MEMORANDUM

No 31/99

Proper Consistency

By
Geir B. Asheim

ISSN: 0801-1117

Department of Economics
University of Oslo

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

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

In co-operation with
The Frisch Centre for Economic Research

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

List of the last 10 Memoranda:

No 30	By Geir B. Asheim: On the Epistemic Foundation for Backward Induction. 26 p.
No 29	By Jon Strand: Labor unions versus individualized bargaining with heterogeneous labor. 36 p.
No 28	By Jon Vislie: The role of foreign ownership in domestic environmental regulation under Asymmetric. 19 p.
No 27	By Erik Biørn: Random Coefficients in Regression Equation Systems: The Case with unbalanced Panel Data. 19 p.
No 26	By Gunnar Bårdsen, Eilev S. Jansen and Ragnar Nymoen: Econometric inflation targeting. 38 p.
No 25	By Cheetal K Chand and Karl O. Moene: Rent Grabbing and Russia's Economic Collapse. 35 p.
No 24	By Rolf Golombek and Espen R. Moen: Do Voluntary Agreements Lead to Cost Efficiency? 45 p.
No 23	By Atle Seierstad: Necessary conditions involving generalized directional derivatives for optimal control of nonsmooth retarded Volterra integral equations in Banach space. 68 p.
No 22	By Karl Ove Moene and Michael Wallerstein: Inequality, Social Insurance and Redistribution. 40 p.
No 21	By Kai Leitemo: Inflation Targeting Strategies in small open Economies. 44 p.

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

PROPER CONSISTENCY

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:

PROPER CONSISTENCY

Abstract:

Proper consistency is defined by the properties that each player takes all opponent strategies into account (is *cautious*) and deems one opponent strategy to be infinitely more likely than another if the opponent prefers the one to the other (*respects preferences*). When there is common certain belief of proper consistency, a most preferred strategy is *properly rationalizable*. Any strategy used with positive probability in a proper equilibrium is properly rationalizable. Only strategies that lead to the backward induction outcome is properly rationalizable in the strategic form of a generic perfect information game. Proper rationalizability can be used to test the robustness of inductive procedures. *JEL* Classification Number: C72.

Date: 6 December 1999.

Acknowledgements: This paper builds in part on joint work with Martin Dufwenberg, who has contributed with helpful suggestions. I also thank Andrés Perea, Frank Schuhmacher and, especially, Ylva Søvik for useful comments.

1. Introduction

Most contributions on the relation between common knowledge/belief of rationality and backward induction in perfect information games perform the analysis in the extensive form of the game. An exception to this rule is Schuhmacher [24] who — based on Myerson's [22] concept of a *proper equilibrium*, but without making equilibrium assumptions — defines the concept of *proper rationalizability* in the strategic form and shows that proper rationalizable play leads to backward induction. Schuhmacher defines the set of properly rationalizable strategy vectors to be the limit of the set of ϵ-properly rationalizable strategy vectors, where the concept of ϵ-proper rationalizability is based on players doing mistakes, but where more costly mistakes are made with a much smaller probability than less costly ones. For a given ϵ, he offers an epistemic foundation for ϵ-proper rationalizability. However, for the limiting concept, i.e. proper rationalizability, there has not been an epistemic foundation available. It is the purpose of the present paper to establish how common certain belief of proper consistency characterizes proper rationalizability in strategic two-player games.

Blume et al. [9] characterize proper equilibrium as a property of preferences. When doing so they represent the preferences of players by the concept of a *lexicographic probability system* (LPS; Blume et al. [8]), which allows a player to deem one opponent strategy to be infinitely more likely than another while still taking the latter strategy into account. In two-player games, their characterization of proper equilibrium can be described by the following two properties.

1. Each player is certain of the LPS of his opponent,
2. Each player's LPS satisfies that the player takes all opponent strategies into account (is cautious) and that the player deems one opponent strategy to be infinitely more likely than another if the opponent prefers the one to the other (respects preferences).

In my characterization of proper rationalizability in two-player games I drop property 1., which is an equilibrium assumption; instead I assume that there is common certain belief of property 2., which I call proper consistency.

Since, in my framework, a player is not certain of the LPS of his opponent, player i's LPS must be defined on $S_j \times T_j$, where S_j denotes the set of opponent strategies and T_j denotes the set of opponent types. Each type of player i is simply an LPS on $S_j \times T_j$. A type t_i is said to be *cautious* if t_i takes into account all strategies of any opponent type that is not deemed Savage-null. A type t_i is said to *respect preferences* if, for any opponent type that is not deemed Savage-null, t_i deems one strategy of the opponent type to be infinitely more likely than another if the opponent type prefers the one to the other. A type t_i is said to be *properly consistent* with the preferences of his opponent if t_i is both cautious and respects preferences. Hence, the present analysis follows Asheim & Dufwenberg [4] (AD) in arguing that in deductive game theory, requirements should be imposed on the beliefs of players rather than their choice. Since the beliefs of players determine their preferences, this amounts to imposing requirements on preferences.[1] A type t_i certainly believes the event that his opponent is of a type that is properly consistent if he deems Savage-null any opponent type that is not properly consistent. There is *common certain belief of proper consistency* at $t = (t_1, t_2)$ if both t_1 and t_2 certainly believes the event that his opponent is of a type that is properly consistent, they both certainly believes that his opponent certainly believes the event that his opponent is of a type that is properly consistent, etc.

A pure strategy s_i is called *properly rationalizable* if there is a set of type vectors $T = T_1 \times T_2$ with common certain belief of proper consistency at $t = (t_1, t_2)$ such that s_i is a most preferred strategy given t_i's LPS. It is the first main result that any pure strategy used with positive probability in a proper equilibrium is properly rationalizable. The second main result is that the present paper's definition of proper rationalizability corresponds to that of Schuhmacher [24]: A pure strategy is properly rationalizable in the sense of the present paper if and only if it is used with positive probability in some properly rationalizable mixed strategy in the sense of Schuhmacher [24]. The third main result is to apply my definition to show that only strategies that lead to the backward induction outcome are properly rationalizable in the normal form of a generic perfect information game. Thus, Schuhmacher's Theorem 2 (which shows that the backward induction outcome obtains with "high" probability for any given "small" ϵ) is strengthened, and an epistemic foundation for the backward induction procedure (as an alternative to Aumann's [5]) is provided. Lastly, it is illustrated through an example how proper rationalizability can be used to test the robustness of inductive procedures.

The analysis is limited to 2-player games. The extension to general (n-player) games raises the issue of whether (and if so, how) each player's belief about the strategy choices of the other players are stochastically independent. This is outside the scope of the present paper.

2. An Illustration

The symmetric game of Fig. 1 is an example where common certain belief of proper consistency is sufficient to determine completely each player's preferences over his or her own strategies. The game is due to Blume et al. ([9], Fig. 1).

[1] Instead of imposing rational choice in the sense that a driver *chooses* to drive on the right side of the road if he believes that his opponent chooses to drive on the right side of the road, AD suggest to impose consistent preferences in the sense that a driver *prefers* to drive on the right side of the road if he believes that his opponent prefers to drive on the right side of the road. This follows a tradition in equilibrium analysis where Nash (perfect/proper) equilibrium is defined as an equilibrium in conjectures (cf. Blume et al. [9]).

	L	C	R
U	1, 1	1, 1	1, 0
M	1, 1	2, 2	2, 2
D	0, 1	2, 2	3, 3

FIGURE 1. Illustrating common certain belief of proper consistency

In this game, caution implies that player 1 prefers M to U since M weakly dominates U. Likewise, player 2 prefers C to L. Since 1 respects the preferences of 2 and, in addition, certainly believes that 2 is cautious, it follows that 1 deems C infinitely more likely than L. This in turn implies that 1 prefers D to U. Likewise, since 2 respects the preferences of 1 and, in addition, certainly believes that 1 is cautious, it follows that 2 prefers R to L. As a consequence, since 1 respects the preferences of 2, certainly believes that 2 respects the preferences of 1, and certainly believes that 2 certainly believes that 1 is cautious, it follows that 1 deems R infinitely more likely than L. Consequently, 1 prefers D to M. A symmetric reasoning entails that 2 prefers R to C. Hence, if there is common certain belief of proper consistency, it follows that the players' preferences over their own strategies are given by

$$U \prec M \prec D$$
$$L \prec C \prec R\,.$$

The facts that D is the unique most preferred strategy for 1 and R is the unique most preferred strategy for 2 means that only D and R are properly rationalizable (cf. Def. 1 of Sect. 4.1). By Prop. 1 of Sect. 4.2, it then follows that the pure strategy vector (D, R) is the unique proper equilibrium, which can easily be checked by inspection. However, note that in the argument above, each player obtains certainty about the preferences of his opponent through deductive reasoning; i.e. such certainty is not assumed as in the concept of proper equilibrium.

The concept of proper rationalizability yields a strict refinement of (ordinary) rationalizability. All strategies for both players are rationalizable, which is implied by the fact that, in addition to (D, R), the pure strategy vectors (U, L) and (M, C) are also Nash equilibria. The concept of proper rationalizability yields even a strict refinement when compared to the Dekel-Fudenberg [14] procedure, which consists of one round of weak elimination followed by iterated strong elimination, and which follows from there being common certain belief of caution and belief of opponent rationality (see e.g. Brandenburger [12] and Börgers [11]). When the Dekel-Fudenberg procedure is employed, only U is eliminated for 1, and only L is eliminated for 2, reflecting that also the pure strategy vector (M, C) is a perfect equilibrium. It is a general result that proper rationalizability refines the Dekel-Fudenberg procedure (cf. Thm. 4 of Herings & Vannetelbosch [18] as well as Remark 1 below).

3. States, Types, Preferences, and Belief

The purpose of this section is to present a framework for strategic games where each player is modeled as a decision maker under uncertainty. The decision-theoretic analysis builds on Blume et al. [8]. 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.

3.1. **A Strategic Game Form.** With $N = \{1, 2\}$ as the set of *players*, let S_i denote player i's finite set of *pure strategies*, and let $z : S \to Z$ map strategy vectors into *outcomes*, where $S = S_1 \times S_2$ is the set of strategy vectors and Z is the set of outcomes. Then $((S_i)_{i \in N}, z)$ is a finite *strategic* two-player *game form*. Write p, r, and s ($\in S$) for pure strategy vectors.

3.2. **States and Types.** When a strategic game form is turned into a decision problem for each player (see Tan & Werlang [29]), the uncertainty faced by a player is 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 [2], Böge & Eisele [10], Mertens & Zamir [20], Brandenburger & Dekel [13], Epstein & Wang [17]) 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. Moreover, since continuity is not imposed, the "regularity" conditions on $\succeq^{t_i}$ consist of *completeness*, *transitivity*, *objective independence*, *nontriviality*, *conditional continuity* and *non-null state independence*, meaning that $\succeq^{t_i}$ is represented by a vNM utility function $v_i^{t_i} : Z \to \mathbb{R}$ that assigns a payoff to any outcome and a lexicographic probability system (LPS) $\lambda^{t_i} = (\mu_1^{t_i}, \dots, \mu_L^{t_i}) \in \mathbf{L}\Delta(S_j \times T_j)$ (cf. Blume et al. [8] and the Appendix). Being a vNM utility function, $v_i^{t_i}$ can be extended to objective randomizations on Z.

The construction is summarized by the following definition.

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

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

3.3. **Certain Belief.** For each player i, i's certain belief can be derived from the belief system. To state this epistemic operator, let, for each player i and each state $\omega \in \Omega$, $t_i(\omega)$ denote the projection of ω on T_i, and let, for any $E \subseteq \Omega$, $E_j^{t_i} := \{(s_j, t_j) \in S_j \times T_j | \exists (s'_1, s'_2, t'_1, t'_2) \in E \text{ s.t. } (s'_j, t'_j) = (s_j, t_j) \text{ and } t'_i = t_i\}$. Associate 'certain belief' of an event with the property that no element of the complement of the event is assigned positive probability by some probability distribution in λ^{t_i}:

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

where $\kappa_j^{t_i} := \text{supp}\lambda^{t_i}$ $(\subseteq S_j \times T_j)$ denotes the set of opponent strategy-type pairs that t_i does not deem *Savage-null*.[2] Say that i certainly believes the event $E \subseteq \Omega$ given ω if $\omega \in K_i E$ (or equivalently, $\kappa_j^{t_i(\omega)} \subseteq E_j^{t_i(\omega)}$). Write $KE := K_1 E \cap K_2 E$. Say that there is mutual certain belief of $E \subseteq \Omega$ given ω if $\omega \in KE$. Write $CKE := KE \cap KKE \cap KKKE \cap \dots$. Say that there is common certain belief of $E \subseteq \Omega$ given ω if $\omega \in CKE$.

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

[2] The term 'certain belief' for this notion is also used by Morris [21].

3.5. **A Strategic Game.** Let, for each i, $u_i : S \to \mathbb{R}$ be a vNM utility function that assigns payoff to any strategy vector. Then $G = (S_i, u_i)_{i \in N}$ is a finite *strategic* 2-player *game*. 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

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

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

4. Consistency of Preferences

Usually requirements in deductive game theory are imposed on choice. E.g. rationality is a requirement on a pair (s_i, t_i), where s_i is said to be a 'rational choice' by t_i if $s_i \in C_i^{t_i}$. See e.g. Epstein ([16], Sect. 6) for a presentation of this approach in a general context.

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. Remark 1 below).
- ... requirements like caution and respect of opponent preferences to be imposed in a straightforward manner. In order to accommodate caution under the 'rational choice' approach, the notion of 'certain belief' must be weakened (cf. Börgers ([11], pp. 266–267) and Epstein ([16], p. 3)). It is unclear how respect of opponent preferences can be accommodated under the 'rational choice' approach.

Here I will focus on showing how 'consistent preferences' as an approach to deductive game theory can be used to define *proper rationalizabiliy* as a non-equilibrium analogue to Myerson's [22] *proper equilibrium* in 2-player games.[3] In the same way as Blume et al. [9] characterize proper equilibrium as an equilibrium in conjectures (which does not entail that players tremble, but that each player takes into account the possibility that the opponent can tremble), the present definition of proper rationalizability stems from requirements on preferences. In particular, it differs from Schuhmacher's [24] main statement of his definition by not modeling players that tremble; instead each player takes into account the possibility that the opponent can tremble, he certainly believes that his opponent takes into account the possibility that he himself can tremble, and so on.[4]

[3] As mentioned in the introduction, an extension to games with more than 2 players raises the issue of independence, which will not be addressed here.

[4] In a concluding discussion, Schuhmacher [24] also considers defining ϵ-proper rationalizability by posing requirements on the players' beliefs rather than on their mixed

4.1. **Proper Consistency.** Proper consistency will be based on 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 (is *cautious*), while the third requirement ensures that each player deems one opponent strategy to be infinitely more likely than another if the opponent prefers the one to the other (*respects preferences*).

To impose these requirements, consider the following events

$$[cau_i] := \{\omega \in \Omega | \kappa_j^{t_i(\omega)} = S_j \times T_j^{t_i(\omega)}\}$$
$$[resp_i] := \{\omega \in \Omega | (r_j, t_j) \gg (s_j, t_j) \text{ acc. to } \succeq^{t_i(\omega)} \text{ whenever } t_j \in T_j^{t_i(\omega)} \text{ and } r_j \succ_{S_i}^{t_j} s_j\},$$

where $T_j^{t_i} := \text{proj}_{T_j} \kappa_j^{t_i}$ denotes the set of opponent types that t_i does not deem Savage-null, and where $\gg$ means 'infinitely more likely' (cf. the Appendix).

- If $\omega \in [cau_i]$, then (s_j, t_j) is not deemed Savage-null acc. to $\succeq^{t_i(\omega)}$ whenever t_j is not deemed Savage-null. This means that, $\forall (s_j, t_j) \in S_j \times T_j^{t_i(\omega)}$, $\omega \notin K_i\{(s'_1, s'_2, t'_1, t'_2) \in \Omega | (s'_j, t'_j) \neq (s_j, t_j)\}$ (cf. Dekel & Gul's [15] definition of caution). It implies that the marginal of $\succeq^{t_i(\omega)}$ on S_j (i.e., $t_i(\omega)$'s preferences over S_i, $\succeq_{S_j}^{t_i(\omega)}$) is admissible on S_j.
- If $\omega \in [resp_i]$, then $t_i(\omega)$ respects the preferences of any opponent type that is not deemed Savage-null.

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

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

Refer to $A^{prop} := A_1^{prop} \cap A_2^{prop}$ as the event of *proper consistency*. The concept of properly rationalizable strategies can now be defined as most preferred strategies in states where there is common certain belief of proper consistency.

Definition 2. A pure strategy r_i for i is *properly rationalizable* in a finite strategic two-player game G if there exists a belief system with $r_i \in C_i^{t_i(\omega)}$ for some $\omega \in CKA^{prop}$.

4.2. **Results.** It follows from Blume et al.'s [9] characterization of proper equilibrium in two-player games that any strategy used with positive probability in a proper equilibrium is properly rationalizable.

Proposition 1. *If* $(x_1, x_2) \in \Delta(S_1) \times \Delta(S_2)$ *is a proper equilibrium in a finite strategic two-player game* G*, then, for each* i*, any* $s_i \in \text{supp} x_i$ *is properly rationalizable.*

---

strategies. His alternative formulation differs from the one offered in this section both in terminology (by having 'types' refer to (s_i, t_i) pairs) and in analysis (by more costly trembles being deemed *much* less likely – rather than *infinitely* less likely – than less costly ones).

Proof. Let (x_1, x_2) be a proper equilibrium. By Blume et al.'s [9] Prop. 5, there exists a pair of preferences, $\succeq^{t_1}$ and $\succeq^{t_2}$, that are represented by $\upsilon_1^{t_1}$ and $\lambda^{t_1} = (\mu_1^{t_1}, \dots) \in \mathbf{L}\Delta(S_2 \times \{t_2\})$, and $\upsilon_2^{t_2}$ and $\lambda^{t_2} = (\mu_1^{t_2}, \dots) \in \mathbf{L}\Delta(S_1 \times \{t_1\})$, respectively — with $\upsilon_1^{t_1} \circ z = u_1$ and, $\forall s_2 \in S_2$, $\mu_1^{t_1}(s_2, t_2) = x_2(s_2)$, and $\upsilon_2^{t_2} \circ z = u_2$ and, $\forall s_1 \in S_1$, $\mu_1^{t_2}(s_1, t_1) = x_1(s_1)$ — such that t_1 and t_2 are cautious and respect preferences. Moreover, $\forall i \in N$, $\text{supp} x_i \subseteq C_i^{t_i}$. Let $\Omega = S \times \{t_1\} \times \{t_2\}$. Then $\Omega = A^{prop} = CKA^{prop}$, implying that, $\forall i \in N$, any $s_i \in C_i^{t_i}$ is properly rationalizable. □

In the construction of the proof, each player certainly believes the preferences of his opponent. This is an equilibrium assumption, which is not satisfied in general when there is common certain belief of proper consistency.

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

Corollary 1. *In any finite strategic two-player game G, there exists a belief system with $CKA^{prop} \neq \emptyset$, implying that there exists, for each i, a nonempty set of properly rationalizable strategies.*

Remark 1. Substitute the event

$$B_i[rat_j] := \{\omega \in \Omega | (r_j, t_j) \in \text{supp}\mu_1^{t_i(\omega)} \text{ implies } r_j \in C_j^{t_j}\}$$

for $[resp_i]$, where $\mu_1^{t_i(\omega)}$ is the primary probability distribution in $t_i(\omega)$'s LPS, $\lambda^{t_i(\omega)}$. Write $A_i := [u_i] \cap [cau_i] \cap B_i[rat_j]$ and $A := A_1 \cap A_2$. Then a strategy r_i surviving the Dekel-Fudenberg procedure can be characterized by the property that there exists a belief system with $r_i \in C_i^{t_i(\omega)}$ for some $\omega \in CKA$ (cf. AD, Prop. 3). Since $[resp_i] \subseteq B_i[rat_j]$, it follows from Def. 2 that proper rationalizability refines the Dekel-Fudenberg procedure. The notation reflects that $B_i[rat_j]$ can be interpreted as the event that i believes (with probability one) that j is rational.

Next it is shown that Def. 2 is equivalent to Schuhmacher's [24] definition of proper rationalizability, thereby establishing an epistemic foundation for the concept that he defines.

Proposition 2. *Consider a finite strategic two-player game G. A pure strategy r_i for i is used with positive probability in some mixed strategy that is properly rationalizable according to the definition of Schuhmacher* [24] *if and only if it is properly rationalizable according to Def. 2.*

Remark 2. Schuhmacher [24] considers a set of type vectors $T = T_1 \times T_2$, where each type t_i of either player i plays a completely mixed strategy $x_i^{t_i}$ and has a probability distribution on $S_j \times T_j$, for which the conditional distribution on $S_j \times \{t_j\}$ coincides with $x_j^{t_j}$ whenever the conditional distribution is defined. He defines ϵ-proper rationalizability through the ϵ-proper trembling condition, which is satisfied by a type t_i of player i if $\epsilon x_i^{t_i}(r_i) \geq x_i^{t_i}(s_i)$ whenever t_i prefers r_i to s_i. His formulation implies that all types of a player agrees not only on the preferences but also on the relative likelihood of the strategies for any given opponent type. In contrast, Def. 2 of Sect. 4.1

requires the types of a player only to agree on the preferences of any given opponent type. This difference implies that expanded type sets must be constructed for the sufficiency part of the proof of Prop. 2.

Proof of Prop. 2. Sufficiency. Suppose that r_1^* is properly rationalizable for 1 according to Def. 2. Hence, there exists a belief system with $r_1^* \in C_1^{t_1(\omega^*)}$ for some $\omega^* \in CKA^{prop}$. Let, $\forall i \in N, T_i' := \{t_i(\omega) | \omega \in CKA^{prop}\}$. Note that, $\forall i \in N$ and $\forall \omega \in CKA^{prop}$, $T_j^{t_i(\omega)} \subseteq T_j'$ since $CKA^{prop} = KCKA^{prop} \subseteq K_i CKA^{prop}$. For each type t_i of either player i, make as many clones of t_i as there are members of T_j': $\forall i \in N$, $T_i'' := \{\mathbf{t}_i(t_i,t_j) | t_i \in T_i'$ and $t_j \in T_j'\}$, where $\mathbf{t}_i(t_i,t_j)$ is the clone of t_i associated with t_j. Let $\mathbf{t}_i(t_i,t_j)$ "share" the preferences of t_i in the sense that

1. the set of opponent types that $\mathbf{t}_i(t_i,t_j)$ does not deem Savage-null, $T_j^{\mathbf{t}_i(t_i,t_j)}$, is equal to $\{\mathbf{t}_j(t_j',t_i) | t_j' \in T_j^{t_i}\}$ ($\subseteq T_j''$ since $T_j^{t_i} \subseteq T_j'$), and
2. the likelihood of $(s_j, \mathbf{t}_j(t_j',t_i))$ according to $\succeq^{\mathbf{t}_i(t_i,t_j)}$ is equal to the likelihood of (s_j, t_j') according to $\succeq^{t_i}$.

Consider any $(t_1,t_2) \in T_1' \times T_2'$. Since $CKA^{prop} \subseteq [u_i]$, $\succeq^{\mathbf{t}_i(t_i,t_j)}$ can be represented by a vNM utility function $v_i^{t_i}$ satisfying $v_i^{t_i} \circ z = u_i$ and an LPS on $S_j \times T_j^{\mathbf{t}_i(t_i,t_j)}$. Since $CKA^{prop} \subseteq [cau_i]$, this LPS yields, for each $t_j' \in T_j^{t_i}$, a partition $\{E_1, \dots, E_{L(t_i,t_j')}\}$ of $S_j \times \{\mathbf{t}_j(t_j',t_i)\}$, where

$$(r_j, \mathbf{t}_j(t_j',t_i)) \gg (s_j, \mathbf{t}_j(t_j',t_i)) \text{ acc. to } \succeq^{\mathbf{t}_i(t_i,t_j)}$$

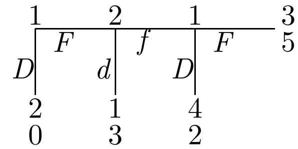
whenever $(r_j, \mathbf{t}_j(t_j',t_i)) \in E_\ell$, $(s_j, \mathbf{t}_j(t_j',t_i)) \in E_{\ell'}$, and $\ell < \ell'$. Since $CKA^{prop} \subseteq [resp_i]$ and $\mathbf{t}_j(t_j',t_i)$ "shares" the preferences of t_j', each $r_j \in E_\ell$ is a most preferred strategy in $E_\ell \cup \dots \cup E_{L(t_i,t_j')}$ for any $\ell \in \{1,\dots,L(t_i,t_j')\}$.

By following the sufficiency part of the proof of Blume et al. ([9], Prop. 5) in their application of [9], Prop. 1, converging sequences $\{x_i^{\mathbf{t}_i(t_i,t_j)}(n)\}_{n=0}^\infty$ of completely mixed strategies and $\{\mu^{\mathbf{t}_i(t_i,t_j)}(n)\}_{n=0}^\infty$ of probability distributions on $S_j \times T_j''$ can be constructed for any type $\mathbf{t}_i(t_i,t_j) \in T_i''$ of either player i, such that, for any n,

1. for any $\mathbf{t}_j(t_j',t_i) \in T_j^{\mathbf{t}_i(t_i,t_j)}$, the conditional distribution of $\mu^{\mathbf{t}_i(t_i,t_j)}(n)$ on $S_j \times \{\mathbf{t}_j(t_j',t_i)\}$ coincides with $x_i^{\mathbf{t}_j(t_j',t_i)}(n)$,
2. $\mathbf{t}_i(t_i,t_j)$ satisfies the $\epsilon(n)$-proper trembling condition, and
3. $C_i^{t_i}$ is the set of most preferred strategies according to $\mu_i^{\mathbf{t}_i(t_i,t_j)}(n)$,

and such that $\epsilon(n) \to 0$ as $n \to \infty$. Add type t_1^* to T_1'' having the property that $\mu^{t_1^*}(n) = \mu^{\mathbf{t}_1(t_1(\omega^*),t_2)}(n)$ for some $t_2 \in T_2'$, but where $x_1^{t_1^*}(n) = \gamma x_1^* + (1-\gamma) x_1^{\mathbf{t}_1(t_1(\omega^*),t_2)}(n)$ with $0 < \gamma < 1$ and $x_1^*(r_1^*) = 1$. For any n, the $\epsilon(n)$-proper trembling condition is satisfied for all types in $T_1'' \cup \{t_1^*\}$ and all types in T_2''. This shows that r_1^* is played with positive probability in some mixed strategy that is properly rationalizable according to the definition of Schuhmacher [24].

Necessity. Suppose there exist converging sequences $\{x_i^{t_i}(n)\}_{n=0}^\infty$ of completely mixed strategies and $\{\mu^{t_i}(n)\}_{n=0}^\infty$ of probability distributions on $S_j \times T_j$ for any type $t_i \in T_i$ of either player i, such that, for any n, the



	d	f
D	2, 0	2, 0
FD	1, 3	4, 2
FF	1, 3	3, 5

Figure 2. A centipede game

$\epsilon(n)$-proper trembling condition is satisfied for all types in T_1 and all types in T_2, and such that $\epsilon(n) \to 0$ as $n \to \infty$. By following the necessity part of the proof of Blume et al. ([9], Prop. 5) in their application of [9], Prop. 2, an LPS on $S_j \times T_j$ can be constructed for any type $t_i \in T_i$ of either player i, such that $r_i \in C_i^{t_i}$ if $x_i^{t_i}(n)(r_i)$ does not converge to 0 as $n \to \infty$, and such that $\Omega = A^{prop} = CKA^{prop}$ if $\Omega = S \times T_1 \times T_2$. This shows that any pure strategy played with positive probability in some mixed strategy that is properly rationalizable according to the definition of Schuhmacher [24] is properly rationalizable according to Def. 2. □

5. Backward Induction

Consider the centipede game of Fig. 2. The backward induction argument in this game goes as follows: If 1's second decision node is reached and 1 chooses a most preferred strategy in the subgame, then 1 will choose D. If 2 knows/believes this and chooses a most preferred strategy in the subgame defined by her decision node, then 2 will choose d. If 1 knows/believes this and chooses a most preferred strategy at the beginning of the game, then 1 will choose D at his first decision node.

It has been hard to provide an unquestionable model of interactive epistemology that supports this simple backward induction argument (cf. Stalnaker [28]). E.g. two widely cited and influential contributions — Aumann [5] and Ben-Porath [7] — reach opposite conclusions on whether common knowledge/belief of rationality implies backward induction. The problem in the centipede game is that if 1 chooses F at his first decision node and thus do not play in accordance with backward induction, then 2 may not believe that 1 will play in accordance with backward induction at 1's last decision node. If so, she may choose f, which in turn opens for the possibility that the strategy FD may be a most preferred strategy for 1.[5]

In the game of Fig. 2 the backward induction argument corresponds to iterated elimination of weakly dominated strategies in the strategic form of the game (see the right part of Fig. 2): First FF is eliminated for 1, then f is eliminated for 2, and finally, FD is eliminated for 1. However, also for this procedure it is hard to provide a convincing model of interactive epistemology (cf. Stahl [27]). The problem in the strategic game of Fig. 2

[5]The epistemic analysis of Battigalli & Siniscalchi [6] provides a foundation for the backward induction *outcome* in the whole game. Their model of interactive epistemology does not, however, support the backward induction *argument*.

is that the elimination of f in the second round removes the reason why FF and not FD was eliminated in the first round. However, it is exactly since FF and not FD was eliminated in the first round that f can be eliminated in the second round. And without the elimination of f, FD cannot be eliminated.

It is straightforward to see how common certain belief of proper consistency implies that players have preferences in accordance with backward induction in the sense that, in any subgame, the backward induction outcome is reached if each player chooses a most preferred strategy: Caution implies that player 1 prefers FD to FF since FD weakly dominates FF. Since 2 respects the preferences of 1 and, in addition, certainly believes that 1 is cautious, it follows that 2 deems FD infinitely more likely than FF. This in turn implies that 2 prefers d to f. As a consequence, since 1 respects the preferences of 2, certainly believes that 2 respects the preferences of 1, and certainly believes that 2 certainly believes that 1 is cautious, it follows that 1 deems d infinitely more likely than f. Consequently, 1 prefers D to FD. Hence, it follows that the players' preferences over their own strategies are given by

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

Note that if there is common certain belief of proper consistency, then, in any subgame, the play of most preferred strategies lead to the backward induction outcome. The remaining part of Sect. 5 shows that this result holds for any generic extensive game of perfect information.

5.1. **Preliminaries.** A finite extensive game of almost perfect information Γ with 2 players and $M-1$ stages can be described as follows. 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, i's finite action set be denoted $A_i(h)$, where i is inactive at h if $A_i(h)$ is a singleton. Write $A(h) := A_1(h) \times A_2(h)$. Define the set of histories at the beginning of stage $m+1$ as follows: $H^{m+1} := \{(h,a)|h \in H^m \text{ and } a \in A(h)\}$. This concludes the induction. Let $H := \bigcup_{m=1}^{M-1} H^m$ denote the set of *subgames* and let $Z := H^M$ denote the set of *outcomes*.

A *pure strategy* for player i is a function s_i that assigns an action in $A_i(h)$ to any $h \in H$. Let S_i denote player i's finite set of pure strategies. Let $z : S \to Z$ map strategy vectors into outcomes.[6] Then $((S_i)_{i\in N}, z)$ is a finite strategic two-player game form. If, for each player i, i's payoff in Γ is

[6] 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 [23]. As there is no need here to differentiate between identical acts, the concept of a plan of action would have sufficed.

determined by the vNM utility function $\upsilon_i : Z \to \mathbb{R}$, then $G = (S_i, u_i)_{i \in N}$ is the strategic game corresponding to Γ, where, for each i, $u_i : S \to \mathbb{R}$ is defined by $u_i = \upsilon_i \circ z$. For any $h \in H \cup Z$, there exists a strategic form structure: Let $S(h) = S_1(h) \times S_2(h)$ denote the set of strategy vectors that are *consistent* with h being reached. If h' is the predecessor of h, then $S(h') \supseteq S(h)$. If $s_i \in S_i$ and $h \in H$, let $s_i|_h$ denote the strategy in $S_i(h)$ satisfying $s_i|_h(h') = s_i(h')$ at any $h' \in H$ except at h' with $S(h') \supset S(h)$ where $s_i|_h(h')$ is dictated by $s_i|_h$ being consistent with h.

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. It is *generic* if, for each i, $\upsilon_i(z) \neq \upsilon_i(z')$ whenever z and z' are different outcomes. Generic extensive games of perfect information have a unique subgame-perfect equilibrium. Moreover, in such games the procedure of backward induction yields in any subgame the unique subgame-perfect equilibrium outcome. 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.

For each type t_i of player i, $\succeq^{t_i}_{S_j(h)}$ is t_i's preferences over i's pure strategies $S_i(h)$ in any subgame $h \in H$. Let $C^{t_i}_i(h) := \{r_i \in S_i(h) | \forall s_i \in S_i(h), r_i \succeq^{t_i}_{S_j(h)} s_i\}$ denote t_i's set of most preferred strategies in the subgame h. Refer to $C^{t_i}_i : H \to 2^{S_i(h)} \backslash \{\emptyset\}$ as t_i's *choice function*. Note that $C^{t_i}_i = C^{t_i}_i(\emptyset)$, and write, for any $h \in H$, $C^t(h) := C^{t_1}_1(h) \times C^{t_2}_2(h)$. By Lemma 1, if s_i is most preferred in a subgame h, then s_i is most preferred in any later subgame that s_i is consistent with.

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

Proof. Suppose that s_i is not a most preferred strategy in h'. Then there exists $r_i \in S_i(h')$ such that $r_i \succ^{t_i}_{S_j(h')} s_i$. It follows from Mailath et al. ([19], Defs. 2 and 3 and the if-part of Theorem 1) that $S(h')$ is a *strategic independence* for i. Hence, r_i can be chosen such that $z(r_i, s_j) = z(s_i, s_j)$ for all $s_j \in S_j \backslash S_j(h')$. This implies that $r_i \succ^{t_i}_{S_j(h)} s_i$, which contradicts that s_i is a most preferred strategy in h. □

5.2. Result on Backward Induction.

In analogy with Aumann's [5] Theorem A, it is established that any vector of most preferred strategies in a subgame of a generic perfect information game, in a state where there is common certain belief of proper consistency, leads to the backward induction outcome in the subgame (Prop. 3). The analogy of Aumann's Theorem B — that for any generic perfect information game, common certain belief of proper consistency is possible; i.e. that the result of Prop. 3 is not empty — has already been established through Cor. 1 since any extensive game of (almost) perfect information Γ has a corresponding strategic game G.

Proposition 3. *Consider a finite generic 2-player extensive game of perfect information* Γ *with corresponding strategic game* G*. If, for some belief system for* G*,* $\omega \in CKA^{prop}$*, then, for each* $h \in H$*,* $C^{t(\omega)}(h) \subseteq S(z(p|_h))$*, where* p *denotes the unique subgame-perfect equilibrium.*

Proof. Some properties of the certain belief operator (see Sect. 3.3) must be established for the proof. It is easy to check that $K_i\Omega = \Omega$ and $K_i\emptyset = \emptyset$, and, for any events E and F, $K_iE \cap K_iF = K_i(E \cap F)$, $K_iE \subseteq K_iK_iE$, and $\neg K_iE \subseteq K_i(\neg K_iE)$, implying that, for any event E, $K_iE = K_iK_iE$. Write $K^0E := E$ and, for each $g \geq 1$, $K^gE := KK^{g-1}E$. Since $K_i(E \cap F) = K_iE \cap K_iF$ and $K_iK_iE = K_iE$, it follows $\forall g \geq 2$,

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

The truth axiom ($K_iE \subseteq E$) is not satisfied, since an event can be certainly believed even though the true state is an element of the complement of the event. However, since $A^{prop} = A_1^{prop} \cap A_2^{prop}$ is an event that concerns the type vector, mutual certain belief of A^{prop} implies that A^{prop} is true: $KA^{prop} = K_1A^{prop} \cap K_2A^{prop} \subseteq K_1A_1^{prop} \cap K_2A_n^{prop} = A_1^{prop} \cap A_2^{prop} = A^{prop}$ since, for each i, $K_iA_i^{prop} = A_i^{prop}$. Hence, it follows that *(i)* $\forall g \geq 1$, $K^gA^{prop} \subseteq K^{g-1}A^{prop}$, and *(ii)* $\exists g' \geq 0$ such that $K^gA^{prop} = CKA^{prop}$ for $g \geq g'$ since Ω is finite.

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

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

($g = 1, \dots, M-2$) Suppose that it has been established for $g' = 0, \dots, g-1$ that if there exists a belief system with $\omega \in K^{g'}A^{prop}$, 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 $\omega \in K^{g-1}A^{prop}$. Then, by Lemma 1 and the premise, $S_j(h) = S_j(h,a)$ and

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

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

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

Now, consider i with a non-singleton action set at h; since Γ has perfect information, there is at most one such i. Let $t_i = t_i(\omega)$ for some $\omega \in K^gA^{prop}$. The preceding argument implies that $C_j^{t_j}(h) \subseteq \bigcap_a S_j(z(p|_{(h,a)}))$ whenever $t_j \in T_j^{t_i}$ since $\omega \in K^gA^{prop} \subseteq K_iK^{g-1}A^{prop}$. 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 $\mathbf{x}_{S_j}$ for the act on S_j that $p_i|_h$ can be viewed as, and write $\mathbf{y}_{S_j}$ for the act on S_j that s_i can be viewed as. Let $\mathbf{x}$ and $\mathbf{y}$ be the acts on $S_j \times T_j$ that satisfy $\mathbf{x}(s_j, t_j) = \mathbf{x}_{S_j}(s_j)$ and $\mathbf{y}(s_j, t_j) = \mathbf{y}_{S_j}(s_j)$ for all (s_j, t_j). Then,

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

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

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

and, thus, $\omega \in K^g A^{prop} \subseteq [resp_i]$ implies that

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

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

Remark 3. The proof of Prop. 3 illustrates the importance of defining proper rationalizability by imposing common *certain belief* of proper consistency, where 'certain belief' of an event means that the complement of the event is deemed Savage-null. Common *belief* of proper consistency, where 'belief' is used in the sense of 'belief with probability one', would not imply backward induction. It is straightforward to show that the belief system for a four-legged centipede game presented in Table 2 of Asheim [3] constitutes an example where common belief of proper consistency is consistent with vectors of most preferred strategies that do not lead to the backward induction outcome.

6. Induction in a Trade Game

The games of Figs. 1 and 2 have in common that the properly rationalizable strategies coincide with those surviving iterated (maximal) elimination of weakly dominated strategies. The present section shows that this conclusion does not hold in general. Rather, it will be argued that the concept of proper rationalizability can be used to test the robustness of iterated (maximal) elimination of weakly dominated strategies and other inductive procedures.

Figure 3 illustrates a simplified version of a trade game introduced by Sonsino et al. [25] for the purpose of experimental study; Søvik [26] has subsequently repeated their experiment in alternative designs. The two players consider to trade a good and have a common and uniform prior over the state of the good. There are no gains from trade. If the state of the good is a, then 1 looses 9 and 2 wins 9 if trade occurs. If the state of the good is b, then 1 wins 6 and 2 looses 6 if trade occurs. Finally, if the state of the good is c, then 1 looses 3 and 2 wins 3 if trade occurs. Player 1 is informed of whether the state of the good is equal to a on the one hand, or

	a	b	c
Player 1	-9	6	-3
Player 2	9	-6	3
	1/3	1/3	1/3

FIGURE 3. A trade game

	yy	yn	ny	nn
YY	-2, 2	-1, 1	-1, 1	0, 0
YN	-3, 3	-3, 3	0, 0	0, 0
NY	1, -1	2, -2	-1, 1	0, 0
NN	0, 0	0, 0	0, 0	0, 0

FIGURE 4. The strategic form of the trade game

in the set $\{b, c\}$ on the other. Player 2 is informed of whether the state of the good is in the set $\{a, b\}$ on the one hand, or equal to c on the other.

As a function of their information, each player can announce to accept trade or not. For player 1 the strategy YN means to accept trade if informed of a and not to accept trade if informed of $\{b, c\}$, etc. For player 2 the strategy yn means to accept trade if informed of $\{a, b\}$ and not to accept trade if informed of c, etc. Trade occurs if and only if both players have accepted trade. This yields the strategic game of Figure 4.

6.1. **An Inductive Procedure.** If player 2 naively believes that player 1 is equally likely to accept trade when informed of a as when informed of $\{b, c\}$, then 2 will wish to accept trade when informed of $\{a, b\}$. However, the following, seemingly intuitive, inductive procedure appears to indicate that 2 should never accept trade if informed of $\{a, b\}$: Player 1 should not accept trade when informed of a since he cannot win by doing so. This eliminates his strategies YY and YN. Player 2, realizing this, should never accept trade when informed of $\{a, b\}$, since — as long as 1 never accepts trade when informed of a — she cannot win by doing so. This eliminates her strategies yy and yn. This in turn means that player 1, realizing this, should never accept trade when informed of $\{b, c\}$, since — as long as 2 never accepts trade when informed of $\{a, b\}$ — he cannot win by doing so. This eliminates his strategy NY. This inductive argument corresponds to iterated (maximal) elimination of weakly dominated strategies, except that the latter procedure eliminates 2's strategies yn and nn in the first round. The argument seems to imply that player 2 should never accept trade if informed of $\{a, b\}$ and that player 1 should never accept trade if informed of $\{b, c\}$. Is this a robust conclusion?

6.2. **Proper Rationalizability in the Trade Game.** The strategic game of Fig. 4 has a set of Nash equilibria that includes the pure strategy vectors

(NN, ny) and (NN, nn), and a set of perfect equilibria that includes the pure strategy vector (NN, ny). However, there is a unique proper equilibrium where player 1 plays NN with probability one, and where player 2 mixes between yy with probability 1/5 and ny with probability 4/5. It is instructive to see why the pure strategy vector (NN, ny) is *not* a proper equilibrium. If 1 assigns probability one to 2 playing ny, then he prefers YN to NY (since the more serious mistake to avoid is to accept trade when being informed of $\{b, c\}$). However, if 2 respects 1's preferences and certainly believes that 1 prefers YN to NY, then she will herself prefer yy to ny, undermining (NN, ny) as a proper equilibrium. The mixture between yy and ny in the proper equilibrium is constructed so that 1 is indifferent between YN and NY.

From Prop. 1 it follows that both yy and yn are properly rationalizable strategies for 2. Moreover, if 1 certainly believes that 2 is of a type with only yy as a most preferred strategy, then NY is a most preferred strategy for 1, implying that NY in addition to NN is a properly rationalizable strategy for 1. That these strategies are in fact properly rationalizable is verified by the belief system of Table 1. In the table the preferences of each type t_i of any player i are represented by a vNM utility function $v_i^{t_i}$ satisfying $v_i^{t_i} \circ z = u_i$ and a 4-level LPS on $S_j \times \{t_j', t_j''\}$, with the first numbers in the parantheses expressing primary probability distributions, the second numbers expressing secondary probability distributions, etc. With $\Omega = S \times \{t_1', t_1''\} \times \{t_2', t_2''\}$, it follows that $\Omega = A^{prop} = CKA^{prop}$. Since each type's preferences over his/her own strategies are given by

$$
\begin{aligned}
t_1' &: NN \succ YN \succ NY \succ YY \\
t_1'' &: NY \succ NN \succ YY \succ YN \\
t_2' &: ny \succ nn \succ yy \succ yn \\
t_2'' &: yy \succ yn \succ ny \succ nn,
\end{aligned}
$$

it follows that NY and NN are properly rationalizable for player 1 and yy and ny are properly rationalizable for player 2. Note that YY and YN for player 1 and yn and nn for player 2 cannot be properly rationalizable since these strategies are weakly dominated and, thus, cannot be most preferred strategies for cautious players.

The lesson to be learned from this analysis, is that is not obvious that deductive reasoning should lead players to refrain from accepting trade in the trade game.[7] By comparison to Prop. 3, the analysis can be used to support the argument that backward induction in generic perfect information games is more convincing than the inductive procedure for the trade game discussed in Sect. 6.1.

[7] The experiments by Sonsino et al. [25] and Søvik [26] show that some subjects do in fact accept trade in a slightly more complicated version of this game.

Table 1. A belief system for the trade game

t_1'	t_2'	t_2''
yy	$(0,0,1,0)$	$(0,0,0,0)$
yn	$(0,0,0,1)$	$(0,0,0,0)$
ny	$(1,0,0,0)$	$(0,0,0,0)$
nn	$(0,1,0,0)$	$(0,0,0,0)$

t_1''	t_2'	t_2''
yy	$(0,0,0,0)$	$(1,0,0,0)$
yn	$(0,0,0,0)$	$(0,1,0,0)$
ny	$(0,0,0,0)$	$(0,0,1,0)$
nn	$(0,0,0,0)$	$(0,0,0,1)$

t_2'	t_1'	t_1''
YY	$(0,0,0,0)$	$(0,0,1,0)$
YN	$(0,0,0,0)$	$(0,0,0,1)$
NY	$(0,0,0,0)$	$(1,0,0,0)$
NN	$(0,0,0,0)$	$(0,1,0,0)$

t_2''	t_1'	t_1''
YY	$(0,0,0,1)$	$(0,0,0,0)$
YN	$(0,1,0,0)$	$(0,0,0,0)$
NY	$(0,0,1,0)$	$(0,0,0,0)$
NN	$(1,0,0,0)$	$(0,0,0,0)$

7. Conclusion

Iterated (maximal) elimination of weakly dominated strategies, backward induction in perfect information games, and other inductive procedures have been subject to critical scrutiny during the last few years. The present paper shows how proper rationalizability — based on the imposition of common certain belief of the proper consistency of preferences with the game and the preferences of the opponent — can be used to test the robustness of such procedures.

It has been shown how proper rationalizability in general supports backward induction in generic perfect information games with 2 players. However, in other games — like the trade game of Sect. 6 — the concept of proper rationalizability points to fundamental reasons why deductive reasoning may *not* coincide with the outcome of iterated (maximal) elimination of weakly dominated strategies.

Appendix. The Decision-Theoretic Framework

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

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

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

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

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

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

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

The following representation result due to Blume et al. ([8], Theorem 3.1) can now be stated. It requires the notion of a *lexicographic probability system* (LPS) which consists of L levels of subjective probability distributions: If $L \geq 1$ and, $\forall \ell \in \{1, \dots, L\}$, $\mu_\ell \in \Delta(F)$, then $\lambda = (\mu_1, ..., \mu_L)$ is an LPS on F. Let $\mathbf{L}\Delta(F)$ denote the set of LPSs on F.

Proposition A1. *If $\succeq_F$ is complete and transitive, and satisfies objective independence, nontriviality, conditional continuity, and non-null state independence, then there exists a vNM utility function $\upsilon : Z \to \mathbb{R}$ and an LPS $\lambda = (\mu_1, ..., \mu_L) \in \mathbf{L}\Delta(F)$ such that $\mathbf{x}_F \succeq_F \mathbf{y}_F$ iff* [8]

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

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

References

1. Anscombe, F.J., Aumann, R.: A definition of subjective probability. Annals of Mathematical Statistics **34** (1963) 199–205.
2. Armbruster, W., Böge, W.: Bayesian game theory. In "Game Theory and Related Topics" (Moeschlin, Pallaschke, Eds.), North-Holland, Amsterdam (1979).
3. Asheim, G.B.: On the epistemic foundation for backward induction. University of Oslo (1999).
4. Asheim, G.B., Dufwenberg, M.: Admissibility and common belief. University of Oslo (1999).
5. Aumann, R.: Backward induction and common knowledge of rationality. Games Econ. Beh. **8** (1995) 6–19.
6. Battigalli, P., Siniscalchi, M: Hierarchies of conditional belief and interactive epistemology in dynamic games. European University Institute working paper ECO No. 98/29.

[8] For two vectors v and w, $v \geq_L w$ iff whenever $w_\ell > v_\ell$, there exists $\ell' < \ell$ such that $v_{\ell'} > w_{\ell'}$.

7. Ben-Porath, E.: Rationality, Nash equilibrium, and backwards induction in perfect information games. Rev. Econ. Stud. **64** (1997) 23–46.
8. Blume, L., Brandenburger, A., Dekel, E.: Lexicographic probabilities and choice under uncertainty. Econometrica **59** (1991) 61–79.
9. Blume, L., Brandenburger, A., Dekel, E.: Lexicographic probabilities and equilibrium refinements. Econometrica **59** (1991) 81–98.
10. Böge, W., Eisele, T.: On the solutions of Bayesian games. Int. J. Game Theory **8** (1979) 193–215.
11. Börgers, T.: Weak dominance and approximate common knowledge. J. Econ. Theory **64** (1994) 265–276.
12. Brandenburger, A.: Lexicographic probabilities and iterated admissibility. In "Economic Analysis of Markets and Games" (Dasgupta, Gale, Hart, Maskin, Eds.), MIT Press, Cambridge, MA (1992).
13. Brandenburger, A., Dekel, E.: Hierarchies of beliefs and common knowledge. J. Econ. Theory **59** (1993) 189–198.
14. Dekel, E., Fudenberg, F.: Rational behavior with payoff uncertainty. J. Econ. Theory **52** (1990) 243–267.
15. Dekel, E., Gul, F.: Rationality and knowledge in game theory. In "Advances in Economics and Econometrics: Theory and Applications" (Kreps, Wallis, Eds.), Cambridge University Press, Cambridge, UK (1997).
16. Epstein, L.G.: Preference, rationalizability and equilibrium. J. Econ. Theory **73** (1997) 1–29.
17. Epstein, L.G., Wang, T.: "Beliefs about beliefs" without probabilities. Econometrica **64** (1996) 1343–1373.
18. Herings, P.J.-J., Vannetelbosch, V.J.: Refinments of rationalizability for normal-form games. Int. J. Game Theory **28** (1999) 53–68.
19. Mailath, G., Samuelson, L., Swinkels, J.: Extensive form reasoning in normal form games. Econometrica **61** (1993) 273–302.
20. Mertens, J.-M., Zamir, S.: Formulation of Bayesian analysis for games of incomplete information. Int. J. Game Theory **14** (1985) 1–29.
21. Morris, S.: Alternative notions of belief. In "Epistemic Logic and the Theory of Decisions in Games" (Bacharach, Gerard-Varet, Mongin, Shin, Eds.), Kluwer, Dordrecht (1997).
22. Myerson, R.: Refinement of the Nash equilibrium concept. International Journal of Game Theory **7** (1978) 73–80.
23. Rubinstein, A.: Comments on the interpretation of game theory. Econometrica **59** (1991) 909–924.
24. Schuhmacher, F.: Proper rationalizability and backward induction. Forthcoming in Int. J. Game Theory.
25. Sonsino, D., Erev, I., Gilat, S.: On rationality, learning and zero-sum betting – An experimental study of the no-betting conjecture. Technion (1999).
26. Søvik, Y.: Impossible bets: An experimental study. University of Oslo (1999).
27. Stahl, D.: Lexicographic rationality, common knowledge, and iterated admissibility. Econ. Letters **47** (1995) 155–159.
28. Stalnaker, R.: Belief revision in games: forward and backward induction. Math. Soc. Sci. **36** (1998) 57–68.
29. Tan, T., Werlang, S.R.C.: The Bayesian foundations of solution concepts of games. J. Econ. Theory **45** (1988) 370–391.