{t_i}$ satisfying $v_i^{t_i} \circ z = u_i$ and an LPS $\lambda_{S_j}^{t_i} = (\mu_1^{t_i}, ..., \mu_L^{t_i}) \in \mathbf{L}\Delta(S_j)$ with more than two levels of subjective probability distributions (i.e. $L > 2$). E.g., in the centipede game of Fig. 1, common certain belief of admissible subgame rationality implies that any type t_2 of player 2 has preferences that can be represented by u_2 and $\lambda_{S_1}^{t_2} = (\mu_1^{t_2}, \mu_2^{t_2}, \mu_3^{t_2})$ where $\text{supp}\mu_1^{t_2} = \{D\}$, $\text{supp}\mu_2^{t_2} = \{D, FD\}$, and $\text{supp}\mu_3^{t_2} = S_1$. Within the 'rational choice' approach one may interpret $\text{supp}\mu_1^{t_i}$ to consist of strategies for j that are "most rational", $\text{supp}\mu_L^{t_i} \backslash \bigcup_{\ell' < L} \text{supp}\mu_{\ell'}^{t_i}$ to consist of strategies for j that are "completely irrational", and $\text{supp}\mu_\ell^{t_i} \backslash \bigcup_{\ell' < \ell} \text{supp}\mu_{\ell'}^{t_i}$, for $\ell = 2, \ldots, L-1$, to consist of strategies for j that are at "intermediate

[14] Note that a hypothetical knowledge operator as suggested by Samet [43] is not needed here since any player considers all paths through the game to be possible.

The present analysis is consistent with the following view expressed by Arló-Costa & Bicchieri ([2], pp. 187–188): "A player who is considering alternative actions ... has yet to move, hence her reasoning is hypothetical, and the statement 'If I were to play A, then ...' is subjunctive, but not a counterfactual. Its antecedent is neither true or false, since nothing has happened yet." "Agents who face an interactive decision problem usually start by assessing all their possible moves, and the possible counter-moves of other players. They engage, that is, in hypothetical reasoning about what could possibly happen after each of their and others' counter moves."

The distinction that I make between preferences and choice echoes a distinction made by Kramarz [33] between "contemplating an action" and actually playing an action.

degrees of rationality". Furthermore, for any $\ell = 2, \dots, L$, t_i deems any strategy in $\bigcup_{\ell' < \ell} \text{supp}\mu_{\ell'}^{t_i}$ infinitely more likely than any strategy not having this property. This illustrates that $(\text{supp}\mu_1^{t_i}, \dots, \text{supp}\mu_L^{t_i} \backslash \bigcup_{\ell' < L} \text{supp}\mu_{\ell'}^{t_i})$ corresponds closely to what Battigalli [11] calls a *rationality ordering* for j.

However, the present construction of such a rationality ordering differs from the one proposed by Battigalli. This difference is along two dimensions:

1. Battigalli considers best responses in reachable subgames only (see his Def. 2.1), while here belief of opponent rationality is held in *all* subgames (cf. 'belief in each subgame of opponent rationality').
2. Battigalli considers best responses given beliefs where opponent strategies that are less than "most rational" are given positive probability, while here each player certainly believes that the opponent is of an admissibly subgame consistent type and believes that he chooses rationally.

This difference has the following consequences:

- Although Battigalli's construction of a rationality ordering also promotes the backward induction outcome in any generic perfect information game, his proof (cf. Battigalli [12]) is not as directly tied to the procedure of backward induction.
- Battigalli's construction of a rationality ordering promotes the forward induction outcome in an extended version of the "Battle-of-the-Sexes" (BoS) game where the BoS game is preceded by 1 being offered an outside option that is preferred by 1 to 2's most preferred outcome in the BoS game. This conclusion is not reached in the present analysis (see Remark 3 of Sect. 5) since there is no choice situation in which 1 under all circumstances will have a particular preference between his BoS strategies.[15]

This also implies that the epistemic foundation for backward induction offered here differs from the epistemic foundation for backward (and forward) induction provided by Battigalli & Siniscalchi [13].

Appendix. 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 *reflexive* 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*

[15]AD show how the concept of admissible consistency can be strengthened so that the forward induction outcome is promoted in the BoS game with an outside option.

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 $\mathbf{x}_F \succ_F \mathbf{y}_F$.
- *continuity* if 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 $\mathbf{x}'_F \succ_F \mathbf{y}_F \succ_F \mathbf{x}''_F$.

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 completeness* if, $\forall f \in F$, $\succeq_{\{f\}}$ is complete.
- *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, $f \in F$ and $\succeq_F$ is conditionally complete, 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 $e \gg f$. Say that $\mathbf{y}_F$ is *maximal* w.r.t. $\succeq_E$ if there is no $\mathbf{x}_F$ such that $\mathbf{x}_F \succ_E \mathbf{y}_F$.

If $\upsilon : Z \to \mathbb{R}$ is a vNM utility function, abuse notation slightly by writing $\upsilon(x) = \sum_{z \in Z} x(z)\upsilon(z)$ whenever $x \in \Delta(Z)$ is an objective randomization. Say that $\mathbf{x}_E$ *strongly dominates* $\mathbf{y}_E$ if, $\forall f \in E$, $\upsilon(\mathbf{x}_E(f)) > \upsilon(\mathbf{y}_E(f))$. Say that $\mathbf{x}_E$ *weakly dominates* $\mathbf{y}_E$ 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 if $\mathbf{x}_F \succ_F \mathbf{y}_F$ whenever $\mathbf{x}_E$ weakly dominates $\mathbf{y}_E$.

The following two representation results can now be stated. The first one — which follows directly from the von Neumann-Morgenstern theorem on expected utility representation — requires the notion of conditional representation: Say that $\succeq_F$ is *conditionally represented* by υ if (a) $\succeq_F$ is nontrivial and (b) $\mathbf{x}_F \succeq_{\{f\}} \mathbf{y}_F$ iff $\upsilon(\mathbf{x}_F(f)) \geq \upsilon(\mathbf{y}_F(f))$ whenever f is not Savage-null.

Proposition A1. *If $\succeq_F$ is reflexive and transitive, and satisfies objective independence, nontriviality, conditional completeness, conditional continuity, and non-null state independence, then there exists a vNM utility function $\upsilon : Z \to \mathbb{R}$ such that $\succeq_F$ is conditionally represented by υ.*

The second result, due to Blume et al. ([18], Theorem 3.1), 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 $\mathbf{L}\Delta(F)$ 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 A2. *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 $\lambda = (\mu_1, ..., \mu_L) \in \mathbf{L}\Delta(F)$*

such that $\mathbf{x}_F \succeq_F \mathbf{y}_F$ *iff*

$$\left(\sum_{f\in F} \mu_\ell(f)\upsilon(\mathbf{x}_F(f))\right)_{\ell=1}^L \geq_L \left(\sum_{f\in F} \mu_\ell(f)\upsilon(\mathbf{y}_F(f))\right)_{\ell=1}^L .$$

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).

References

[1] Anscombe, F.J., Aumann, R.: A definition of subjective probability. Annals of Mathematical Statistics **34** (1963) 199–205.

[2] Arló-Costa, H., and Bicchieri, C.: Games and conditionals. In "Theoretical Aspects of Rationality and Knowledge, Proceedings of the Seventh Conference" (Gilboa, Ed.) (1998).

[3] Armbruster, W., Böge, W.: Bayesian game theory. In "Game Theory and Related Topics" (Moeschlin, Pallaschke, Eds.), North-Holland, Amsterdam (1979).

[4] Asheim, G.B.: Proper consistency. University of Oslo (1999).

[5] Asheim, G.B., Dufwenberg, M.: Admissibility and common belief. University of Oslo (1999).

[6] Aumann, R.: Backward induction and common knowledge of rationality. Games Econ. Beh. **8** (1995) 6–19.

[7] Aumann, R.: On the centipede game Games Econ. Beh. **23** (1998) 97–105.

[8] Aumann, R., Brandenburger, A.: Epistemic conditions for Nash equilibrium. Econometrica **63** (1995) 1161–1180.

[9] Balkenborg, D., Winter, E.: A necessary and sufficient epistemic condition for player backward induction. J. Math. Econ. **27** (1997) 325–345.

[10] Basu, K.: On the non-existence of a rationality definition for extensive games. Int. J. of Game Theory **19** (1990), 33–44.

[11] Battigalli, P.: Strategic rationality orderings and the best rationalization principle. Games Econ. Beh. **13** (1996) 178–200.

[12] Battigalli, P.: On rationalizability in extensive games. J. Econ. Theory **74** (1997) 40–61.

[13] Battigalli, P., Siniscalchi, M: Hierarchies of conditional belief and interactive epistemology in dynamic games. European University Institute working paper ECO No. 98/29.

[14] Ben-Porath, E.: Rationality, Nash equilibrium, and backwards induction in perfect information games. Rev. Econ. Stud. **64** (1997) 23–46.

[15] Bicchieri, C.: Self-refuting theories of strategic interaction: A paradox of common knowledge. Erkenntnis **30** (1989) 69–85.

[16] Binmore, K.: Modelling rational players I. Econ. and Phil. **3** (1987) 179–214.

[17] Binmore, K.: Backward induction and rationality. DP 95–10, University College London (1995).

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

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

[20] Böge, W., Eisele, T.: On the solutions of Bayesian games. Int. J. Game Theory **8** (1979) 193–215.

[21] Bonanno, G.: Branching time logic, perfect information games and backward induction. Univeristy of California, Davis (1998).

[22] Börgers, T.: Weak dominance and approximate common knowledge. J. Econ. Theory **64** (1994) 265–276.

[23] Brandenburger, A.: Lexicographic probabilities and iterated admissibility. In "Economic Analysis of Markets and Games" (Dasgupta, Gale, Hart, Maskin, Eds.), MIT Press, Cambridge, MA (1992).

[24] Brandenburger, A.: On the existence of a 'complete' belief model. Harvard Business School Working Paper 99-056 (1998).

[25] Brandenburger, A., Dekel, E.: Hierarchies of beliefs and common knowledge. J. Econ. Theory **59** (1993) 189–198.

[26] Brandenburger, A., Keisler, H.J.: An impossibility theorem on beliefs in games. Harvard Business School Working Paper 00-010 (1999).

[27] Dekel, E., Fudenberg, F.: Rational behavior with payoff uncertainty. J. Econ. Theory **52** (1990) 243–267.

[28] Dekel, E., Gul, F.: Rationality and knowledge in game theory. In "Advances in Economics and Econometrics: Theory and Applications" (Kreps, Wallis, Eds.), Cambridge University Press, Cambridge, UK (1997).

[29] Epstein, L.G.: Preference, rationalizability and equilibrium. J. Econ. Theory **73** (1997) 1–29.

[30] Epstein, L.G., Wang, T.: "Beliefs about beliefs" without probabilities. Econometrica **64** (1996) 1343–1373.

[31] Gul, F.: Rationality and coherent theories of strategic behavior. J. Econ. Theory **70** (1997) 1–31.

[32] Kaneko, M.: On "paradoxes" in the centipede and chain-store games I: Nonepistemic considerations. IPPS-DP 810, University of Tsukuba (1999).

[33] Kramarz, F.: How agents plan their actions in games: a model of players' reasoning with common knowledge of rationality. WP 9322, INSEE (1993).

[34] Mailath, G., Samuelson, L., Swinkels, J.: Extensive form reasoning in normal form games. Econometrica **61** (1993) 273–302.

[35] Mertens, J.-M., Zamir, S.: Formulation of Bayesian analysis for games of incomplete information. Int. J. Game Theory **14** (1985) 1–29.

[36] Monderer, D., Samet, D.: Approximating common knowledge with common beliefs. Games Econ. Beh. **1** (1989) 170–190.

[37] Morris, S.: Alternative notions of belief. In "Epistemic Logic and the Theory of Decisions in Games" (Bacharach, Gerard-Varet, Mongin, Shin, Eds.), Kluwer, Dordrecht (1997).

[38] Osborne, M.J., Rubinstein, A.: "A Course in Game Theory", MIT Press, Cambridge, MA (1994).

[39] Rabinowicz, W.: Grappling with the centipede: Defence of backward induction for BI-terminating games. Econ. and Phil. **14** (1997) 95–126.

[40] Reny, P.J.: Common belief and the theory of games with perfect information. J. Econ. Theory **59** (1993) 257–274.

[41] Rosenthal, R.: Games of perfect information, predatory pricing and the chain-store paradox. J. Econ. Theory **25** (1981) 92–100.

[42] Rubinstein, A.: Comments on the Interpretation of Game Theory. Econometrica **59** (1991) 909–924.

[43] Samet, D.: Hypothetical knowledge and games with perfect information. Games Econ. Beh. **17** (1996) 230–251.

[44] Schuhmacher, F.: Proper rationalizability and backward induction. Forthcoming in Int. J. Game Theory.

[45] Stalnaker, R.: Belief revision in games: forward and backward induction. Math. Soc. Sci. **36** (1998) 57–68.

[46] Tan, T., Werlang, S.R.C.: The Bayesian foundations of solution concepts of games. J. Econ. Theory **45** (1988) 370–391.

[47] Vilks, A.: Knowledge of the game, relative rationality, and backward induction without counterfactuals. Handelhochschule Leipzig (1998).