Asheim, Geir B.

Working Paper

On the epistemic foundation for backward induction

Memorandum, No. 1999,30

Provided in Cooperation with:
Department of Economics, University of Oslo

Suggested Citation: Asheim, Geir B. (1999) : On the epistemic foundation for backward induction, Memorandum, No. 1999,30, University of Oslo, Department of Economics, Oslo

This Version is available at:
http://hdl.handle.net/10419/63154

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
MEMORANDUM

No 30/99

On the Epistemic Foundation for Backward Induction

By
Geir B. Asheim

Department of Economics
University of Oslo
List of the last 10 Memoranda:

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

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] impossibility 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 \emph{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 per-
fect 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 sub-
game 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 cer-
tain 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 prob-
ability 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 in-
consistent 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 pref-
erences.\footnote{Instead of imposing rational choice in the sense that a driver \emph{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 \emph{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 present analysis allows, but do not require, subjective prob-
obabilities, 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 frame-
work in which extensive games will be analyzed. AD’s characterization of
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: \( H^{m+1} := \{ (h, a) \mid h \in H^m \text{ and } a \in A(h) \} \). 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 \( S(h') \supseteq S(h) \) 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 \( v_i^{t_i} : Z \to \mathbb{R} \) that assigns a payoff to any outcome (cf. the Appendix).\(^5\) Being a vNM utility function, \( v_i^{t_i} \) 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)
  \( \omega \in \Omega \), \( t_i(\omega) \) denote the projection of \( \omega \) on \( T_i \), and let, for any event \( E \subseteq \Omega \), \( E_i^{t_i} := \{(s_j, t_j) \in S_j \times T_j | \exists (s_i', s_j, t_i', t_j') \in E \) \( s.t. \) \( (s_j', t_j') = (s_j, t_j) \) 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 \( \lambda^{t_i} = (\mu_1^{t_i}, \ldots, \mu_j^{t_i}) \in L(\Delta(S_j \times T_j)) \). Then an event is ‘certainly believed’ if no element of the complement is assigned positive probability by some probability distribution in \( \lambda^{t_i} \):

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

where \( \kappa_j^{t_i} := \text{supp} \lambda_j^{t_i} (\subseteq S_j \times T_j) \). 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 \( \beta_j^{t_i} := \text{supp} \mu_1^{t_i} (\subseteq S_j \times T_j) \). It follows that \( K_i E \subseteq B_i E \) (i.e. ‘certain belief’ implies ‘belief’) since \( \beta_j^{t_i} = \text{supp} \mu_1^{t_i} \subseteq \kappa_j^{t_i} = \text{supp} \lambda_j^{t_i} = \cup_{\ell=1}^{t_i} \text{supp} \mu_1^{t_i} \).

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

\[
k_j^{t_i} := \{(s_j, t_j) \in S_j \times T_j | (s_j, t_j) \text{ is not Savage-null acc. to } \geq^{t_i} \}
\]

\(^5\)If conditional completeness is strengthened to completeness, then it follows from Blume et al. [18] that \( \succeq^{t_i} \) is represented by \( v_i^{t_i} \) and a lexicographic probability system (LPS) \( \lambda^{t_i} = (\mu_1^{t_i}, \ldots, \mu_j^{t_i}) \in L(\Delta(S_j \times T_j)) \) (cf. Prop. A2 of the Appendix). If, in addition, conditional continuity is strengthened to continuity, then \( \succeq^{t_i} \) is represented by \( v_i^{t_i} \) 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 \( \text{supp} \mu^t_i \) is the set of opponent strategy-type pairs that \( t_i \) does not deem Savage-null.

To generalize \( \beta_j^t \) (and thus \( B_iE \)) to incomplete preferences, say that \( \succeq^t_i \) is admissible on \( \beta_j \), where \( \emptyset \neq \beta_j \subseteq S_j \times T_j \), if \( x \succeq^t_i y \) whenever \( x \beta_j \) weakly dominates \( y \beta_j \). If \( \succeq^t_i \) is admissible on \( \beta_j \), then any \( (s'_j, t'_j) \in \beta_j \) is deemed infinitely more likely than any \( (s''_j, t''_j) \in S_j \times T_j \setminus \beta_j \). 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 \( \beta'_j \) and \( \beta''_j \). Since, in addition, \( \succeq^t_i \) is admissible on \( \kappa_j^t \), 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 \( \beta_j^t \):\(^8\)

\[ \succeq^t_i \text{ is admissible on } \beta_j^t \text{ and } \beta_j \supseteq \beta_j^t \text{ whenever } \succeq^t_i \text{ is admissible on } \beta_j. \]

This generalizes the case of complete preferences, since in that case \( \text{supp} \mu^t_i \) 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.e. \( \beta_j^t \subseteq \kappa_j^t \). If \( \beta_j^t \neq \kappa_j^t \), 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

\[ B_iE \cap B_iF = B_i(E \cap F) \]
\[ K_iE \cap K_iF = K_i(E \cap F) \]
\[ B_i\emptyset = \emptyset \]
\[ B_iE \subseteq K_iB_iE \]
\[ K_iE \subseteq K_iK_iE \]
\[ -B_iE \subseteq K_i(-B_iE) \]
\[ -K_iE \subseteq K_i(-K_iE). \]

Since \( K_iE \subseteq B_iE \) implies that \( K_i\emptyset = \emptyset, B_i\Omega = \Omega \), \( B_iE \subseteq B_iB_iE \) and \( -B_iE \subseteq B_i(-B_iE) \), 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 \( \omega \) if \( \omega \in B_iE \) (or equivalently, \( \beta_j^t(\omega) \subseteq E_j^t(\omega) \)). Write \( BE := B_iE \cap B_2E \). Say that there is mutual belief of \( E \subseteq \Omega \) given \( \omega \) if \( \omega \in BE \). Write \( CBE := BE \cap BB E \cap BBB E \cap \ldots \). Say that there common belief of \( E \subseteq \Omega \) given \( \omega \) if \( \omega \in CBE \). Say that \( i \) certainly believes the event \( E \subseteq \Omega \) given \( \omega \) if \( \omega \in K_iE \) (or equivalently, \( \kappa_j^t(\omega) \subseteq E_j^t(\omega) \)). Write \( KE := K_1E \cap K_2E \). Say that there is mutual certain belief of \( E \subseteq \Omega \) given \( \omega \) if \( \omega \in KE \). Write \( CKE := KE \cap KK E \cap KK KE \cap \ldots \). Say that there is common certain belief of \( E \subseteq \Omega \) given \( \omega \) 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 $x_{S_j}$ 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 $x_{S_j}$ 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

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

is non-empty and supports any maximal mixed strategy. Refer to $C^{t_i}(h)$ as $t_i$’s choice set in the subgame $h$, and refer to $C^{t_i}(h) : H \rightarrow 2^{S_i(h)} \backslash \{\emptyset\}$ as $t_i$’s choice function. Write $C^{t_i} := C^{t_i}(\emptyset)$, and write, for any $h \in H$, $C^{t_i}(h) := C^{t_i}(h) \times C^{t_i}(h)$.

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 $s_i \in C^{t_i}(h)$, then $s_i \in C^{t_i}(h')$ 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 $x_{S_j}$ such that $x_{S_j} \succeq^{t_i}_{S_j(h')} y_{S_j}$, where $x_{S_j}$ assigns $z(x_i, s_j)$ to any $s_j \in S_j$ with $x_i \in \Delta(S_i(h'))$, and where $y_{S_j}$ 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, $x_{S_j}$ can be chosen such that $x_{S_j}(s_j) = y_{S_j}(s_j)$ for all $s_j \in S_j \backslash S_j(h')$. This implies that $x_{S_j} \succeq^{t_i}_{S_j(h')}$, 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 $\kappa_i^{t_i} \cap S_j(h) \times T_j \neq \emptyset$ (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 $i$, $u_i : Z \rightarrow \mathbb{R}$ be a vNM utility function that assigns payoff to any outcome. Then the pair of the extensive game form and the vNM utility functions $(u_i)_{i \in N}$ is a finite extensive game of almost perfect information, $\Gamma$. Let $G = (S_i, u_i)_{i \in N}$ be the corresponding finite strategic game, where for each $i$, the vNM utility function $u_i : S \rightarrow \mathbb{R}$ is defined by $u_i = u_i \circ z$ (i.e., $u_i(s) = u_i(z(s))$ 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

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

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 \(s_i \in C_{t_i}^i\), and where the event that \(i\) is rational is defined as\(^9\)

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

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:

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

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

\(^9\)See e.g. Epstein ([29], Sect. 6) for a presentation of this approach in a general context.
where $T^t_i := \text{proj}_{T_i} \kappa^t_j$ denotes the set of opponent types that $t_i$ does not deem Savage-null.\(^\text{10}\)

- If $\omega \in [\text{cau}]$, 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, $\forall (s_j, t_j) \in S_j \times T^t_j(\omega)$, $\omega \notin K_i((s'_j, s'_2, t'_1, t'_2) \in \Omega((s'_j, t'_j) \neq (s_j, t_j))$ (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 $S_i$, $\succeq_{t_i(\omega)}$) is admissible on $S_j$.

- If $\omega \in B_i[\text{rat}_j]$, then $i$ believes given $\omega$ that $j$ is rational.

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

$$A_i := [u_i] \cap [\text{cau}] \cap B_i[\text{rat}_j].$$

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 $\omega \in \text{CKA}$.

### 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 $i$, $v_i(z) \neq v_i(z')$ 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 $\omega \in [\text{cau}] \cap B_i[\text{rat}_j]$ and $\succeq_{t_i(\omega)}$ is complete, then $\succeq_{t_i(\omega)}$ can be represented by $v_i^{t_i(\omega)}$ and an LPS $\lambda^{t_i(\omega)}(\mu_1^{t_i(\omega)}, \ldots, \mu_{L_j}^{t_i(\omega)}) \in L\Delta(S_j \times T_j)$ 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 $r_j \in C_j^{t_j(\omega)}$. Note that $\omega \in [\text{cau}] \cap B_i[\text{rat}_j]$ does not imply — but is consistent with — $\succeq_{t_i(\omega)}$ being complete, while $\omega \in [\text{cau}] \cap B_i[\text{rat}_j]$ is not consistent with $\succeq_{t_i(\omega)}$ being continuous.
that common (true) knowledge of (all subgame) 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 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 $\nu_i$ satisfying $\nu_i(z) = u_i$ 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

\begin{table}[h]
\centering
\begin{tabular}{c c c c}
\hline
$t'_1$: & $t''_1$: & $t'_2$: & $t''_2$: \\
\hline
\hline
$D$ & $(\frac{4}{5}, \frac{1}{10})$ & $(0, \frac{1}{10})$ & $(\frac{3}{5}, \frac{9}{10})$ & $(0, \frac{1}{10})$ \\
$f$ & $(0, \frac{1}{10})$ & $(\frac{1}{5}, \frac{1}{10})$ & $(0, \frac{1}{10})$ & $(\frac{2}{5}, \frac{3}{10})$ \\
\hline
$t'_2$: & $t''_2$: & $t'_1$: & $t''_1$: \\
\hline
\hline
$D$ & $(\frac{1}{2}, \frac{1}{2})$ & $(0, \frac{1}{5})$ & $(1, \frac{1}{2})$ & $(0, 0)$ \\
$FD$ & $(0, \frac{4}{5})$ & $(\frac{1}{2}, \frac{1}{2})$ & $(0, \frac{1}{2})$ & $(0, 0)$ \\
$FF$ & $(0, \frac{1}{8})$ & $(0, \frac{1}{5})$ & $(0, \frac{1}{4})$ & $(0, 0)$ \\
\hline
\end{tabular}
\caption{A belief system for the game of Fig. 1.}
\end{table}

\footnote{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 choosesrationally 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
then it follows from the above definitions that 

\[ \beta \omega \text{ given } \beta \text{ sense that } \nu \text{ s}

that is not implied by \( \nu \beta \). If \( \nu \beta \) is admissible on \( \beta \), then any \( (s_j', t_j') \in \beta \) is deemed infinitely more likely than any \( (s_j'', t_j'') \in \beta \). 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 \( \nu \beta \) is admissible on both \( \beta_j' \) and \( \beta_j'' \). If the set of opponent strategy-type pairs that \( t_i \) does not deem Savage-null in \( h \), \( \kappa_j \cap \Omega_j(h) \) is non-empty — which is implied by \( t_i \) satisfying ‘caution’ — then \( \nu \beta \) is admissible on \( \kappa_j \cap \Omega_j(h) \). Hence, under this assumption, there exists a unique smallest (w.r.t. set inclusion) set on which \( \nu \beta \) is admissible; let this set be denoted \( \beta_j^{i}(h) \):

\[ \nu \beta \subseteq \beta_j^{i}(h) \text{ and } \beta_j \supseteq \beta_j^{i}(h) \text{ whenever } \nu \beta \text{ is admissible on } \beta_j \cdot \]

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

\[ B_i(h)E := \{ \omega \in \Omega | \beta_j^{i}(\omega) \subseteq \beta_j^{i}(h) \} \]

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

\[ B_i(h)E \cap B_i(h)F = B_i(h)(E \cap F) \]

\[ B_i(h)\emptyset = \{ \omega \in \Omega | \kappa_j^{\omega} \cap \Omega_j(h) = \emptyset \} \]

\[ B_i(h)E \subseteq K_iB_i(h)E \]

\[ \neg B_i(h)E \subseteq K_i(\neg B_i(h)E) \]

Since \( K_iE \subseteq B_i(h)E \) implies that \( B_i(h)E \Omega = \Omega \), \( B_i(h)E \subseteq B_i(h)B_i(h)E \) 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 \( \beta_j^{i}(\omega) \neq \emptyset \) in any state \( \omega \) where \( i \) deems it possible that \( h \) can be reached (i.e., \( \kappa_j^{\omega} \cap \Omega_j(h) \neq \emptyset \), meaning that \( \nu \beta \) 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 \( s_j \in C_j^{i}(h) \), and let the event that \( j \) is rational in \( h \) be defined as

\[ [\text{rat}_j(h)] := \{(s_1, s_2, t_1, t_2) \in \Omega \mid s_j \in C_j^{i}(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 \( h \):

\[ B_i(h)[\text{rat}_j(h)] = \{ \omega \in \Omega \mid (r_j, t_j) \in \beta_j^{i}(\omega) \text{ implies } r_j \in C_j^{i}(h) \} \]
If \( \omega \in \bigcap_{h \in H} B_i(h)[rat_j(h)] \), then, conditional on any subgame \( h \), \( i \) believes given \( \omega \) 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 \( \Gamma \) of almost perfect information with corresponding strategic game \( G \). Say that \( i \) is admissibly subgame consistent (with \( \Gamma \) and the preferences of his opponent) given \( \omega \) 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, \( \bigcap_{h \in H} B_i(h)[rat_j(h)] \) 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 $\Gamma$ with corresponding strategic game $G$. If, for some belief system for $G$, $\omega \in \text{CKA}^*$, 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 $\forall g \geq 2$,

$$
K^gE = K_iK^{g-1}E \cap K_2K^{g-1}E
\subseteq K_1K_iK^{g-2}E \cap K_2K_2K^{g-2}E
= K_1K^{g-2}E \cap K_2K^{g-2}E = K^{g-1}E.
$$

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 $A^* := A^*_0 \cup A^*_1$ is an event that concerns the type vector: $KA^* = K_iA^* \cap K_2A^* \subseteq K_1A^*_1 \cap K_2A^*_2 = A^*_1 \cap A^*_2 = A^*$ 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 $K^gA^* = \text{CKA}^*$ for $g \geq g'$ since $\Omega$ is finite.

In view of these properties, it is sufficient to show for any $g = 0, \ldots, M-2$ 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 $C^{t_j}(h) = S_j(h) = S_j(z(p|h))$. Now, consider $i$ with a non-singleton action set at $h$; since $\Gamma$ 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 $C^{t_i}(h) = S_i(z(p|h))$ since $\Gamma$ is generic and $\omega \in A^* \subseteq [u_i] \cap [cau_i]$.

($g = 1, \ldots, M-2$) Suppose that it has been established for $g' = 0, \ldots, g-1$ that if there exists a belief system with $\omega \in K^gA^*$, then $C^{t(\omega)}(h') \subseteq S(z(p|h'))$ for any $h' \in H^{M-1-g'}$. 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

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

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

$$
C^{t_j}(h) \subseteq \bigcap_a S_j(z(p|h, a)) \subseteq S_j(z(p|h)).
$$

Now, consider $i$ with a non-singleton action set at $h$; since $\Gamma$ 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 $C^{t_j}(h) \subseteq \bigcap_a S_j(z(p|h, a)))$ whenever $t_j \in T^{t_j}_i$ 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., $z(s_i, p_{j|h}) \neq$...
z(p|h) and \( s_i(h') = p_i|_h(h') \) whenever \( S_i(h) \supset S_i(h') \). Write \( x_{S_j} \) for the act on \( S_j \) that \( p_i|_h \) can be viewed as, and write \( y_{S_j} \) for the act on \( S_j \) that \( s_i \) can be viewed as. Let \( x \) and \( y \) be the acts on \( S_j \times T_j \) that satisfy \( x(s_j,t_j) = x_{S_j}(s_j) \) and \( y(s_j,t_j) = y_{S_j}(s_j) \) for all \( (s_j,t_j) \). Then (recalling that \( \Omega_j(\cdot) := S_j(\cdot) \times T_j )\),

\[
x_{\cap a_1(S_j(z(p|_h,a)))} \text{ strongly dominates } y_{\cap a_1(S_j(z(p|_h,a)))}
\]

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

\[
x_{C_j^{t_i}(h) \times \{t_j\}} \text{ strongly dominates } y_{C_j^{t_i}(h) \times \{t_j\}},
\]

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

\[
x \succ^{t_i} S_j(h) \times T_j y \quad \text{and} \quad x_{S_j} \succ^{t_i} S_j(h) y_{S_j}.
\]

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 \( v_i^{t_i} o z = u_i \) 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, \( KA^* = A^* = S \times \{t_1', t_1''\} \times \{t_2', t_2''\} \), while \( KKA^* = \emptyset \). Hence, preferences that yield maximal

![Figure 2. A four-legged centipede game](image-url)
By letting $\forall H \in \Omega$ and $\forall \text{probability}$ the properties of a subgame-perfect equilibrium, $\Gamma$ the opponent. $d$ $\lambda$, $A$, $\Omega$: Construct a belief system with only one type of each player, and write $\text{Proof.}$. $certain belief$ $common belief$ $of admissible subgame consistency$. $\Omega$. By (1) and (2), it follows that $\Omega$. Let $\Omega$. By (1) and (2), it follows that $\Omega$. Let $\Omega$. Let $\Omega$. Let $\Omega$. Let $\Omega$. By letting $\geq t_i$ be represented by a vNM utility function $u_i$ satisfying $u_i \circ z = u_i$ and the LPS $\lambda_i$, then (1) $\forall i \in N$, $\exists v_i \mid [\text{cau}] = \Omega$. Let $\lambda_i$ denote the marginal of $\lambda_i$ on $S_j$, and let, $\forall h \in H$, $\lambda_i(S_j(h)) = (\mu_i(S_j(h)) \cdots, \mu_i(S_j(h))$ denote the conditional of $\lambda_i$ on $S_j(h)$ (see Blume et al. [18], Def. 4.2). By the properties of a subgame-perfect equilibrium, $\forall h \in H$, $\mu_i(S_j(h)) = 1$ and $p_j(h) \in C_i(h)$. Hence, since likewise $p_j(h) \in C_i(h)$, we have that (2) $\forall i \in N$, $\exists v_i \mid H_i(h) = \Omega$. By (1) and (2), it follows that $\Omega$. $\Box$

<table>
<thead>
<tr>
<th>$t'_1$ :</th>
<th>$t''_1$ :</th>
<th>$t'''_1$ :</th>
<th>$t'_2$ :</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td>$d$</td>
<td>$d$</td>
<td>$d$</td>
</tr>
<tr>
<td>$fd$</td>
<td>$fd$</td>
<td>$fd$</td>
<td>$fd$</td>
</tr>
<tr>
<td>$ff$</td>
<td>$ff$</td>
<td>$ff$</td>
<td>$ff$</td>
</tr>
<tr>
<td></td>
<td>$D$</td>
<td>$D$</td>
<td>$D$</td>
</tr>
<tr>
<td></td>
<td>$FD$</td>
<td>$FD$</td>
<td>$FD$</td>
</tr>
<tr>
<td></td>
<td>$FF$</td>
<td>$FF$</td>
<td>$FF$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$t'_2$ :</th>
<th>$t''_1$ :</th>
<th>$t''_1$ :</th>
<th>$t''_1$ :</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t'_2$</td>
<td>$t''_1$</td>
<td>$t''_1$</td>
<td>$t''_1$</td>
</tr>
<tr>
<td>$t''_1$</td>
<td>$t''_1$</td>
<td>$t''_1$</td>
<td>$t''_1$</td>
</tr>
<tr>
<td>$t''_1$</td>
<td>$t''_1$</td>
<td>$t''_1$</td>
<td>$t''_1$</td>
</tr>
</tbody>
</table>

Table 2. A belief system for the game of Fig. 2.
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], Borgers [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.\footnote{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].}

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
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 \( \lambda_{S_i}^{t_i} = (\mu_1^{t_i}, ..., \mu_L^{t_i}) \in L\Delta(S_j) \) 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”, \( \text{supp} \mu_1^{t_i} \cup \bigcup_{\ell < i} \text{supp} \mu_1^{t_{\ell}} \) to consist of strategies for \( j \) that are “completely irrational”, and \( \text{supp} \mu_1^{t_i} \cup \bigcup_{\ell < i} \text{supp} \mu_{i'}^{t_{\ell}} \), 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 \( \ell = 2, \ldots, L \), \( t_i \) deems any strategy in \( \bigcup_{\ell' < \ell} \text{supp} \mu_{t_i}^{\ell'} \) infinitely more likely than any strategy not having this property. This illustrates that \((\text{supp} \mu_{t_1}^{1}, \ldots, \text{supp} \mu_{t_L}^{L}\setminus \bigcup_{\ell' < L} \text{supp} \mu_{t_i}^{\ell'})\) 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 \( x_F : F \to \Delta(Z) \) is called an act on \( F \). Write \( x_F \) and \( y_F \) for acts on \( F \). A reflexive and transitive binary relation on the set of acts on \( F \) is denoted by \( \succeq_F \), where \( x_F \succeq_F y_F \) means that \( x_F \) is preferred or indifferent to \( y_F \). As usual, let \( \succ_F \) (preferred to) and \( \sim_F \) (indifferent to).

\(^{15}\)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 $x'_F \succ_F (\text{respectively } \sim_F) x''_F$ iff $\gamma x'_F + (1 - \gamma) y_F \succ_F (\text{respectively } \sim_F) \gamma x''_F + (1 - \gamma) y_F$, whenever $0 < \gamma < 1$ and $y_F$ is arbitrary.
- **nontriviality** if there exist $x_F$ and $y_F$ such that $x_F \succ_F y_F$.
- **continuity** if there exist $0 < \gamma < \delta < 1$ such that $\delta x'_F + (1 - \delta) x''_F \succ_F y_F \succ_F x'_F + (1 - \gamma) x''_F$ whenever $x_F \succ_F y_F \succ_F x'_F$.

If $E \subseteq F$, let $x_E$ denote the restriction of $x_F$ to $E$. Define the **conditional** binary relation $\succeq_E$ by $x'_F \succeq_E x''_F$ if, for arbitrary $y_F$, $(x'_F, y_E) \succeq_F (x''_F, y_E)$, where $-E$ denotes $F \setminus E$. Say that the state $f \in F$ is **Savage-null** if $x_F \sim_f y_F$ for all acts $x_F$ and $y_F$ 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 $\delta x'_F + (1 - \delta) x''_F \succ_f y_F \succ_f \gamma x'_F + (1 - \gamma) x''_F$ whenever $x_F \succ_f y_F \succ_f x'_F$.
- **non-null state independence** if $x_F \succ_{\{e\}} y_F$ iff $x_F \succ_f y_F$ whenever $e$ and $f$ are not Savage-null and $x_F$ and $y_F$ satisfy $x_F(e) = x_F(f)$ and $y_F(e) = y_F(f)$.

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 $x_F \succ_{\{e\}} y_F$ implies $(x_{-\{f\}}, x_{-\{f\}}) \succ_{\{e\}} (y_{-\{f\}}, y_{-\{f\}})$ for all $x_F, y_F$. According to this definition, $f$ may, but need not, be Savage-null if $e \gg f$. Say that $y_F$ is **maximal w.r.t.** $\succeq_E$ if there is no $x_F$ such that $x_F \succ_E y_F$.

If $v : Z \to \mathbb{R}$ is a vNM utility function, abuse notation slightly by writing $v(x) = \sum_{z \in Z} x(z)v(z)$ whenever $x \in \Delta(Z)$ is an objective randomization. Say that $x_E$ strongly dominates $y_E$ if, $\forall f \in E$, $v(x_E(f)) > v(y_E(f))$. Say that $x_E$ weakly dominates $y_E$ if, $\forall f \in E$, $v(x_E(f)) \geq v(y_E(f))$, with strict inequality for some $e \in E$. Say that $\succeq_F$ is **admissible** on $E$ if $x_F \succeq_F y_F$ whenever $x_E$ weakly dominates $y_E$.

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 $v$ if (a) $\succeq_F$ is nontrivial and (b) $x_F \succeq_{\{f\}} y_F$ iff $v(x_F(f)) \geq v(y_F(f))$ 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 $v : Z \to \mathbb{R}$ such that $\succeq_F$ is conditionally represented by $v$.

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 $\lambda = (\mu_1, \ldots, \mu_L)$ 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 $v : Z \to \mathbb{R}$ and an LPS $\lambda = (\mu_1, \ldots, \mu_L) \in \mathbf{L}\Delta(F)$
such that $x_F \succeq_F y_F$ iff

$$
\left( \sum_{f \in F} \mu(\ell) v(x_F(f)) \right)_{\ell=1}^L \succeq_L \left( \sum_{f \in F} \mu(\ell) v(y_F(f)) \right)_{\ell=1}^L.
$$

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, $x_{F_1} \succeq_{F_1} y_{F_1}$ iff $x_F \succeq_F y_F$ whenever $x_{F_1}(f_1) = x_F(f_1, f_2)$ and $y_{F_1}(f_1) = y_F(f_1, f_2)$ for all $(f_1, f_2)$.

References


