The Economic Journal, **113** (*April*), 305–325. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

DEDUCTIVE REASONING IN EXTENSIVE GAMES*

Geir B. Asheim and Martin Dufwenberg

We justify the application to extensive games of a model of deductive reasoning based on three key features: 'caution', 'full belief of opponent rationality', and 'no extraneous restrictions on beliefs'. We apply the model to several examples, and show that it yields novel economic insights. The approach supports forward induction, without necessarily promoting backward induction.

In many economic contexts decision makers interact and take actions that extend through time. A bargaining party makes an offer, which is observed by the adversary, and accepted, rejected or followed by a counter-offer. Firms competing in markets choose prices, levels of advertisement, or investments with the intent of thereby influencing the future behaviour of competitors. One could add many examples. The standard economic model for analysing such situations is that of an *extensive game*. This paper is concerned with the following question: what happens in an extensive game if players reason deductively by trying to figure out one another's moves? We have in Asheim and Dufwenberg (2002) (AD) proposed a model for deductive reasoning leading to the concept of 'fully permissible sets', which can be applied to many strategic situations. In this paper we argue that the model is appropriate for analysing extensive games and we apply it to several such games.

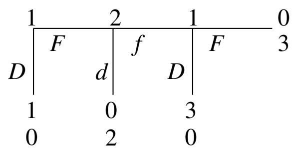
1. Motivation

There is already a literature exploring the implications of deductive reasoning in extensive games, but the answers provided differ and the issue is controversial. Much of the excitement concerns whether or not deductive reasoning implies backward induction in games where that principle is applicable. We next discuss this issue, since it provides a useful backdrop against which to introduce and motivate our own approach.

Consider the 3-stage 'Take-it-or-leave-it', introduced by Reny (1993) (a version of Rosenthal's (1981) centipede game), and shown in Figure 1 together with its pure strategy reduced strategic form (PRSF).[1] What would 2 do in Γ_1 if called upon to play? Backward induction implies that 2 would choose *d*, which is consistent with the following idea: 2 chooses *d* because she 'figures out' that 1 would choose *D* at

* During our work with this paper and its companion piece 'Admissibility and common belief', we have received helpful comments from many scholars, including Pierpaolo Battigalli and Larry Samuelson. We are very grateful to David de Meza and two referees for their constructive and insightful comments. Asheim gratefully acknowledges the hospitality of Harvard University and financial support from the Research Council of Norway.

[1] We need not consider what players plan to do at decision nodes that their own strategy precludes them from reaching (Section 3.2).



	d	f
D	1, 0	1, 0
FD	0, 2	3, 0
FF	0, 2	0, 3

Fig. 1. Γ_1 *and its PRSF*

the last node. Many models of deductive reasoning support this story, starting with Bernheim's (1984) concept of 'subgame perfect rationalisability' and Pearce's (1984) concept of 'extensive form rationalisability' (EFR). More recently, Battigalli and Siniscalchi (2002) provide a rigorous epistemic foundation for EFR, and Aumann (1995) proposes an alternative epistemic model that also implies backward induction.

However, showing that backward induction can be given some kind of underpinning does not imply that the underpinning is convincing. Indeed, scepticism concerning backward induction can be expressed by means of Γ_1. Suppose that each player believes that the opponent will play in accordance with backward induction; i.e., 1 believes that 2 chooses d if asked to play, and 2 believes that 1 plays D at his initial note. Then player 1 prefers playing D to any of his two other strategies FD and FF. Moreover, if 2 is certain that 1 believes that 2 chooses d if she were asked to play, then 2 realises that 1 has not chosen in accordance with his preferences if she after all *is* asked to play. Why then should 2 believe that 1 will make any particular choice between his two less preferred strategies, FD and FF, at his last node? So why then should 2 prefer d to f?

This kind of perspective on the 'Take-it-or-leave-it' game is much inspired by the approach proposed by Ben-Porath (1997), where similar objections against backward inductive reasoning are raised. We shall discuss his contribution in some detail, since the key features of our approach can be appreciated via a comparison to his model. Applied to Γ_1, Ben-Porath's model captures the following intuition: each player has an initial belief about the opponent's behaviour. If this belief is contradicted by the play (a 'surprise' occurs) he may subsequently entertain any belief consistent with the path of play. The only restriction imposed on updated beliefs is Bayes' rule. In Γ_1, Ben-Porath's model allows player 2 to make any choice. In particular, 2 may choose f if she initially believes with probability one that player 1 will choose D, and conditionally on D *not* being chosen assigns sufficient probability on FF. This entails that if 2 initially believes that 1 will comply with backward induction, then 2 need not follow backward induction herself.

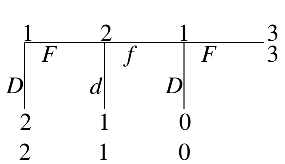
In Γ_1, our analysis captures much the same intuition as Ben-Porath's approach, and it has equal cutting power in this game. However, it yields a more structured solution as it is concerned with what strategy subsets that are deemed to be the *set* of rational choices for each player. While agreeing with Ben-Porath that deductive reasoning may lead to each of D and FD being rational for 1 and each of d and f being rational for 2, our concept of full permissibility predicts that 1's *set* of rational choices is either $\{D\}$ or $\{D, F\}$, and 2's *set* of rational choices is either $\{d\}$ or $\{d, f\}$. This has appealing features. If 2 is certain that 1's set is $\{D\}$, then – unless 2 has an assessment of the relative likelihood of 1's less preferred strategies *FD* and *FF* – one cannot conclude that 2 prefers *d* to *f* or *vice versa*; this justifies {*d*, *f*} as 2's set of rational choices. On the other hand, if 2 considers it possible that 1's set is {*D*, *FD*}, then *d* weakly dominates *f* on this set and justifies {*d*} as 2's set of rational choices. Similarly, one can justify that *D* is preferred to *FD* if and only if 1 considers it impossible that 2's set is {*d*, *f*}.

This additional structure is important for the analysis of Γ_2, illustrated in Figure 2. This game is due to Reny (1992, Fig. 1) and has appeared in many contributions. Suppose in this game that each player believes that the opponent will play in accordance with backward induction by choosing *FF* and *f* respectively. Then both players will prefer *FF* and *f* to any alternative strategy. Moreover, as will be shown in Section 4.1, our analysis implies that {*FF*} and {*f*} are the unique sets of rational choices.

Ben-Porath's approach, by contrast, does not have such cutting power in Γ_2, as it entails that deductive reasoning may lead to each of the strategies *D* and *FF* being rational for 1 and each of the strategies *d* and *f* being rational for 2. The intuition for why the strategies *D* and *d* are admitted is as follows: *D* is 1's unique best strategy if he believes with probability one that 2 plays *d*. Player 1 is justified in this belief in the sense that *d* is 2's best strategy if she initially believes with probability one that 1 will choose *D*, and if called upon to play 2 revises this belief so as to believe with sufficiently high probability (e.g., probability one) that 1 is using *FD*. This belief revision is consistent with Bayes' rule, and so is acceptable.

Ben-Porath's approach is a very important contribution to the literature, since it is a natural next step if one accepts the above critique of backward induction. Yet we shall argue below that it is *too* permissive, using Γ_2 as an illustration. The problem is that Ben-Porath does not impose certain reasonable constraints on how players reason about the likelihood of opponent choices. To see how our approach does impose such constraints, let us present the three key features of our modeling:

1. *Caution.* A player should prefer one strategy to another if the former weakly dominates the latter. Such admissibility of a player's preferences on the set of *all* opponent strategies is implicit in any procedure that starts out by eliminating all weakly dominated strategies. In extensive games, 'caution' ensures that each player takes into account any information set that is consistent with the player's own strategy.
2. *Full belief of opponent rationality.* A player should prefer one strategy to another if the former weakly dominates the latter on the set of rational



	d	*f*
D	2, 2	2 , 2
FD	1, 1	0 , 0
FF	1, 1	3 , 3

Fig. 2. Γ_2 *and its PRSF*

choices for the opponent. Such admissibility of a player's preferences on a *particular subset* of opponent strategies is essentially satisfied by procedures, like EFR and 'iterated elimination of weakly dominated strategies' (IEWDS), that promote forward induction. It is equivalent to deeming any opponent strategy that is a rational choice infinitely more likely, in the sense of Blume *et al.* (1991, Def. 5.1), than any opponent strategy not having this property.
3. *No extraneous restrictions on beliefs.* A player should prefer one strategy to another *only if* the former weakly dominates the latter on the set of all opponent strategies or on the set of rational choices for the opponent. Such equal treatment of different opponent strategies that are all rational – or all irrational – have in principle been argued by Samuelson (1992, p. 311), Gul (1997), and Mariotti (1997).

A player's preferences over his own strategies, which depend on his payoff function and his beliefs about the opponent's choice, leads to a *choice set* (i.e., a set of maximal strategies). A player's preferences is said to be *fully admissibly consistent* with the preferences of his opponent if one strategy is preferred to another if and only if the former weakly dominates the latter

- on the union of opponent choice sets that are deemed possible, or
- on the set of all opponent strategies,

thereby satisfying the three key features presented above. A subset of strategies is a *fully permissible set* if and only if it can be a choice set when there is common certain belief of full admissible consistency, where an event is 'certainly believed' if the complement is deemed impossible (or more precisely: is Savage-null). Thus, this solution concept determines a collection of strategy subsets – a family of choice sets – for each player.

We can use Γ_2 of Figure 2 to illustrate the three key features. Assume that 1 deems d infinitely more likely than f, while 2 deems D infinitely more likely than FD and FD infinitely more likely than FF. Then the players rank their strategies as follows:

$$\begin{array}{ll} \text{1's preferences:} & D \succ FF \succ FD \\ \text{2's preferences:} & d \succ f. \end{array}$$

This is in fact precisely the justification of the strategies D and d given above when applying Ben-Porath's approach to Γ_2. Here, 'caution' is satisfied since all opponent strategies are taken into account; in particular, FF is preferred to FD as the former strategy weakly dominates the latter. Moreover, 'full belief of opponent rationality' is satisfied since each player deems the opponent's maximal strategy infinitely more likely that any non-maximal strategy. However, the requirement of 'no extraneous restrictions on beliefs' is not satisfied since the preferences of 2 introduce extraneous restrictions on beliefs by deeming one of 1's non-maximal strategies, FD, infinitely more likely than another non-maximal strategy, FF. When we return to G_2 in Section 4.1, we show how the additional imposition of 'no extraneous restrictions on beliefs' means that deductive reasoning leads to the conclusion that $\{FF\}$ and $\{f\}$ are the players' choice sets in this game.

In Section 2, the formal definition of the concept of fully permissible sets is reproduced from AD. The definition applies to games represented in strategic form. However, in Section 3 we prove results, drawing on work by Mailath *et al.* (1993), that justify the claim that interesting implications of deductive reasoning in a given extensive game can be derived by applying the concept to the strategic form of that game.

Sections 4 and 5 are concerned with such applications, with the aim of showing how our solution concept gives new and economically relevant insights into the implications of deductive reasoning in extensive games. The material is organised around two central themes: *backward* and *forward induction.* Other support for forward induction through the concept of EFR and the procedure of IEWDS precludes outcomes in conflict with backward induction (Battigalli, 1997). In contrast, we will show how the concept of fully permissible sets promotes forward induction in the 'Battle-of-the-Sexes-with-an-Outside-Option' and 'Burning Money' games as well as an economic application from organisation theory, while not insisting on the backward induction outcome in games (like Γ_1 and the 3-period prisoners' dilemma) where earlier contributions, like Basu (1990), Reny (1993) and others, have argued on theoretical grounds that this is problematic. Still, we will show that the backward induction outcome *is* obtained in Γ_2, and that our concept has considerable bite in the 3-period prisoners' dilemma game.

Lastly, in Section 6 we compare our approach to related work.

2. Concepts

Below we make a self-contained presentation of the concept of fully permissible sets, reproduced from AD. This definition is given in terms of an elimination procedure – iterative elimination of choice set under full admissible consistency (IECFA) – that iteratively eliminates strategy subsets that cannot be choice sets under full admissible consistency. In AD we also provide an epistemic characterisation of this procedure (as indicated in Section 1). Readers that are interested in details are referred to AD. In the present paper we will only apply the elimination procedure, although we will at some places interpret results in a way consistent with the underlying epistemic foundation. As the purpose here is to apply IECFA to extensive games, we start by introducing such games. We refer to standard texts for the general formalism of extensive games and state only those basic and derived notions that will be needed.

2.1. *Extensive Games*

A finite extensive game Γ (without nature) includes a set of players $I \in \{1, 2\}$ (we assume 2 players for convenience), a set of terminal nodes Z, and, for each player i, a vNM utility function v_i: $Z \to \Re$ that assigns payoff to any outcome. For any player i, there is a finite collection of information sets H_i, with a finite set of actions $A(h)$ being associated with each $h \in H_i$. A pure strategy for player i is a function s_i that to any $h \in H_i$ assigns an action in $A(h)$. Let S_i denote player i's finite set of pure strategies, and let $S = S_1 \times S_2$. Write p_i, r_i and $s_i (\in S_i)$ for pure strategies and x_i and

y_i ($\in \Delta(S_i)$) for mixed strategies. Define u_i: $S \rightarrow \Re$ by $u_i(s) = v_i(z)$, where z is the terminal node reached when $s = (s_1, s_2)$ is used, and refer to $G = (S_1, S_2, u_1, u_2)$ as the strategic form of the extensive game Γ. Since u_i is a vNM utility function, we may extend u_i to mixed strategies: $u_i(x_i, s_j) = \sum_{s_i \in S_i} x_i(s_i) u_i(s_i, s_j)$. For any $h \in \cup_{i \in I} H_i$, let S^h denote the set of strategy vectors for which h is reached. As Γ is a 2-player game with perfect recall, S^h is rectangular: $S^h = S_1^h \times S_2^h$.

2.2. *Fully Permissible Sets*

Each player i has preferences over his own strategies, which depend on u_i and his beliefs about the strategy choice of his opponent. Player i's *choice set* is the set of pure strategies that are maximal w.r.t. i's preferences over his own strategies: s_i ($\in S_i$) is in i's choice set if and only if there is no x_i ($\in \Delta(S_i)$) such that x_i is preferred to s_i. For the class of preferences considered here, i's choice set is non-empty and supports any maximal mixed strategy (subsection 4.3 of AD).

Let S_j' ($\subseteq S_j$) be a set of opponent strategies. Say that *x_i weakly dominates y_i on S_j'* if, $\forall s_j \in S_j'$, $u_i(x_i, s_j) \geq u_i(y_i, s_j)$, with strict inequality for some $s_j \in S_j'$. Interpret Q_j ($\subseteq S_j$) as the set of strategies that player i deems to be the set of rational choices for his opponent. Assume that player i's preferences over his own strategies are given by: x_i is preferred to y_i if and only if x_i weakly dominates y_i on Q_j or S_j. Player i's choice set, $C_i(Q_j)$, is then equal to $S_i \backslash D_i(Q_j)$, where, for any ($\emptyset \neq$) $Q_j \subseteq S_j$,

$$D_i(Q_j) := \{s_i \in S_i | \exists x_i \in \Delta(S_i) \text{ s.t. } x_i \text{ weakly dom. } \mathrm{S}_i \text{ on } Q_j \text{ or } S_j\}.$$

Let $\Sigma = \Sigma_1 \times \Sigma_2$, where $\Sigma_i := 2^{S_i} \backslash \{\emptyset\}$ denotes the collection of non-empty subsets of S_i. Write π_i, ρ_i, and σ_i ($\in \Sigma_i$) for subsets of pure strategies. For any ($\emptyset \neq$) $\Xi = \Xi_1 \times \Xi_2 \subseteq \Sigma$, write $\alpha(\Xi) := \alpha_1(\Xi_2) \times \alpha_2(\Xi_1)$, where

$$\alpha_i(\Xi_j) := \{\pi_i \in \Sigma_i | \exists (\emptyset \neq) \Psi_j \subseteq \Xi_j \text{ s.t. } \pi_i = C_i(\cup_{\sigma_j \in \Psi_j} \sigma_j)\}.$$

Hence, $\alpha_i(\Xi_j)$ is the collection of strategy subsets that can be choice sets for player i if i's preferences are characterised by the following property: one strategy is preferred to another if and only if the one weakly dominates the other *either* on the union of the strategy subsets in a non-empty subcollection of Ξ_j *or* on the set of all opponent strategies.

We can now define the concept of a fully permissible set.

DEFINITION 1. *Consider the sequence defined by* $\Xi(0) = \Sigma$ *and* $\forall g \geq 1$, $\Xi(g) = \alpha[\Xi(g-1)]$. *A non-empty strategy set* π_i *is said to be a fully permissible set for i if* $\pi_i \in \bigcap_{g=0}^{\infty} \Xi_i(g)$.

Let $\Pi = \Pi_1 \times \Pi_2$ denote the *collection* of vectors of fully permissible sets. Since the game is finite, $\Xi(g)$ converges to Π in a finite number of iterations. IECFA is the procedure that in round g eliminates sets in $\Xi(g-1) \backslash \Xi(g)$ as possible choice sets. A choice set of player i survives elimination round g if and only if it is a choice set w.r.t. preferences that are characterised by the following property: one strategy is preferred to another if and only if the one weakly dominates the other *either* on the union of some (or all) of opponent choice sets that have survived the procedure up till round $g - 1$ *or* on the set of all opponent strategies. A fully permissible set is a choice set which will survive in this way for any g. It follows from the analysis of AD that strategy subsets that this algorithm has not eliminated by round g can be interpreted as choice sets that are compatible with $g - 1$ order of mutual certain belief of full admissible consistency.

The algorithm of Definition 1 – IECFA – is an elimination procedure, and in this regard it is reminiscent of procedures that iteratively eliminates dominated strategies. However, IECFA does *not* eliminate strategies. Rather, it eliminates *sets* of strategies that cannot be choice sets under full admissible consistency. It is therefore that IECFA starts with each player's collection of all non-empty strategy subsets, and then iteratively eliminates subsets in this collection. It is important that the appropriate interpretation of IECFA in terms of surviving choice sets be borne in mind.

We reproduce from AD the following proposition, which characterises the strategy subsets that survive IECFA and thus are fully permissible.

PROPOSITION 1. *(i)* $\forall i \in I,\ \Pi_i \neq \emptyset$. *(ii)* $\Pi = \alpha(\Pi)$. *(iii)* $\forall i \in I$, $\pi_i \in \Pi_i$ *if and only if there exists* $\Xi = \Xi_1 \times \Xi_2$ *with* $\pi_i \in \Xi_i$ *such that* $\Xi \subseteq \alpha(\Xi)$.

Proposition 1(i) establishes existence, but not uniqueness, of each player's fully permissible set(s). Games with multiple strict Nash equilibria illustrate the possibility of such multiplicity; by Proposition 1(iii), any strict Nash equilibrium corresponds to a vector of fully permissible sets. Other (quite different) examples of games with multiple fully permissible sets are provided in Sections 4 and 5, namely Γ_1, a 3-period prisoners' dilemma (Γ_3), and Γ_5. Proposition 1(ii) means that Π is a fixed point in terms of a collection of vectors of strategy sets. By Proposition 1(iii) it is the largest such fixed point.

3. Justifying Extensive Form Application

The concept of fully permissible sets, presented in Section 2 of the present paper and epistemically characterised in AD, is designed to analyse the implications of deductive reasoning in strategic form games. In this paper, we propose that this concept can be fruitfully applied for analysing any extensive game through its strategic form. In fact, we propose that it is legitimate to confine attention to the game's pure strategy reduced strategic form (Definition 2), which is computationally more convenient. In this Section we prove two results which, taken together, justify such applications.

3.1. *Dynamic Consistency*

Proposition 2 addresses the dynamic consistency problem inherent in applying AD's strategic form analysis to games with an explicit sequential structure. Consider any strategy that is maximal given preferences that satisfy that one strategy is preferred to another if and only if the one weakly dominates the other on Q_j – the set of strategies that player i deems to be the set of rational choices for his opponent – or S_j – the set of all opponent strategies. Hence, the strategy is maximal at the outset of a corresponding extensive game. We prove that this strategy is still maximal when the preferences have been updated upon reaching any information set that the choice of this strategy does not preclude.

Assume that player i's preferences over his own strategies are given by: x_i is preferred to y_i if and only if x_i weakly dominates y_i on Q_j or S_j. Let, for any $h \in H_i$, $Q_j^h := Q_j \cap S_j^h$ denote the set of strategies in Q_j that are consistent with the information set h being reached. If x_i, y_i $(\in \Delta(S_i^h))$, then i's preferences conditional on the information set $h \in H_i$ being reached is given by: x_i is preferred to y_i if and only if x_i weakly dominates y_i on Q_j^h or S_j^h (where it follows from the definition that weak dominance on Q_j^h is not possible if $Q_j^h = \emptyset$). Furthermore, i's choice set conditional on $h \in H_i$, $C_i^h(Q_j)$, is equal to $S_i^h \backslash D_i^h(Q_j)$, where, for any $(\emptyset \neq)$ $Q_j \subseteq S_j$,

$$D_i^h(Q_j) := \{s_i \in S_i^h | \exists x_i \in \Delta(S_i^h) \text{ s.t.} x_i \text{ weakly dom. } s_i \text{ on } Q_j^h \text{ or } S_j^h\}.$$

By the result below, if s_i is maximal at the outset of an extensive game, then it is also maximal at later information sets for i that s_i does not preclude.

PROPOSITION 2. *Let* $(\emptyset \neq)$ $Q_j \subseteq S_j$. *If* $s_i \in C_i(Q_j)$, *then* $s_i \in C_i^h(Q_j)$ *for any* $h \in H_i$ *with* $S_i^h \ni s_i$.

Proof. Suppose that $s_i \in S_i^h \backslash C_i^h(Q_j) = D_i^h(Q_j)$. Then there exists $x_i \in \Delta(S_i^h)$ such that x_i weakly dominates s_i on Q_j^h or S_j^h. By Mailath *et al.* (1993, Defs. 2 and 3 and the if-part of Theorem 1), S^h is a *strategically independent set* for i. Hence, x_i can be chosen such that $u_i(x_i, s_j) = u_i(s_i, s_j)$ for all $s_j \in S_j \backslash S_j^h$. This implies that x_i weakly dominates s_i on Q_j or S_j, implying that $s_i \in D_i(Q_j) = S_i \backslash C_i(Q_j)$.

By the assumption of 'caution', each player i takes into account the possibility of reaching any information set for i that the player's own strategy does not preclude from being reached.

3.2. *Reduced Strategic Form*

It follows from Proposition 3 that it is in fact sufficient to consider the pure strategy reduced strategic form when deriving the fully permissible sets of the game. The following definition is needed.

DEFINITION 2. *Let* r_i, $s_i \in S_i$. *Then* r_i *and* s_i *are equivalent if, for each* $k \in I$, $u_k(r_i, s_j) = u_k(s_i, s_j)$ *for all* $s_j \in S_j$. *The pure strategy reduced strategic form* (*PRSF*) *of G is obtained by letting, for each i, each class of equivalent pure strategies be represented by exactly one pure strategy.*

Since the maximality of one of two equivalent strategies implies that the other is maximal as well, the following observation holds: if r_i and s_i are equivalent and π_i is a fully permissible set for i, then $r_i \in \pi_i$ if and only if $s_i \in \pi_i$. To see this formally, note that if $r_i \in \pi_i$ for some fully permissible set π_i, then, by Proposition 1(ii), there exists $(\emptyset \neq)\Psi_j \subseteq \Pi_j$ such that $r_i \in \pi_i = C_i(\cup_{\sigma_j \in \Psi_j} \sigma_j)$. Since r_i and s_i are equivalent,

$s_i \in C_i(\cup_{\sigma_j \in \Psi_j} \sigma_j) = \pi_i$. This observation explains why the following result can be established.

PROPOSITION 3. *Let* $\tilde{G} = (\tilde{S}_k, \tilde{u}_k)_{k \in I}$ *be a strategic form game where* $r_i, s_i \in \tilde{S}_i$ *are two equivalent strategies for i. Consider* $G = (S_k, u_k)_{k \in I}$, *where* $S_i = \tilde{S}_i \backslash \{r_i\}$ *and* $S_j = \tilde{S}_j$ *for* $j \neq i$, *and where, for all* $k \in I$, u_k *is the restriction of* $\tilde{u}_k$ *to* $S = S_1 \times S_2$. *Let, for each* $k \in I$, $\Pi_k(\tilde{\Pi}_k)$ *denote the collection of fully permissible sets for k in G* $(\tilde{G})$. *Then* Π_i *is obtained from* $\tilde{\Pi}_i$ *by removing* r_i *from any* $\tilde{\pi}_i \in \tilde{\Pi}_i$ *with* $s_i \in \tilde{\pi}_i$, *while, for* $j \neq i$, $\Pi_j = \tilde{\Pi}_j$.

Proof. By Proposition 1(iii) it suffices to show that

1. If $\tilde{\Xi} \subseteq \alpha(\tilde{\Xi})$ for $\tilde{G}$, then $\Xi \subseteq \alpha(\Xi)$ for G, where Ξ_i is obtained from $\tilde{\Xi}_i$ by removing r_i from any $\tilde{\pi}_i \in \tilde{\Xi}_i$ with $s_i \in \tilde{\pi}_i$, while, for $j \neq i$, $\Xi_j = \tilde{\Xi}_j$.
2. If $\Xi \subseteq \alpha(\Xi)$ for G, then $\tilde{\Xi} \subseteq \alpha(\tilde{\Xi})$ for $\tilde{G}$, where $\tilde{\Xi}_i$ is obtained from Ξ_i by adding r_i to any $\pi_i \in \Xi_i$ with $s_i \in \pi_i$, while, for $j \neq i$, $\tilde{\Xi}_j = \Xi_j$.

Part 1. Assume $\tilde{\Xi} \subseteq \alpha(\tilde{\Xi})$. By the observation preceding Proposition 3, if $\tilde{\pi}_i \in \tilde{\Pi}_i$, then $r_i \in \tilde{\pi}_i$ if and only if $s_i \in \tilde{\pi}_i$. Pick any $k \in I$ and any $\tilde{\pi}_k \in \tilde{\Pi}_k$. Let ℓ denote the other player. By the definition of $\alpha_k(\cdot)$, there exists $(\emptyset \neq)\tilde{\Psi}_\ell \subseteq \tilde{\Pi}_\ell$ such that $\tilde{\pi}_k = C_k(\cup_{\tilde{\sigma}_\ell \in \tilde{\Psi}_\ell} \tilde{\sigma}_\ell)$. Construct Ψ_i by removing r_i from any $\tilde{\sigma}_i \in \tilde{\Psi}_i$ with $s_i \in \tilde{\sigma}_i$ and replace $\tilde{S}_i$ by S_i, while, for $j \neq i$, $\Psi_j = \tilde{\Psi}_j$ and $S_j = \tilde{S}_j$. Then it follows from the definition of $C_k(\cdot)$ that $C_k(\cup_{\sigma_\ell \in \Psi_\ell} \sigma_\ell) = \tilde{\pi}_k \backslash \{r_k\}$ if $k = i$ and $C_k(\cup_{\sigma_\ell \in \Psi_\ell} \sigma_\ell) = \tilde{\pi}_k$ if $k \neq i$. Since, $\forall k \in I$, $(\emptyset \neq)\Psi_k \subseteq \Xi_k$, we have that $\Xi \subseteq \alpha(\Xi)$. *Part 2* is shown similarly.

Proposition 3 means that the PRSF is sufficient for analysing common certain belief of full admissible consistency, which is the epistemic foundation for the concept of fully permissible sets. Consequently, in the strategic form of an extensive game, it is unnecessary to specify actions at information sets that a strategy precludes from being reached. Hence, instead of fully specified strategies, it is sufficient to consider what Rubinstein (1991) calls *plans of action.* For a generic extensive game, the set of plans of action is identical to the strategy set in the PRSF.

In the following two sections we apply the concept of fully permissible sets to extensive games. We organise the discussion around two themes: backward and forward induction. Motivated by Propositions 2 and 3, we analyse each extensive game via its PRSF (Definition 2), given in conjunction to the extensive form. In each example, each plan of action that appears in the underlying extensive game corresponds to a distinct strategy in the PRSF.

4. Backward Induction

Does deductive reasoning in extensive games imply backward induction? In this Section we show that the answer provided by the concept of fully permissible sets is 'sometimes, but not always'.

4.1. *'Sometimes'*

There are many games where Ben-Porath's approach does not capture backward induction while our approach does (and the converse is not true). Ben-Porath

(1997) assumes 'initial common certainty of rationality' (initial CCR) in extensive games with perfect information. He proves that in generic games (with no payoff ties at terminal nodes for any player) the outcomes consistent with that assumption coincide with those that survive the Dekel and Fudenberg (1990) procedure (where one round of elimination of all weakly dominated strategies is followed by iterated elimination of strongly dominated strategies).

It is a general result that the concept of fully permissible sets refines the Dekel and Fudenberg procedure (see AD, Proposition 3.2). Game Γ_2 of Figure 2 shows that the refinement may be strict even for generic extensive games with perfect information, and indeed that fully permissible sets may respect backward induction where Ben-Porath's solution does not. The strategies surviving the Dekel–Fudenberg procedure, and thus consistent with initial CCR, are D and FF for player 1 and d and f for player 2. In Section 1 we gave an intuition for why the strategies D and d are possible. This is, however, at odds with the implications of common certain belief of full admissible consistency. Applying IECFA to the PRSF of Γ_2 yields:

$$\begin{aligned}
\Xi(0) &= \Sigma_1 \times \Sigma_2 \\
\Xi(1) &= \{\{D\},\{FF\},\{D,FF\}\} \times \Sigma_2 \\
\Xi(2) &= \{\{D\},\{FF\},\{D,FF\}\} \times \{\{f\},\{d,f\}\} \\
\Xi(3) &= \{\{FF\},\{D,FF\}\} \times \{\{f\},\{d,f\}\} \\
\Xi(4) &= \{\{FF\},\{D,FF\}\} \times \{\{f\}\} \\
\Pi = \Xi(5) &= \{\{FF\}\} \times \{\{f\}\}.
\end{aligned}$$

Interpretation: $\Xi(1)$: 'Caution' implies that FD cannot be a maximal strategy (i.e., an element of a choice set) for 1 since it is weakly dominated (in fact, even strongly dominated). $\Xi(2)$: Player 2 certainly believes that only $\{D\}$, $\{FF\}$ and $\{D, FF\}$ are candidates for 1's choice set. By 'full belief of opponent rationality' and 'no extraneous restrictions on beliefs' this excludes $\{d\}$ as 2's choice set, since d weakly dominates f only on $\{FD\}$ or $\{D, FD\}$. $\Xi(3)$: 1 certainly believes that only $\{f\}$ and $\{d, f\}$ are candidates for 2's choice set. By 'full belief of opponent rationality' and 'no extraneous restrictions on beliefs' this excludes $\{D\}$ as 1's choice set, since D weakly dominates FD and FF only on $\{d\}$. $\Xi(4)$: Player 2 certainly believes that only $\{FF\}$ and $\{D, FF\}$ are candidates for 1's choice set. By 'full belief of opponent rationality' this implies that 2's choice set is $\{f\}$ since f weakly dominates d on both $\{FF\}$ and $\{D,FF\}$. $\Xi(5)$: 1 certainly believes that 2's choice set is $\{f\}$. By 'full belief of opponent rationality' this implies that $\{FF\}$ is 1's choice set since FF weakly dominates D on $\{f\}$. No further elimination of choice sets is possible, so $\{FF\}$ and $\{f\}$ are the respective players' unique fully permissible sets.

4.2. *'Not Always'*

While fully permissible sets capture backward induction in Γ_2 and other games, the concept does not capture backward induction in certain games where the procedure has been considered controversial.[2] Much of the controversy centres on the following paradoxical aspect: Why should a player believe that an opponent's future play will satisfy backward induction if the opponent's previous play is incompatible with backward induction? A prototypical game for casting doubt on backward induction is the take-it-or-leave-it game Γ_1, which we next analyse in detail.

Applying our algorithm IECFA to the PRSF of Γ_1 of Figure 1 yields:

$$\begin{aligned}
\Xi(0) &= \Sigma_1 \times \Sigma_2 \\
\Xi(1) &= \{\{D\}, \{FD\}, \{D, FD\}\} \times \Sigma_2 \\
\Xi(2) &= \{\{D\}, \{FD\}, \{D, FD\}\} \times \{\{d\}, \{d, f\}\} \\
\Pi = \Xi(3) &= \{\{D\}, \{D, FD\}\} \times \{\{d\}, \{d, f\}\}.
\end{aligned}$$

Interpretation: $\Xi(1)$: *FF* cannot be a maximal strategy for 1 since it is strongly dominated. $\Xi(2)$: Player 2 certainly believes that only {*D*}, {*FD*} and {*D*, *FD*} are candidates for 1's choice set. This excludes $\{f\}$ as 2's choice set since $\{f\}$ is 2's choice set only if 2 deems {*FF*} or {*D*, *FF*} possible. $\Xi(3)$: 1 certainly believes that only $\{d\}$ and $\{d, f\}$ are candidates for 2's choice set, implying that {*FD*} cannot be 1's choice set. No further elimination of choice sets is possible and the collection of vectors of fully permissible sets is as specified.

Note that backward induction is *not* implied. To illustrate why, we focus on player 2 and explain why $\{d, f\}$ may be a choice set for her. Player 2 certainly believes that 1's choice set is {*D*} or {*D*, *FD*}. This leaves room for two basic cases. First, suppose 2 deems {*D*, *FD*} possible. Then $\{d\}$ must be her choice set, since she must consider it infinitely more likely that 1 uses *FD* than that he uses *FF*. Second, and more interestingly, suppose 2 does not deem {*D*, *FD*} possible. Then conditional on 2's node being reached 2 certainly believes that 1 is not choosing a maximal strategy. As player 2 does not assess the relative likelihood of strategies that are not maximal (the requirement of 'no extraneous restrictions on beliefs'), $\{d, f\}$ is her choice set in this case. Note that even in the case when 2 deems {*D*} to be the only possible choice set for 1, she still considers it possible that 1 may choose one of his non-maximal strategies *FD* and *FF* (the requirement of 'caution'), although each of these strategies is in this case deemed infinitely less likely than the unique maximal strategy *D*.

Applied to (the PRSF of) Γ_1, our concept permits two fully permissible sets for each player. How can this multiplicity of fully permissible sets be interpreted? The following interpretation corresponds to the underlying formalism: the concept of fully permissible sets, when applied to Γ_1, allows for two different 'types' of each player. Consider player 2. Either she may consider that {*D*, *FD*} is a possible choice set for 1, in which case her choice set will be $\{d\}$ so that she complies with backward induction. Or she may consider {*D*} to be the only possible choice set for 1, in

[2] Many authors have debated whether backward induction is a problematic procedure. A classic reference is Selten (1978) and later contributions include Aumann (1995), Basu (1990), Ben-Porath (1997), Bicchieri (1989), Binmore (1987), Bonanno (1991), Dufwenberg and Lindén (1996), Gul (1997), Reny (1993), Rosenthal (1981) and Stalnaker (1998).

which case 2's choice set is $\{d, f\}$. Intuitively, if 2 is certain that 1 is a backward inducter, then 2 need not be a backward inducter herself! In this game, our model captures an intuition that is very similar to that of Ben-Porath's model.

Reny (1993) defines a class of 'belief consistent' games, and argues on epistemic grounds that backward induction is problematic only for games that are not in this class. It is interesting to note that the game (Γ_2) where our concept of fully permissible sets differs from Ben-Porath's analysis by promoting backward induction, is belief-consistent. In contrast, the game (Γ_1) where the present concept coincides with his by *not* yielding backward induction, is *not* belief-consistent. There are examples of games that are *not* belief consistent, where full permissibility still implies backward induction, meaning that belief consistency is not necessary for this conclusion. It is, however, an as-yet unproven conjecture that belief consistency is sufficient for the concept of fully permissible sets to promote backward induction.

We now compare our results to the very different findings of Aumann (1995) (Stalnaker, 1998, Sect. 5). In Aumann's model, where it is crucial to specify full strategies (rather than plans of actions), common knowledge of rational choice implies in Γ_1 that all strategies for 1 but *DD* (where he takes a payoff of 1 at his first node and a payoff of 3 at his last node) are impossible. Hence, it is impossible for 1 to play *FD* or *FF* and thereby ask 2 to play. However, in the counterfactual event that 2 is asked to play, she optimises as if player 1 at his last node follows his only possible strategy *DD*, implying that it is impossible for 2 to choose f (Aumann's Sections 4b, 5b, and 5c). Thus, in Aumann's analysis, if there is common knowledge of rational choice, then each player chooses the backwards induction strategy. By contrast, in our analysis player 2 being asked to play is seen to be incompatible with 1 playing *DD* or *DF*. For the determination of 2's preference over her strategies it is the relative likelihood of *FD* versus *FF* that is important to her. As seen above, this assessment depends on whether she deems $\{D, FD\}$ as a possible candidate for 1's choice set.

4.3. *Prisoners' Dilemma*

We close this section by considering a finitely repeated prisoners' dilemma game. Such a game does not have perfect information, but it can still be solved by backward induction to find the unique subgame perfect equilibrium (no one cooperates in the last period, given this no one cooperates in the penultimate period etc.). This solution has been taken to be counterintuitive (Pettit and Sugden, 1989). We consider the case of a 3-period prisoners' dilemma game (Γ_3) and show that, again, the concept of fully permissible sets does not capture backward induction. However, the fully permissible sets nevertheless have considerable cutting power. Our solution refines the Dekel–Fudenberg procedure and generates some special 'structure' on the choice sets that survive.

The payoffs of the stage game are given as follows, using Aumann's (1987, pp. 468–9) description: each player decides whether he will receive 1 (*defect*) or the other will receive 3 (*cooperate*). There is no discounting. Hence, the action *defect* is strongly dominant in the stage game, but still, each player is willing to *cooperate* in one stage if this induces the other player to *cooperate* instead of *defect* in the next stage. It follows from Proposition 3 that we need only consider, what Rubinstein (1991) calls, plans of action.

There are six plans of actions for each player that survive the Dekel and Fudenberg procedure. In any of these, a player always defects in the 3rd stage, and does not always cooperate in the 2nd stage. The six plans of actions for each player i are denoted s_i^{NT}, s_i^{NV}, s_i^{NE}, s_i^{RT}, s_i^{RV} and s_i^{RE}, where N denotes that i is *nice* in the sense of *cooperating* in the 1st stage, where R denotes that i is *rude* in the sense of *defecting* in the 1st stage, where T denotes that i plays *tit-for-tat* in the sense of *cooperating* in the 2nd stage if and only $j \neq i$ has *cooperated* in the 1st stage, where V denotes that i plays *inverse tit-for-tat* in the sense of *defecting* in the 2nd stage if and only if $j \neq i$ has *cooperated* in the 1st stage, and where E denotes that i is *exploitive* in the sense of *defecting* in the 2nd stage independently of what $j \neq i$ has played in the 1st stage. The strategic form after elimination of all other plans of actions is given in Figure 3. Note that none of these plans of actions are weakly dominated in the full strategic form.

AD (Proposition 3.2) show that any fully permissible set is a subset of the set of strategies surviving the Dekel and Fudenberg procedure. Hence, only subsets of $\{s_i^{NT}, s_i^{NV}, s_i^{NE}, s_i^{RT}, s_i^{RV}, s_i^{RE}\}$ can be i's choice set under common certain belief of full admissible consistency. Furthermore, under common certain belief of full admissible consistency, we have for each player i that

- any choice set that contains s_i^{NT} must also contain s_i^{NE}, since s_i^{NT} is a maximal strategy only if s_i^{NE} is a maximal strategy,
- any choice set that contains s_i^{NV} must also contain s_i^{NE}, since s_i^{NV} is a maximal strategy only if s_i^{NE} is a maximal strategy,
- any choice set that contains s_i^{RT} must also contain s_i^{RE}, since s_i^{RT} is a maximal strategy only if s_i^{RE} is a maximal strategy,
- any choice set that contains s_i^{RV} must also contain s_i^{RE}, since s_i^{RV} is a maximal strategy only if s_i^{RE} is a maximal strategy,

Given that the choice set of the opponent satisfies these conditions, this implies that

- if s_i^{NE} is included in i's choice set, only the following sets are candidates for i's choice set: $\{s_i^{NT}, s_i^{NE}, s_i^{RT}, s_i^{RE}\}$, $\{s_i^{NV}, s_i^{NE}, s_i^{RV}, s_i^{RE}\}$, or $\{s_i^{NE}, s_i^{RE}\}$. The reason is that s_i^{NE} is a maximal strategy only if i considers it possible that j's choice set contains s_j^{NT} (and hence, s_j^{NE}) or s_j^{RT} (and hence, s_j^{RE}).

	s_2^{NT}	s_2^{NV}	s_2^{NE}	s_2^{RT}	s_2^{RV}	s_2^{RE}
s_1^{NT}	7, 7	4, 8	4, 8	5, 5	2, 6	2, 6
s_1^{NV}	8, 4	5, 5	5, 5	4, 8	1, 9	1, 9
s_1^{NE}	8, 4	5, 5	5, 5	5, 5	2, 6	2, 6
s_1^{RT}	5, 5	8, 4	5, 5	3, 3	6, 2	3, 3
s_1^{RV}	6, 2	9, 1	6, 2	2, 6	5, 5	2, 6
s_1^{RE}	6, 2	9, 1	6, 2	3, 3	6, 2	3, 3

Fig. 3. *Reduced Form of the 3-period PD game* (Γ_3)

- if s_i^{RE}, but not s_i^{NE}, is included in i's choice set, only the following sets are candidates for i's choice set: $\{s_i^{RT}, s_i^{RE}\}$, $\{s_i^{RV}, s_i^{RE}\}$, or $\{s_i^{RE}\}$. The reason is that s_i^{RE} is a maximal strategy only if i considers it possible that j's choice set contains s_j^{NV}, s_j^{NE}, s_j^{RV}, or s_j^{RE}.

This in turn implies that

- i's choice set does not contain s_i^{NV} or s_i^{RV} since any candidate for j's choice set contains s_j^{RE}, implying that s_i^{NE} is preferred to s_i^{NV} and s_i^{RE} is preferred to s_i^{RV}.

Hence, the only candidates for i's choice set under common certain belief of full admissible consistency are $\{s_i^{NT}, s_i^{NE}, s_i^{RT}, s_i^{RE}\}$, $\{s_i^{NE}, s_i^{RE}\}$, $\{s_i^{RT}, s_i^{RE}\}$, and $\{s_i^{RE}\}$. Moreover, it follows from Proposition 1(iii) that all these sets are indeed fully permissible since

- $\{s_i^{NT}, s_i^{NE}, s_i^{RT}, s_i^{RE}\}$ is i's choice set if he deems $\{s_j^{RT}, s_j^{RE}\}$, but not $\{s_j^{NE}, s_j^{RE}\}$ and $\{s_j^{NT}, s_j^{NE}, s_j^{RT}, s_j^{RE}\}$, as possible candidates for j's choice set,
- $\{s_i^{NE}, s_i^{RE}\}$ is i's choice set if he deems $\{s_j^{NT}, s_j^{NE}, s_j^{RT}, s_j^{RE}\}$ as a possible candidate for j's choice set,
- $\{s_i^{RT}, s_i^{RE}\}$ is i's choice set if he deems $\{s_j^{RE}\}$ as the only possible candidate for j's choice set,
- $\{s_i^{RE}\}$ is i's choice set if he deems $\{s_j^{NE}, s_j^{RE}\}$, but not $\{s_j^{RT}, s_j^{RE}\}$ and $\{s_j^{NT}, s_j^{NE}, s_j^{RT}, s_j^{RE}\}$, as possible candidates for j's choice set.

While play in accordance with strategies surviving the Dekel and Fudenberg procedure does not provide any prediction other than both players *defecting* in the 3rd stage, the concept of fully permissible sets has more bite. In particular, a player *cooperates* in the 2nd stage only if the opponent has *cooperated* in the 1st stage. This implies that only the following paths can be realised if players choose strategies in fully permissible sets:

((*cooperate*, *cooperate*), (*cooperate*, *cooperate*), (*defect*, *defect*))
((*cooperate*, *cooperate*), (*cooperate*, *defect*), (*defect*, *defect*)) and v.v.
((*cooperate*, *defect*), (*defect*, *cooperate*), (*defect*, *defect*)) and v.v.
((*cooperate*, *cooperate*), (*defect*, *defect*), (*defect*, *defect*))
((*cooperate*, *defect*), (*defect*, *defect*), (*defect*, *defect*)) and v.v.
((*defect*, *defect*), (*defect*, *defect*), (*defect*, *defect*)).

That the path ((*cooperate, defect*), (*cooperate, defect*), (*defect, defect*)) or v.v. cannot be realised if players choose strategies in fully permissible sets can be interpreted as an indication that the present analysis seems to produce some element of reciprocity in the 3-period prisoners' dilemma game.

5. Forward Induction

AD analyse the 'Battle-of-the-Sexes-with-an-Outside-Option' (BoSwOO) game (introduced by Kreps and Wilson (1982) who credit Elon Kohlberg) and the 'Burning money' game (van Damme 1989, Figure 5; Ben-Porath and Dekel 1992, Figure 1.2) in the strategic form and show how the concept of fully permissible sets yields forward induction outcomes. Here we consider extensive form versions of these games, then analyse a modification of BoSwOO due to Dekel and Fudenberg (1990), and finally consider an economic application from organisation theory.

5.1. *Battle-of-the-Sexes With Variations*

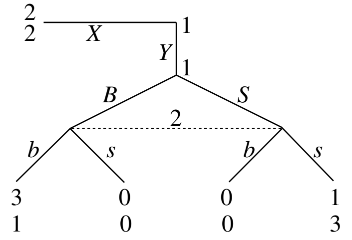
We start with BoSwOO (cf. Γ_4 of Figure 4). Applying IECFA to the PRSF of Γ_4 yields:

$$\begin{aligned}
\Xi(0) &= \Sigma_1 \times \Sigma_2 \\
\Xi(1) &= \{\{X\},\{B\},\{X,B\}\} \times \Sigma_2 \\
\Xi(2) &= \{\{X\},\{B\},\{X,B\}\} \times \{\{b\},\{b,s\}\} \\
\Xi(3) &= \{\{B\},\{X,B\}\} \times \{\{b\},\{b,s\}\} \\
\Xi(4) &= \{\{B\},\{X,B\}\} \times \{\{b\}\} \\
\Pi = \Xi(5) &= \{\{B\}\} \times \{\{b\}\}.
\end{aligned}$$

The profile (B, b) corresponds to the usual forward induction outcome. Propositions 2 and 3 together is our justification for claiming that common certain belief of full admissible consistency captures forward induction in the same way in any extensive game underlying the PRSF of Γ_4.

Analogous remarks apply to the Burning money game – which depicts a situation where a 'Battle-of-the-Sexes' game is preceded by an opportunity for one player to 'waste a unit of payoff' (see, e.g., Ben-Porath and Dekel (1992) for a more detailed discussion) – but with a twist. The player who cannot burn money will have two strategies in her unique fully permissible set; at the information set reached if money is burnt a maximal strategy may prescribe any action. By contrast IEWDS – the usual route to forward induction outcomes – permits only one specific action. This has been taken, as troublesome as it may seem, to suggest that burning is viewed as a 'signal of a rational player's intentions', despite burning in the end being an action a rational player would never use. Our solution of the Burning money game is robust to this critique. See AD for more discussion and details.

We next turn now to a game introduced by Dekel and Fudenberg (1990, Fig. 7.1), which is discussed also by Hammond (1993), and which is reproduced



	b	s
X	2, 2	2, 2
B	3, 1	0, 0
S	0, 0	1, 3

Fig. 4. Γ_4 (= *BoSwOO*) *and its PRSF*

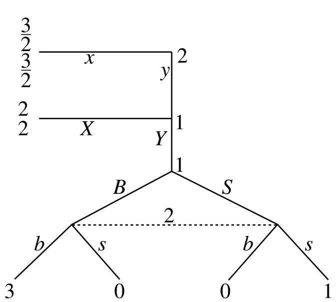
here as Γ_5 of Figure 5. It is a modification of Γ_4 which introduces an 'extra outside option' for player 2. In this game there may seem to be a tension between forward and backward induction: For player 2 not to choose x may seem to suggest that 2 'signals' that she seeks a payoff of at least 3/2, in contrast to the payoff of 1 that she gets when the subgame structured like Γ_4 is considered in isolation (as seen in the analysis of Γ_4). However, this intuition is not quite supported by the concept of fully permissible sets. Applying our algorithm IECFA to the PRSF of Γ_5 yields:

$$\begin{aligned}
\Xi(0) &= \Sigma_1 \times \Sigma_2 \\
\Xi(1) &= \{\{X\},\{B\},\{X,B\}\} \times \{\{x\},\{s\},\{x,b\},\{x,s\},\{b,s\},\{x,b,s\}\} \\
\Xi(2) &= \{\{X\},\{B\},\{X,B\}\} \times \{\{x\},\{x,b\},\{b,s\}\} \\
\Xi(3) &= \{\{B\},\{X,B\}\} \times \{\{x\},\{x,b\},\{b,s\}\} \\
\Pi = \Xi(4) &= \{\{B\},\{X,B\}\} \times \{\{x\},\{x,b\}\}.
\end{aligned}$$

The only way for X to be a maximal strategy for player 1 is that he deems $\{x\}$ as the only possible candidate for 2's choice set, in which case 1's choice set is $\{X,B\}$. Else $\{B\}$ is 1's choice set. Furthermore, 2 can have a choice set different from $\{x\}$ only if she deems $\{X, B\}$ as a possible candidate for 1's choice set. Intuitively this means that if 2's choice set differs from $\{x\}$ (i.e., equals $\{x, b\}$), then she deems it possible that 1 considers it impossible that b is a maximal strategy for 2. Since it is only under such circumstances that b is a maximal element for 2, perhaps this strategy is better thought of in terms of 'strategic manipulation' than in terms of 'forward induction'. Note that the concept of fully permissible sets has more bite than the Dekel–Fudenberg procedure; in addition to the strategies appearing in fully permissible sets also s survives the Dekel–Fudenberg procedure.

5.2. *An Economic Application*

Finally, we apply the concept of fully permissible sets to an economic model from organisation theory. Schotter (2000, ch. 8) discusses incentives schemes for firms and the moral hazard problems that may plague them. 'Revenue-sharing contracts', for example, often invite free-riding behaviour by the workers, and so lead



	x	b	s
X	$\frac{3}{2}, \frac{3}{2}$	2, 2	2, 2
B	$\frac{3}{2}, \frac{3}{2}$	3, 1	0, 0
S	$\frac{3}{2}, \frac{3}{2}$	0, 0	1, 3

Fig. 5. Γ_5 *and its PRSF*

to inefficient outcomes. However, Schotter points to 'forcing contracts' – incentive schemes of a kind introduced by Holmström (1982) – as a possible remedy: each worker is paid a bonus if and only if the collective of workers achieve a certain level of total production. If incentives are set right, then there is a symmetric and efficient Nash equilibrium in which each worker exerts a substantial effort. Each worker avoids shirking because he feels that his role is 'pivotal', believing that any reduction in effort leads to a loss of the bonus.

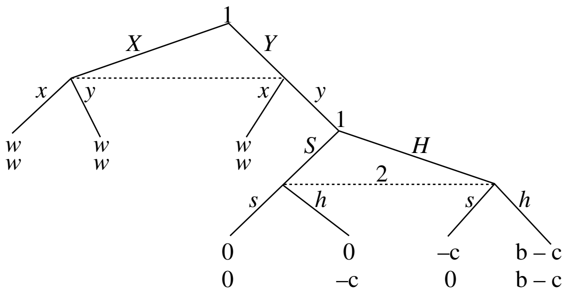
However, forcing contracts are often problematic in that there typically exists a Nash equilibrium in which no worker exerts any effort at all. How serious is this problem? Schotter offers the following argument in support of the forcing-contract (p. 302): 'While the no-work equilibrium for the forcing-contract game does indeed exist, it is unlikely that we will ever see this equilibrium occur. If workers actually accept such a contract and agree to work under its terms, we must conclude that they intend to exert the necessary effort and that they expect their coworkers to do the same. Otherwise, they would be better off obtaining a job elsewhere at their opportunity wage and not wasting their time pretending that they will work hard.'

Schotter appeals to intuition, but his argument has a forward induction flavour to it. We now show how the concept of fully permissible sets lends support. Consider the following situation involving a forcing contract: a firm needs two workers to operate. The workers simultaneously choose shirking at zero cost of effort, or high effort at cost $c > 0$. They get a bonus $b > c$ if and only if both workers choose high effort. As indicated above, this situation can be modelled as a game with two Nash equilibria (S, s) and (H, h), where (H, h) Pareto-dominates (S, s). However, let this game be a subgame of a larger game. In line with Schotter's intuitive discussion, add a preceding stage where each worker simultaneously decides whether to indicate willingness to join the firm with the forcing contract, or to work elsewhere at opportunity wage w, $0 < w < b - c$. The firm with the forcing contract is established if and only if both workers indicate willingness to join it.

This situation is depicted by the extensive game Γ_6. Again, we analyse the PRSF (Figure 6). Application of IECFA yields:

$$\begin{aligned}
\Xi(0) &= \Sigma_1 \times \Sigma_2 \\
\Xi(1) &= \{\{X\}, \{H\}, \{X, H\}\} \times \{\{x\}, \{h\}, \{x, h\}\} \\
\Xi(2) &= \{\{H\}, \{X, H\}\} \times \{\{h\}, \{x, h\}\} \\
\Pi = \Xi(3) &= \{\{H\}\} \times \{\{h\}\}.
\end{aligned}$$

Interpretation: $\Xi(1)$: Shirking cannot be a maximal strategy for either worker since it is weakly dominated. $\Xi(2)$: This excludes the possibility that a worker's choice set contains only the outside option. $\Xi(3)$: Since each worker certainly believes that hard work *is*, while shirking *is not*, an element of the opponent's choice set, it follows that each worker deems it infinitely more likely that the opponent chooses hard work rather than shirking. This means that, for each



	x	s	h
X	w, w	w, w	w, w
S	w, w	0, 0	0, $-c$
H	w, w	$-c$, 0	b-c, b-c

Fig. 6. Γ_6 *and its PRSF*

worker, only hard work is in his choice set, a conclusion that supports Schotter's argument.

6. Discussion

Here we have explored the implications of AD's concept of fully permissible sets in extensive games. AD characterise this concept as choice sets under common certain belief of full admissible consistency. Full admissible consistency entails that one strategy is preferred to another if and only if the former weakly dominates the latter on the union of the choice sets that are deemed possible for the opponent, or on the set of all opponent strategies.

Samuelson (1992) also presents an epistemic analysis of admissibility that leads to a collection of choice sets for each player. He requires that a player's choice set equals the set of strategies that are not weakly dominated on the union of choice sets that are deemed possible for the opponent, implying that our requirement of 'caution' is not imposed. Moreover, he defines possibility relative to a knowledge operator that satisfies the truth axiom, while we – in our analysis – allow a player to deem an opponent choice set to be impossible (or more precisely, Savage-null) even when it is the true choice set of the opponent. For example, in the 'Take-it-or-leave-it' game (Γ_1) our analysis includes a type of player 2 who deems it impossible for player 1 to have the choice set $\{D, FD\}$ even when this is the true choice set of player 1. This explains why we in contrast to Samuelson obtain general existence (Proposition 1(i)).

Battigalli (1996) has shown how Pearce's (1984) EFR corresponds to the 'best rationalisation principle'. This implies that some opponent strategies are neither completely rational nor completely irrational, but are considered to be at intermediate degrees of rationality. The present analysis, in contrast, differentiates only between whether a strategy is maximal (i.e., a rational choice) or not. In particular, although a strategy that is weakly dominated on the set of all opponent strategies is a 'stupid' choice, it need not be more 'stupid' than any remaining admissible strategy, as this depends on the interactive analysis of the game.

Battigalli and Siniscalchi (2002) have provided an interesting epistemic foundation for EFR. As discussed in AD, there are several differences between their model and the epistemic analysis we present in our companion paper.[3] However, the fact that Battigalli and Siniscalchi present an epistemic model for extensive games in contrast to the strategic form analysis of AD is of less importance for the applicability to extensive games than it might appear. The reason is that Battigalli and Siniscalchi consider rational choice *only* at information sets that are not precluded from being reached by the player's own strategy. This corresponds exactly to the implication of AD's analysis when 'caution' is imposed (Proposition 2 of the present paper).

This latter observation is related to the fact that our requirement of 'full belief of opponent rationality' is concerned with strategy choices of the opponent *in the whole game* only, not with choices among the remaining available strategies *at each and every information set*. To illustrate this point, look back at Γ_1 and consider a type of player 2 who deems $\{D\}$ as the only possible choice set for 1. Conditional on 2's node being reached it is clear that 1 cannot be choosing a strategy that is maximal given his preferences. Conditional on 2's node being reached, our modelling then imposes no constraint on 2's assessment of likelihood concerning which non-maximal strategy *FF* or *FD* that 1 has chosen. This crucially presumes that 2 assesses the likelihood of different strategies as chosen by player 1 *in the whole game*.

It is possible to model players being concerned with opponent choices *at all information sets*. In Γ_1 this would amount to the following when player 2 is of a type who deems $\{D\}$ as the only possible choice set for 1: conditional on 2's node being reached she realises that 1 cannot be choosing a *strategy* which is maximal given his preferences. Still, 2 considers it infinitely more likely that 1 at his last node chooses a strategy that is maximal among his remaining available strategies given his conditional preferences at that node. In Section 1 we argued, with Ben-Porath (1997), that it is not intuitively clear that this is reasonable, a view which permeates the working hypotheses on which the current work in grounded. Yet, the alternative approach is logically conceivable, and research on this basis is illuminating and worthwhile. However, contributions that to go in this alternative direction[4] appear to promote concepts that imply backward induction without yielding forward induction, and thus lead to implications that are significantly different from those of the present paper.

University of Oslo
Stockholm University

Date of receipt of first submission: March 2000
Date of receipt of final typescript: May 2002

[3] For example, while both contributions invoke non-monotonic epistemic operators to capture forward induction, Battigalli and Siniscalchi in contrast to AD use such an operator for the interactive epistemology and must employ a belief-complete epistemic model.

[4] See, e.g., Asheim (2002) and Asheim and Perea (2002) for epistemic models where each player believes that his opponent chooses rationally at all information sets. The former of these papers is related to Bernheim's (1984) 'subgame rationalisability', while the latter demonstrates how it – in accordance with Bernheim's conjecture – is possible to define 'sequential rationalisability'. See also Schuhmacher (1999) and Asheim (2001) for closely related strategic form analyses that define and characterise 'proper rationalisability' as a non-equilibrium analog to Myerson's (1978) 'proper equilibrium'.

References

Asheim, G.B. (2001). 'Proper rationalizability in lexicographic beliefs', *International Journal of Game Theory*, vol. 30, pp. 453–78.

Asheim, G.B. (2001*a*). 'On the epistemic foundation for backward induction', *Mathemathical Social Sciences*, vol. 44, pp. 121–44.

Asheim, G.B. and Dufwenberg, M. (2002). 'Admissibility and common belief', Forthcoming in *Games and Economic Behavior.*

Asheim, G.B. and Perea, A. (2002). 'Sequential and quasi-perfect rationalizability in extensive games', mimeo, University of Oslo and University of Maastricht.

Aumann, R. (1987). 'Game theory', in (J. Eatwell, M. Milgate and P. Newman, eds.) *The New Palgrave: A Dictionary of Economics*, pp. 460–82, London and Basingstoke: Macmillan Press.

Aumann, R. (1995). 'Backward induction and common knowledge of rationality', *Games and Economic Behavior*, vol. 8, pp. 6–19.

Basu, K. (1990). 'On the non-existence of a rationality definition for extensive games', *International Journal of Game Theory*, vol. 19, pp. 33–44.

Battigalli, P. (1996). 'Strategic rationality orderings and the best rationalization principle', *Games and Economic Behavior*, vol. pp. 13, 178–200.

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

Battigalli, P. and Siniscalchi, M. (2002). 'Strong belief and forward induction reasoning', *Journal of Economic Theory*, vol. 16, pp. 356–91.

Ben-Porath, E. (1997). 'Rationality, Nash equilibrium, and backwards induction in perfect information games', *Review of Economic Studies*, vol. 64, pp. 23–46.

Ben-Porath, E. and Dekel, E. (1992). 'Coordination and the potential for self-sacrifice', *Journal of Economic Theory*, vol. 57, pp. 36–51.

Bernheim, D. (1984). 'Rationalizable strategic behavior', *Econometrica*, vol. 52, pp. 1007–28.

Bicchieri, C. (1989). 'Self-refuting theories of strategic interaction: a paradox of common knowledge', *Erkenntnis*, vol. 30, pp. 69–85.

Binmore, K. (1987). 'Modelling rational players I', *Economics and Philosophy*, vol. 3, pp. 179–214.

Blume, L., Brandenburger, A., and Dekel, E. (1991) 'Lexicographic probabilities and choice under uncertainty', *Econometrica*, vol. 59, pp. 61–79.

Bonanno, G. (1991). 'The logic of rational play in games of perfect information', *Economics and Philosophy*, vol. 7, pp. 37–65.

Dekel, E. and Fudenberg, D. (1990). 'Rational behavior with payoff uncertainty', *Journal of Economic Theory*, vol. 52, pp. 243–67.

Dufwenberg, M. and Lindén, J. (1996). 'Inconsistencies in extensive games: common knowledge is not the issue', *Erkenntnis*, vol. 45, pp. 103–14.

Gul, F. (1997). 'Rationality and coherent theories of strategic behavior', *Journal of Economic Theory*, vol. 70, pp. 1–31.

Hammond, P. (1993). 'Aspects of rationalizable behavior', in (K.G. Binmore, A. Kirman and P. Tani, eds.) *Frontiers of Game Theory*, pp. 277–305, Cambridge, MA: MIT Press.

Holmström, B. (1982). 'Moral hazard in teams', *Bell Journal of Economics*, vol. 13, pp. 324–41.

Kreps, D.M. and Wilson, R. (1982). 'Sequential equilibria', *Econometrica*, vol. 50, pp. 863–94.

Mailath, G., Samuelson, L. and Swinkels, J. (1993). 'Extensive form reasoning in normal form games', *Econometrica*, vol. 61, pp. 273–302.

Mariotti, M. (1997). 'Decisions in games: why there should be a special exemption from Bayesian rationality', *Journal of Economic Methodology*, vol. 4, pp. 43–60.

Myerson, R. (1978). 'Refinement of the Nash equilibrium concept', *International Journal of Game Theory*, vol. 7, pp. 73–80.

Pearce, D.G. (1984). 'Rationalizable strategic behavior and the problem of perfection', *Econometrica*, vol. 52, pp. 1029–50.

Pettit, P. and Sugden, R. (1989). 'The backward induction paradox', *Journal of Philosophy*, vol. 4, pp. 169–82.

Reny, P.J. (1992). 'Backward induction, normal form perfection and explicable equilibria', *Econometrica*, vol. 60, pp. 627–49.

Reny, P.J. (1993). 'Common belief and the theory of games with perfect information', *Journal of Economic Theory*, vol. 59, pp. 257–74.

Rosenthal, R. (1981). 'Games of perfect information, predatory pricing and the chain-store paradox', *Journal of Economic Theory*, vol. 25, pp. 92–100.

Rubinstein, A. (1991). 'Comments on the interpretation of game theory', *Econometrica*, vol. 59, pp. 909–24.

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

Schotter, A. (2000). *Microeconomics: A Modern Approach*, 3rd ed., Boston: Addison Wesley Longman.
Schuhmacher, F. (1999). 'Proper rationalizability and backward induction', *International Journal of Game Theory*, vol. 28, pp. 599–615.
Selten, R. (1978). 'The chain store paradox', *Theory and Decision*, vol. 9, pp. 127–59.
Stalnaker, R. (1998). 'Belief revision in games: forward and backward induction', *Mathematical Social Sciences*, vol. 36, pp. 57–68.
van Damme, E. (1989). 'Stable equilibria and forward induction', *Journal of Economic Theory*, vol. 48, pp. 476–96.