GAMES AND ECONOMIC BEHAVIOR **23,** 54–74 (1998)
ARTICLE NO. GA970602

Imperfectly Observable Commitments in *n*-Player Games*

Werner Güth

Humboldt-Universität zu Berlin, Institut für Wirtschaftstheorie III, Spandauer Str. 1, D-10178 Berlin, Germany

Georg Kirchsteiger†

CentER, Tilburg University, Warandelaan 2, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

and

Klaus Ritzberger‡

Institute for Advanced Studies, Department of Economics, Stumpergasse 56, A-1060 Vienna, Austria

Received January 11, 1996

In a two-stage game where followers can observe moves by leaders only with noise, pure subgame perfect Nash equilibria of the limiting game without noise may not survive arbitrarily small noise. Still, for almost all games, there exists a subgame perfect equilibrium outcome of the game with no noise that is approximated by equilibrium outcomes of games with small noise. This, however, depends crucially on generic payoffs and does not necessarily hold for all subgame perfect equilibria of the game without noise. *Journal of Economic Literature* Classification Number: C72.

Key Words: commitments; imperfect observability; subgame perfection.

1. INTRODUCTION

In a recent paper Bagwell (1995) has challenged the validity of applying backwards induction to perfect information two-stage games. Consider

* Helpful comments by two anonymous referees and an associate editor are gratefully acknowledged. Remaining errors and shortcomings are ours.
† E-mail: g.kirchsteiger@kub.nl.
‡ E-mail: ritzbe@ihs.ac.at.

0899-8256/98 $25.00

any (finite) two-player game with generic payoffs, where one player, the "leader," moves first, and then, after having observed a noisy signal about the leader's choice, the "follower" moves. Bagwell shows that in such games the pure strategy Nash equilibrium outcomes coincide with those of the associated "one-shot" game in which both players move simultaneously. This has been interpreted as showing that any precommitment effect is eliminated when there is a slight amount of noise associated with the observation of the leader's choice.[1]

But subsequent analysis by van Damme and Hurkens (1997) indicates that it is the focus on pure strategy equilibria which seems to eliminate the precommitment effect. Call an outcome *accessible* if it is induced by a subgame perfect Nash equilibrium in the game with perfect observability and every game with sufficiently small noise in the observations has a (possibly mixed) equilibrium inducing an outcome close to the accessible one. Van Damme and Hurkens show that for any generic two-player game the (unique) subgame perfect equilibrium outcome of the game without noise is accessible. They also show that the approximating mixed equilibria in the games with noise get selected by an equilibrium selection theory in the spirit of Harsanyi and Selten (1988).

The analysis conducted so far is restricted to two-player games with generic payoffs. Therefore, the issue arises whether the results depend on these restrictions. The present paper addresses this in a general setup of n-player games with arbitrary payoffs. Any such "one-shot" game can be transformed into an extensive form game by splitting the player set into two non-empty subsets, the "leaders" and the "followers," such that leaders move first, followers observe a possibly noisy signal about the leaders' strategy combination, and finally make their choices simultaneously. Therefore, in this framework even the game with perfect observability may have several subgame perfect equilibria. Payoffs need not necessarily be generic, so extensive forms underlying each of the two interactions among leaders and among followers, respectively, are allowed for.[2]

We show that in this general setup only one part of Bagwell's (1995) observation continues to hold. In particular, equilibrium outcomes of the "one-shot" game in which leaders play pure remain as equilibrium out-

[1]Such an interpretation, of course, depends on the validity of noise in the signals that the follower receives. Consider, for instance, a monopolist setting a price for a number of objects which consumers may buy at the quoted price. If there is noise in the observations of the price, then some consumers may find themselves accepting a bargain that the monopolist has never offered. So, whether or not the introduction of noise into the observations of the leader's choice is a valid procedure, depends on the situation which is modeled.

[2]What is not allowed for is that one player moves at both stages, i.e., one of the leaders and one of the followers are agents of the same original player. Under the informational assumptions of the present model this would violate perfect recall.

comes in the games with noisy signals. But the converse is not necessarily true. Equilibria of the games with imperfect observability at which leaders play pure need not always induce equilibrium outcomes of the "one-shot" game. The result by van Damme and Hurkens (1997), on the other hand, holds true for all games with generic payoffs. For any partition of the player set into leaders and followers and for almost all "one-shot" games there exists an accessible outcome.

By means of examples we also show that these results are binding in the following sense: For *degenerate* payoffs there exist two-player games which have *no* accessible outcome at all. Moreover, there are games with *degenerate* payoffs and with more than two players for which even a *set-valued* generalization of accessibility fails to exist. Finally, games with several followers and with *generic* payoffs may have subgame perfect equilibrium outcomes which are *not* accessible.

The existence of an accessible outcome for almost all games is driven by the presence of a Nash equilibrium outcome which is robust against all payoff perturbations. Such robustness is a stronger criterion than *strategic stability* (Kohlberg and Mertens, 1986) which requires only robustness against strategy perturbations. But the latter may not span the payoffs generated by noisy signals.

The paper is organized as follows: Section 2 contains notation and the description of the model. Section 3 explores the survival of "one-shot" equilibrium outcomes when signals are noisy and leaders play pure strategies. Section 4 demonstrates the existence of accessible outcomes for almost all games and contains the negative examples described above. Section 5 concludes.

2. THE MODEL

Let $\Gamma = ((S_i)_{i\in N}, (u_i)_{i\in N})$ be a finite *normal form game* with players $i \in N = \{1, \ldots, n\}$, (finite) pure strategy sets S_i, $S = \times_{i\in N} S_i$, and payoff functions u_i: $S \to \Re$. Mixed strategy sets are denoted by $\Delta(S_i)$, the space of mixed strategy combinations by $\Theta(S) = \times_{i\in N}\Delta(S_i)$, and payoff functions for the mixed extension U_i: $\Theta(S) \to \Re$ are defined as usual, $\forall\, i \in N$. The game Γ will be referred to as the one-shot game.

Now partition the player set N into two non-empty sets, the set of "leaders" I and the set of "followers" II, denote $S^I = \times_{i\in I} S_i$ and $S^{II} = \times_{i\in II} S_i$, and define the *extensive form game* G as follows: First all players $i \in I$ choose simultaneously their pure strategies $s_i \in S_i$. Then all players $i \in II$ get to see $s^I \in S^I$ and choose simultaneously their pure strategies $s_i \in S_i$, $\forall\, i \in II$. Payoffs at terminal nodes are as in Γ. The normal form game that corresponds to G is denoted by Γ_G.

Next, introduce (possibly) noisy signals about the leaders' strategy combination. For each follower $i \in II$ let T_i be a set of *signals* with the same number of elements as S^I, $|T_i| = |S^I|$, $\forall\, i \in II$. Denote by T the product $T = \times_{i\in II} T_i$ and let Λ be the set of all conditional probability distributions $\lambda = (\lambda(t \mid s))_{(t,\,s)\in T\times S^I}$ over vectors of signals of followers given a strategy combination of leaders. The set Λ will be the parameter set for the games defined below.

Associated with any $\lambda \in \Lambda$ there is an *extensive form* game $G(\lambda)$ defined as follows: First leaders $i \in I$ choose simultaneously their strategies $s_i \in S_i$, $\forall\, i \in I$. Then, given $s^I \in S^I$, a chance move selects a vector of signals $t \in T$ according to the probability distribution $\lambda(t \mid s^I)$, $\forall\, t \in T$. Finally, each follower $i \in II$ gets to see her signal $t_i \in T_i$ and chooses, simultaneously with other followers, her strategy $s_i \in S_i$, $\forall\, i \in II$, in response to her signal. Payoffs V_i^λ in $G(\lambda)$ are derived from the payoffs of the one-shot game.

Formally, for any $(\sigma^I, \sigma^{II}) \in \Theta(S)$ denote $\sigma^I(s^I) = \prod_{i\in I} \sigma_i^I(s_i^I)$, for all $s^I = (s_i^I)_{i\in I} \in S^I$, and $\sigma^{II}(s^{II}) = \prod_{i\in \mathrm{II}} \sigma_i^{II}(s_i^{II})$, for all $s^{II} = (s_i^{II})_{i\in II} \in S^{II}$. Given some vector of signals $t = (t_i)_{i\in II} \in T$ a *behavior strategy* combination of followers $b = (b_i)_{i\in II} = ((b_i(\tau_i))_{\tau_i\in T_i})_{i\in II}$ evaluated at the signal vector, $b(t) = (b_i(t_i))_{i\in II} = ((b_i(s_i \mid t_i))_{s_i\in S_i})_{i\in II}$, induces a probability distribution on S^{II} by $b(s^{II} \mid t) = \prod_{i\in II} b_i(s_i^{II} \mid t_i)$, $\forall\,(s^{II}, t) \in S^{II} \times T$, where $s^{II} = (s_i^{II})_{i\in II}$. A typical (behavior) strategy combination for all players will be denoted by $\sigma = (\sigma^I, b) \in \Theta$, where Θ is the space of all behavior strategy combinations. With these definitions the payoff function for player $i \in N$ in $G(\lambda)$ is given by

$$V_i^\lambda(\sigma^I, b) = \sum_{s^I\in S^I} \sigma^I(s^I) \sum_{t\in T} \lambda(t \mid s^I) \sum_{s^{II}\in S^{II}} b(s^{II} \mid t)\, u_i(s^I, s^{II}). \tag{1}$$

To define outcomes induced by strategy combinations, let $\Delta(S)$ be the set of probability distributions on S and define ϕ: $\Theta(S) \to \Delta(S)$ by

$$\phi(\sigma) = \left\{\varphi \in \Delta(S) \mid \varphi(s) = \sigma^I(s^I)\, \sigma^{II}(s^{II}),\ \forall\, s = (s^I, s^{II}) \in S\right\}, \tag{2}$$

which gives the *outcome* induced by $\sigma \in \Theta(S)$ in the one-shot game Γ. For each $\lambda \in \Lambda$ the function ϕ_λ: $\Theta \to \Delta(S)$ defined by

$$\phi_\lambda(\sigma) = \left\{\varphi \in \Delta(S) \mid\ \varphi(s) = \sigma^I(s^I) \sum_{t\in T} \lambda(t \mid s^I)\, b(s^{II} \mid t), \quad \forall\, s = (s^I, s^{II}) \in S\right\} \tag{3}$$

gives the *outcome* induced by $\sigma \in \Theta$ and $\lambda \in \Lambda$ in the game $G(\lambda)$. The inverse of ϕ from Eq. (2) gives the *marginals* corresponding to an outcome by

$$\phi^{-1}(\varphi) = \left(\phi_i^{-1}(\varphi)\right)_{i\in N} = \left(\left(\sum_{s_{-i}\in S_{-i}} \varphi(s_{-i}, s_i)\right)_{s_i\in S_i}\right)_{i\in N}, \tag{4}$$

where $S_{-i} = \times_{j\in N\setminus\{i\}} S_j$.

Finally, let us distinguish the following classes of signal distributions. First, define $\mathrm{int}(\Lambda)$ as the set of all $\lambda \in \Lambda$ such that for each follower $i \in II$ one has $\lambda_i(t_i \mid s) := \sum_{t_{-i} \in T_{-i}} \lambda(t_{-i}, t_i \mid s) > 0$, $\forall (t_i, s) \in T_i \times S^I$, where $T_{-i} = \times_{j \in II \setminus \{i\}} T_j$. Signals are *noisy* if $\lambda \in \mathrm{int}(\Lambda)$, i.e., if for each follower every signal always occurs with positive probability.

Second, let $m_\lambda(s) = \arg\max_{t \in T} \lambda(t \mid s)$ denote the most likely vector of signals given $s \in S^I$ and call the signals *perfectly informative* if $\lambda(m_\lambda(s) \mid s) = 1$, $\forall s \in S^I$, and for each follower m_λ projects one-to-one onto the signal space T_i. When signals are perfectly informative, every follower can deduce with certainty from her signal both the leaders' strategy combination *and* the signals of all other followers, because signals are perfectly correlated.

Third, call the signals *completely uninformative* if for each $i \in II$ there is a probability distribution $\hat{\lambda}_i$: $T_i \to \Re_+$, such that $\lambda(t \mid s) = \prod_{i \in II} \hat{\lambda}_i(t_i)$, $\forall t = (t_i)_{i \in II} \in T$, $\forall s \in S^I$. When signals are completely uninformative, the distribution of signals is independent of the leaders' strategy combination *and* signals are uncorrelated across followers.

Since the set of λ's generating perfectly informative signals and the set of λ's generating completely uninformative signals are disjoint compact subsets of Λ, there exists a continuous function d: $\Lambda \to [0, 1]$ such that $d(\lambda) = 0$ if and only if λ is perfectly informative, and $d(\lambda) = 1$ if and only if λ is completely uninformative. Signals are the more *informative* the smaller $d(\lambda)$ is.

Remark. The present model can be modified to accomodate a more explicit treatment of extensive form interaction among leaders. If leaders interact according to a fixed tree several strategy combinations $s^I \in S^I$ may lead to the same outcome (of the interaction among leaders) and followers may only receive signals about those outcomes, rather than about the full strategy combination of leaders. The required modification for such cases may be sketched as follows: The tree induces an equivalence relation $\sim$ on S^I defined by $s \sim s'$ if and only if s and s' induce the same outcome (among leaders), $\forall s, s' \in S^I$. For each follower $i \in II$ let T_i be a set with the same number of elements as the quotient space $Z := S^I / \sim$ and let $T = \times_{i \in II} T_i$. Then λ assigns for each equivalence class $z \in Z$ a probability distribution $(\lambda(t \mid z))_{t \in T}$ on T. Followers again observe their component of the vector $t \in T$ and choose strategies simultaneously. In the payoff functions $\lambda(t \mid [s^I])$ replaces $\lambda(t \mid s^I)$, where $[s^I]$ denotes the equivalence class of $s^I \in S^I$. All other definitions can be adapted accordingly and the analysis would proceed along the same lines as for the present model.

3. ONE-SHOT EQUILIBRIUM OUTCOMES

The first result shows that if λ is completely uninformative, then $G(\lambda)$ is "strategically equivalent" to Γ, and if λ is perfectly informative, then $G(\lambda)$

is "strategically equivalent" to G. Still, of course, the extensive form game $G(\lambda)$ differs from G even in terms of the tree. The number of terminal nodes of G is $|S| = |S^I|\,|S^{II}|$, while the number of terminal nodes of $G(\lambda)$ is $|S^I|^{|II|+1}\,|S^{II}|$, for all $\lambda \in \Lambda$. Moreover, G has $|S^I|$ proper subgames, while $G(\lambda)$ has no proper subgame. But the normal forms of G and $G(\lambda)$ have the same strategy sets, i.e., they differ only by the payoff function.

PROPOSITION 1. (a) *If $\lambda \in \Lambda$ is completely uninformative*, $d(\lambda) = 1$, *then the equilibrium outcomes of $G(\lambda)$ and Γ coincide*.

(b) *If λ is perfectly informative*, $d(\lambda) = 0$, *then the equilibrium outcomes of $G(\lambda)$ and G coincide*.

Proof. (a) Assume $d(\lambda) = 1$. Then, for any $\sigma \in \Theta$, straightforward calculations using Eqs. (3) and (4) show that $\varphi = \phi_\lambda(\sigma)$ implies for any follower $i \in II$ that $\phi_i^{-1}(\varphi) = \sum_{t_i \in T_i} \hat{\lambda}_i(t_i)\, b_i(t_i)$ and

$$\prod_{i \in II}\left[\sum_{t_i \in T_i} \hat{\lambda}_i(t_i)\, b_i(s_i^{II} \mid t_i)\right] = \sum_{t \in T} \lambda(t \mid s)\, b(s^{II} \mid t)$$

for all $s^{II} \in S^{II}$ and all $s \in S^I$. Therefore, $\prod_{i \in N} \phi_i^{-1}(\phi_\lambda(\sigma))(\bar{s}_i) = \phi_\lambda(\sigma)(\bar{s})$, $\forall\, \bar{s} \in S$, and by Eq. (1) one obtains $V_i^\lambda(\sigma) = U_i(\phi^{-1}(\phi_\lambda(\sigma)))$, $\forall\, \sigma \in \Theta$, $\forall\, i \in N$. This implies directly that σ is an equilibrium of $G(\lambda)$ if and only if $\phi^{-1}(\phi_\lambda(\sigma))$ is an equilibrium of Γ.

(b) If $d(\lambda) = 0$, then for each $s \in S^I$ the most likely vector of signals $m_\lambda(s) \in T$ projects one-to-one onto each T_i and occurs with probability one given the leaders' strategy combination. Thus, one can replace the followers' strategy combinations in Γ_G by their compositions with m_λ (a relabelling of strategies), so (the normal form of) $G(\lambda)$ and Γ_G are the same game. ∎

Of course, if $d(\lambda) < 1$ followers may correlate their strategies and, therefore, such a game $G(\lambda)$ need *not* be "strategically equivalent" to Γ (see Example 1 below).

If signals are noisy, then in $G(\lambda)$ all information sets $t_i \in T_i$ are reached with positive probability for all followers $i \in II$. Therefore, any Nash equilibrium of $G(\lambda)$ induces a *sequential* equilibrium (Kreps and Wilson, 1982) of $G(\lambda)$, for any $\lambda \in \text{int}(\Lambda)$. Sequential equilibria of $G(\lambda)$ encompass *beliefs* for all followers. In the present model those are beliefs about the leaders' strategy combination *and* about the signals received by other followers. They are given by

$$\mu_i(s^I, t_{-i} \mid t_i) = \frac{\sigma^I(s^I)\, \lambda(t_{-i}, t_i \mid s^I)}{\sum_{s \in S^I} \sigma^I(s) \sum_{\tau_{-i} \in T_{-i}} \lambda(\tau_{-i}, t_i \mid s)} \tag{5}$$

for all $(s^I, t_{-i}) \in S^I \times T_{-i}$ and for all information sets $t_i \in T_i$, whenever this quantity is defined, and arbitrary otherwise, $\forall\, i \in II$. If signals are noisy, then the quantity in Eq. (5) is always defined.

Beliefs are governed, within the present model, by the signals which followers observe *and* by what leaders (are supposed to) do in equilibrium. At a Nash equilibrium every player behaves optimally *given* the strategy combination of the opponents. The latter information on the opponents' strategy combination represents a "prior" which followers update when they observe their signals. But if the "prior" is concentrated at a pure strategy combination of leaders, updating by Bayes' rule does not change the followers' beliefs about the leaders' strategy combination (though possibly about the signals received by other followers). This observation underlies the following proposition, which is conceptually different from Bagwell's (1995), because it makes explicit the conditions for an equilibrium of $G(\lambda)$ to induce an equilibrium outcome of Γ. Let $E(\Gamma)$ resp. $E(G(\lambda))$ be the set of (possibly mixed) Nash equilibria of Γ resp. $G(\lambda)$.

PROPOSITION 2. *For all noisy signals, i.e.,* $\forall\, \lambda \in \operatorname{int}(\Lambda)$, *there exists a Nash equilibrium* $(s^I, b) \in E(G(\lambda))$ *at which leaders play pure,* $s^I \in S^I$, *and followers ignore signals,* $b(t) = b(t')$, $\forall\, t, t' \in T$, *if and only if there exists a Nash equilibrium* $(s^I, \sigma^{II}) \in E(\Gamma)$ *at which leaders play pure,* $s^I \in S^I$, *with the same outcome,* $\phi_\lambda(s^I, b) = \phi(s^I, \sigma^{II})$.

Proof. Let $(s^I, b) \in E(G(\lambda))$ for some $\lambda \in \operatorname{int}(\Lambda)$ be such that $b(t) = \sigma^{II}$, $\forall\, t \in T$. Then for all leaders $i \in I$ from Eq. (1)

$$\begin{aligned} V_i^\lambda(s_{-i}^I, s_i, b) &= \sum_{t \in T} \lambda(t \mid s_{-i}^I, s_i) \sum_{s^{II} \in S^{II}} b(s^{II} \mid t)\, u_i(s_{-i}^I, s_i, s^{II}) \\ &= \sum_{s^{II} \in S^{II}} \sigma^{II}(s^{II})\, u_i(s_{-i}^I, s_i, s^{II}) = U_i(s_{-i}^I, s_i, \sigma^{II}) \\ &\leq V_i^\lambda(s^I, b) = U_i(s^I, \sigma^{II}) \qquad \forall\, s_i \in S_i. \end{aligned}$$

From Eq. (5) for all $i \in II$ beliefs at information sets $t_i \in T_i$ are given by

$$\mu_i(s^I, t_{-i} \mid t_i) = \frac{\lambda(t_{-i}, t_i \mid s^I)}{\sum_{\tau_{-i} \in T_{-i}} \lambda(\tau_{-i}, t_i \mid s^I)} \qquad \forall\, t_{-i} \in T_{-i}$$

and $\mu_i(s, t_{-i} \mid t_i) = 0$, $\forall\, s \in S^I \setminus \{s^I\}$, implying from the equilibrium property that $b_i(t_i) = \sigma_i^{II}$, $\forall\, t_i \in T_i$, is a best reply against $(s^I, \sigma^{II}) \in \Theta(S)$, $\forall\, i \in II$. Therefore, $(s^I, \sigma^{II}) \in E(\Gamma)$ and $\phi_\lambda(s^I, b) = \phi(s^I, \sigma^{II})$.

Conversely, if $(s^I, \sigma^{II}) \in E(\Gamma)$, then setting $b(t) = \sigma^{II}$, for all $t \in T$, implies $(s^I, b) \in E(G(\lambda))$ and $\phi(s^I, \sigma^{II}) = \phi_\lambda(s^I, b)$, $\forall\, \lambda \in \operatorname{int}(\Lambda)$. ■

From the "if" part of Proposition 2 every outcome of an equilibrium of the one-shot game Γ at which leaders play pure, $s^I \in S^I$, remains an equilibrium outcome of $G(\lambda)$ for all noisy signals. But the converse need, in general, *not* be true. There can be equilibria of $G(\lambda)$, for some $\lambda \in \operatorname{int}(\Lambda)$, at which leaders play pure, $s^I \in S^I$, which do *not* induce an equilibrium outcome of Γ (nor a subgame perfect equilibrium outcome of G). This may happen, because with sufficiently informative signals followers may use their observations to coordinate on different strategy combinations after different signals. This possibility is illustrated by the first example, showing that it is necessary for Proposition 2 for followers to ignore signals. In this example followers' behavior approximates a Nash equilibrium of G which is not subgame perfect.

EXAMPLE 1. Consider the game Γ given in Table I for some $z \in [0, 1]$, where $S_i = \{s_i^1, s_i^2\}$, $\forall i \in N$, and $u = (u_1, u_2, u_3)$. To define G let $I = \{1\}$ and $II = \{2, 3\}$ and let signal spaces be given by $T_i = \{1, 2\}$, $\forall i = 2, 3$.

For this example assume that signals are *conditionally independent* in the sense that $\lambda(t \mid s_1) = \lambda_2(t_2 \mid s_1)\,\lambda_3(t_3 \mid s_1)$, $\forall (t, s_1) \in T \times S_1$. This assumption is adopted here to avoid that equilibria of $G(\lambda)$, for some $\lambda \in \text{int}(\Lambda)$, at which leaders play pure fail to induce one-shot equilibrium outcomes merely because followers receive (conditionally) correlated signals and thus can correlate their strategies.

Abbreviate notation by $\lambda_i(1 \mid s_1^2) = \varepsilon_i$, and $b_i(s_i^1 \mid t) = x_{it}$, $\forall t \in T_i$, for $i = 2, 3$. Assume that there is some $\varepsilon > 0$ such that $\lambda_i(t \mid s_1^{3-t}) < \varepsilon$, $\forall t = 1, 2$, for $i = 2, 3$, i.e., the "wrong" signals occur with probability less than ε.

Now consider the strategy combination $(s_1^2, \bar{x}_2, \bar{x}_3)$ given by $\bar{x}_{i1} = 1$ and

$$\bar{x}_{i2} = \frac{1 - 2\varepsilon_i}{2(1 - \varepsilon_i)} \qquad \forall i \in II = \{2, 3\}.$$

If signals are conditionally independent, payoffs can be written as

$$V_i^\lambda(\sigma) = \sum_{s^I \in S^I} \sigma^I(s^I) \sum_{s^{II} \in S^{II}} \prod_{j \in II} \left[\sum_{t_j \in T_j} \lambda_j(t_j \mid s^I) b_j(s_j^{II} \mid t_j) \right] u_i(s^I, s^{II})$$

for all $i \in N$ and any strategy combination $\sigma \in \Theta$. Therefore, at the strategy combination $(s_1^2, \bar{x}_2, \bar{x}_3)$ one has

$$\begin{aligned} V_2^\lambda(s_1^2, x_2, \bar{x}_3) = {} & [\varepsilon_2\, x_{21} + (1 - \varepsilon_2)x_{22}][2\,\varepsilon_3\,\bar{x}_{31} + 2(1 - \varepsilon_3)\bar{x}_{32} - 1] \\ & + 1 - \varepsilon_3\,\bar{x}_{31} - (1 - \varepsilon_3)\bar{x}_{32} = \tfrac{1}{2} \qquad \forall x_2 \in [0, 1]^2 \end{aligned}$$

TABLE I

$u(s_1^1, \cdot)$	s_3^1	s_3^2	$u(s_1^2, \cdot)$	s_3^1	s_3^2
s_2^1	$(-2, 0, 0)$	$(3z - 1, 0, 1)$	s_2^1	$(2, 1, 0)$	$(-4, 0, 1)$
s_2^2	$(2z, 1, 0)$	$(4 - 5z, 2, 2)$	s_2^2	$(-3, 0, 1)$	$(5, 1, 0)$

and

$$V_3^{\lambda}(s_1^2, \bar{x}_2, x_3) = [\varepsilon_3 x_{31} + (1-\varepsilon_3)x_{32}][1 - 2\,\varepsilon_2\,\bar{x}_{21} - 2(1-\varepsilon_2)\bar{x}_{22}] + \varepsilon_2\,\bar{x}_{21} + (1-\varepsilon_2)\bar{x}_{22} = \tfrac{1}{2} \qquad \forall\, x_3 \in [0,1]^2$$

Since every strategy is a best reply for followers, in particular $\bar{x}_i$ is a best reply against $(s_1^2, \bar{x}_2, \bar{x}_3)$, for all $i \in II$. Moreover, if ε is sufficiently small, then the leader $i = 1$ will obtain a payoff close to -2 from deviating to s_1^1 against $(\bar{x}_2, \bar{x}_3)$, while she obtains precisely 0 from playing s_1^2. Hence, $(s_1^2, \bar{x}_2, \bar{x}_3) \in E(G(\lambda))$, independently of $z \in [0,1]$.

But the outcome $\phi_\lambda(s_1^2, \bar{x}_2, \bar{x}_3) = (s_1^2, \frac{1}{2} \circ s_2^1 + \frac{1}{2} \circ s_2^2, \frac{1}{2} \circ s_3^1 + \frac{1}{2} \circ s_3^2)$ is *never* an equilibrium outcome of Γ, because if both $i = 2$ and $i = 3$ randomize uniformly player $i = 1$ obtains $\frac{1}{4} > 0$ from deviating to s_1^1 in Γ, independently of z. Moreover, if $z < \frac{4}{5}$, then $\phi_\lambda(s_1^2, \bar{x}_2, \bar{x}_3)$ is *not* a subgame perfect equilibrium outcome of G, while for $z > \frac{4}{5}$ it is the *unique* subgame perfect equilibrium outcome of G.

The example not withstanding, there *are* cases where an equilibrium of some $G(\lambda)$, with noisy signals, at which leaders play pure does indeed induce an equilibrium outcome of Γ. One such case implies Bagwell's (1995) result in the following sense. Suppose G has a single follower, $II = \{i\}$, and this single follower has a unique optimal choice in every proper subgame of G. This is, for instance, the case in the two-player games with distinct payoffs studied by Bagwell (1995) and by van Damme and Hurkens (1997). In such a case beliefs concern exclusively the leaders' strategy combination (and not also the signals of other followers). Therefore, if leaders play pure, $s^I \in S^I$, at some equilibrium of $G(\lambda)$, for $\lambda \in \mathrm{int}(\Lambda)$, the follower's beliefs at information set $t_i \in T_i$ are $\mu_i(s^I \mid t_i) = 1$ and $\mu_i(s \mid t_i) = 0$, $\forall\, s \in S^I \setminus \{s^I\}$, for all $t_i \in T_i$. Hence, the follower must play her unique optimal choice from the subgame of G which starts after s^I at *all* of her information sets. Consequently, the follower does not condition on signals and the "only if" part of Proposition 2 implies that the induced outcome is an equilibrium outcome of the one-shot game Γ.

In other cases extra assumptions yield a coincidence of the pure strategy equilibrium outcomes of Γ and some $G(\lambda)$ with noisy signals. Below such a coincidence is demonstrated by combining generic payoffs with conditionally independent but sufficiently informative noisy signals. Recall from Example 1 that signals are *conditionally independent* if $\lambda(t \mid s) = \prod_{i \in II} \lambda_i(t_i \mid s)$, $\forall\,(t, s) \in T \times S^I$. Moreover, say that a statement holds for *almost all* games Γ, or for *generic* Γ, if there is an open dense set with full Lebesgue measure in the space of payoff assignments $u = (u_i)_{i \in N}$ for which the statement is true.

PROPOSITION 3. *For almost all games* Γ *there exists* $\varepsilon > 0$ *such that if signals are noisy and conditionally independent and* $d(\lambda) < \varepsilon$, *then the pure strategy equilibrium outcomes of* $G(\lambda)$ *and* Γ *coincide.*

Proof. It is not difficult to show that for almost all games Γ there exists some threshold $\gamma \in (0,1)$ such that if $\phi(\sigma)(s) > 1-\gamma$, then every follower $i \in II$ has a unique (and, therefore, pure) best reply against $\sigma \in \Theta(S)$, $\forall s \in S$. Fix such a generic game Γ and let $\gamma \in (0,1)$ denote the associated threshold. Since d is continuous, for every $\delta \in (0,1)$ there is $\varepsilon \in (0,1)$ such that if $d(\lambda) < \varepsilon$, then $\lambda(m_\lambda(s) \mid s) > 1-\delta$, $\forall s \in S^I$.

Assume that (noisy) signals are conditionally independent and let $\varepsilon \in (0,1)$ be so small that $d(\lambda) < \varepsilon$ implies $\lambda(m_\lambda(s) \mid s) > (1-\gamma)^{1/|II|}$, $\forall s \in S^I$. Now consider a pure strategy equilibrium of $G(\lambda)$ at which leaders play $s^I \in S^I$, denote $m_\lambda(s^I) = (t_i^*)_{i\in II}$, and let $s_i^{II} \in S_i$ denote the (pure) choice that follower $i \in II$ takes at information set $t_i^* \in T_i$, $\forall i \in II$. A follower's beliefs are from Eq. (5) and conditional independence given by

$$\mu_i(s^I, t_{-i} \mid t_i) = \prod_{j\in II\setminus\{i\}} \lambda_j(t_j \mid s^I)$$

and $\mu_i(s, t_{-i} \mid t_i) = 0$, $\forall s \in S^I \setminus \{s^I\}$, independently of $t_i \in T_i$. Since under conditional independence $\lambda(m_\lambda(s^I) \mid s^I) > (1-\gamma)^{1/|II|}$ implies $\lambda_i(t_i^* \mid s^I) > (1-\gamma)^{1/|II|}$, $\forall i \in II$, it follows that $\mu_i(s^I, t_{-i}^* \mid t_i) > 1-\gamma$, $\forall t_i \in T_i$, $\forall i \in II$. Therefore, at all her information sets follower $i \in II$ expects the strategy combination (s^I, s_{-i}^{II}) of her opponents to occur with probability larger than $1-\gamma$. Consequently, she will choose at all her information sets her unique (pure) best reply against $(s^I, s^{II}) \in S$, which, by the equilibrium property, must be $s_i^{II} \in S_i$. Since this holds for all followers, at the equilibrium $b_i(s_i^{II} \mid t_i) = 1$, for all $t_i \in T_i$ and all $i \in II$, i.e., followers ignore signals. Then the "only if" part of Proposition 2 implies that $(s^I, s^{II}) \in E(\Gamma)$. Conversely, if $(s^I, s^{II}) \in E(\Gamma)$, then the "if" part of Proposition 2 implies that this induces an equilibrium outcome for $G(\lambda)$, for all noisy signals. ∎

Proposition 3 appears as the most direct generalization of Bagwell's (1995) result. It seems to indicate that pure subgame perfect equilibria of G do not survive the introduction of a small amount of noise in signals, unless they are equilibria of the one-shot game Γ. But Proposition 3, like Bagwell's (1995) result, relies heavily on pure strategies.

4. ACCESSIBLE OUTCOMES

Van Damme and Hurkens (1997) have shown that for generic two-player games there is always a (mixed) equilibrium of $G(\lambda)$, for $\lambda \in \text{int}(\Lambda)$, the *outcome* of which converges to the pure subgame perfect equilibrium outcome of G as $d(\lambda)$ converges to zero. Such an approximation by (equilibrium) *outcomes* is the best one can hope for. Already for Bagwell's (1995) original example it can be shown that *in strategy space* no equilibrium of a game $G(\lambda)$ with noisy signals is close to the (unique) subgame perfect Nash equilibrium of G. And this is a robust example.

Even the approximation in outcomes, however, relies on generic payoffs, because if payoffs are degenerate, then the game without noise may have a multiplicity of subgame perfect equilibria. Among those multiple equilibrium outcomes there may be no single outcome which is "accessible," even if there are only two players, as the following example shows.

EXAMPLE 2. Let Γ be given by the 2×2 game in Table II, where again $S_i = \{s_i^1, s_i^2\}$, $\forall\, i \in N = \{1, 2\}$.

Define G by setting $I = \{1\}$ and $II = \{2\}$ and let the signal space be $T = T_2 = \{1, 2\}$. The game G has a whole set of subgame perfect equilibria: The leader randomizes with probability $\sigma_1(s_1^1) = \sigma_1 \in [0, 1]$ on strategy s_1^1 and the follower chooses s_2^1 in the subgame starting after s_1^1 and chooses s_2^2 in the subgame starting after s_1^2. Denoting $b_2(s_2^1 \mid t) = x_t$, $\forall\, t \in T = \{1, 2\}$, and $\lambda(t \mid s_1^j) = \lambda_{tj}$, $\forall\, (t, j) \in \{1, 2\}^2$, the follower's conditional payoff given signal $t \in T$ is proportional to

$$V_2^\lambda(\sigma_1, x_t \mid t) \propto 2\, x_t[\sigma_1(\lambda_{t1} + \lambda_{t2}) - \lambda_{t2}] + 2(1 - \sigma_1)\lambda_{t2} - \sigma_1 \lambda_{t1}$$

(where "$\propto$" means "proportional to"). For sufficiently informative signals, $\lambda_{tj} > 0$, $\forall\, (t, j) \in \{1, 2\}^2$, but $\lambda_{tj} \approx 0$, $\forall\, j \neq t$, $\lambda_{tt} \approx 1$, $\forall\, t \in \{1, 2\}$, the follower's behavior in equilibrium thus satisfies

$$(x_1, x_2) = \begin{cases} (0, 0) & \text{if } \sigma_1 < \lambda_{12}/(\lambda_{11} + \lambda_{12}) \\ (1, 0) & \text{if } \lambda_{12}/(\lambda_{11} + \lambda_{12}) < \sigma_1 < \lambda_{22}/(\lambda_{21} + \lambda_{22}) \\ (1, 1) & \text{if } \lambda_{22}/(\lambda_{21} + \lambda_{22}) < \sigma_1. \end{cases}$$

Denoting $X_j = \sum_{t \in T} \lambda_{tj}\, x_t$, $\forall\, j \in T$, the leader's payoff can be written as

$$V_1^\lambda(\sigma) = \sigma_1[1 - X_1 - X_2] + X_2.$$

Thus $\sigma_1 = 1$ implies $x_1 = x_2 = 1$, which in turn implies that $\sigma_1 = 0$ is the leader's unique best choice; similarly, $\sigma_1 = 0$ implies $x_1 = x_2 = 0$, which in turn implies that $\sigma_1 = 1$ is the leader's unique best choice. Therefore, in

TABLE II

(u_1, u_2)	s_2^1	s_2^2
s_1^1	$(0, 1)$	$(1, -1)$
s_1^2	$(1, 0)$	$(0, 2)$

equilibrium $0 < \sigma_1 < 1$ implies $X_1 + X_2 = 1$. If now $\lambda_{11} + \lambda_{12} > 1$, then in equilibrium $\sigma_1 = \lambda_{12}/(\lambda_{11} + \lambda_{12}) \approx 0$. If, on the other hand, $\lambda_{11} + \lambda_{12} < 1$, then in equilibrium $\sigma_1 = \lambda_{22}/(\lambda_{21} + \lambda_{22}) \approx 1$. Note that the payoffs to the follower are different in the two cases and that in both cases the corresponding equilibria of $G(\lambda)$ are unique. Consequently, no *single* equilibrium outcome of G is approximated by the (outcomes induced by) mixed equilibria of the games with informative but noisy signals.

But it turns out that generic payoffs are the only necessary condition for the result by van Damme and Hurkens (1997) to hold. While their proof for the existence of an "accessible" outcome relies on a single follower and on inequalities describing the (single) leader's optimal choice, the generalization below uses well-known results from game theory to establish an analogous conclusion for the general n-player case.

Formally, call an outcome $\varphi \in \Delta(S)$ *accessible* if

(i) there exists a subgame perfect (Selten, 1965) Nash equilibrium of G which induces $\varphi \in \Delta(S)$

and

(ii) for any sequence $\{\lambda_r\}_{r=1}^{\infty}$ which satisfies $\lambda_r \in \text{int}(\Lambda)$, $\forall r$, and $d(\lambda_r) \rightarrow_{r\rightarrow\infty} 0$, there exists an associated sequence $\{\sigma^r\}_{r=1}^{\infty}$ with $\sigma^r \in E(G(\lambda_r))$, $\forall r$, such that $\phi_{\lambda_r}(\sigma^r) \rightarrow_{r\rightarrow\infty} \varphi \in \Delta(S)$.

With this definition one can now state:

THEOREM 1. *For all partitions of the player set into leaders and followers and almost all games* Γ *there exists an accessible outcome*.

Proof. First observe that the number of terminal nodes of the extensive form game G equals the number of strategy combinations of the normal form game Γ, $\prod_{i\in N} |S_i|$. Hence, the mapping from Γ to G for a fixed partition of the player set into leaders and followers is one-to-one and onto. Kreps and Wilson (1982) show that for a fixed extensive form there is, in the space of payoff assignments to terminal nodes, an open dense set with full Lebesgue measure such that for all payoffs in this set the number of equilibrium outcomes is finite.[3] It follows that for such *generic* payoffs the outcome is constant across each of the finitely many connected components of the set of Nash equilibria of the associated normal form (Kohlberg and Mertens, 1986, Proposition 1). Hence, for any fixed partition of the player set into leaders and followers, there is a generic set of payoff assignments to G and, therefore, a generic set of normal form games Γ such that the

[3]Note that this set of "generic" games in general differs from the one used in the proof of Proposition 3.

outcome is constant across each of the finitely many connected components of the set of equilibria of Γ_G.

Moreover, every game has an *essential* component[4] (Kohlberg and Mertens, 1986, Proposition 1; Ritzberger, 1994, Theorem 4) which contains a hyperstable set (Kohlberg and Mertens, 1986, p. 1022) by Zorn's lemma (cf. van Damme, 1987, p. 266). Every hyperstable set contains a proper equilibrium (Myerson, 1978; Kohlberg and Mertens, 1986, Proposition 3). Since every proper equilibrium induces a sequential equilibrium in any extensive form corresponding to the given normal form (van Damme, 1984; Kohlberg and Mertens, 1986, p. 1009), among the outcomes associated with an essential component there is at least one that corresponds to a sequential equilibrium of any associated extensive form game.

For perfectly informative signals, $d(\lambda) = 0$, the normal form of $G(\lambda)$ and the normal form Γ_G are identical up to a relabelling of the followers' strategies by Proposition 1(b). Let C be an essential component of Nash equilibria for Γ_G. If Γ is generic in the sense described above, there is a single outcome corresponding to C which is induced by some subgame perfect Nash equilibrium of G, because C contains a proper equilibrium. But since C is essential, every (normal form of a) game $G(\lambda)$ with $\lambda \in \text{int}(\Lambda)$ such that $d(\lambda)$ is sufficiently small has a Nash equilibrium close to C, because d is continuous. Since from Eq. (3) the mapping $(\lambda, \sigma) \mapsto \phi_\lambda(\sigma)$ is continuous, the outcomes induced by those Nash equilibria of (the normal form of) $G(\lambda)$ close to C must be close to the outcome induced by (all equilibria in) C. ■

The technique of this proof relying on set-valued solution concepts may be suggestive of a set-valued generalization of accessibility. After all, in the game from Example 2 there at least exists a closed (and connected) *set* of outcomes that satisfies the definition of accessibility *as a set*.[5] While it seems possible that the existence of such a set-valued generalization of accessibility may hold for all two-player games, the next example shows that this is false in the general case where payoffs may be degenerate.

What happens in the example below is that subgame perfect equilibrium outcomes of G are induced by a (small) part of a connected component of Nash equilibria, the latter being an essential component. The degeneracy of the example rests with the fact that outcomes vary across the essential component. By choosing a particular signal distribution for $G(\lambda)$ one chooses effectively a particular payoff perturbation of Γ_G. The equilibria

[4]Roughly, a connected component of Nash equilibria is *essential* if every nearby game in payoff space has an equilibrium close to the component.

[5]In a suitably modified definition of an *accessible set* of outcomes, of course, the outcomes induced by equilibria of $G(\lambda)$, for $\lambda \in \text{int}(\Lambda)$, need not converge. Only the (Hausdorff) distance to the accessible set would have to converge to zero.

TABLE III

u^{11}	s_4^1	s_4^2	u^{12}	s_4^1	s_4^2
s_3^1	$(0, 0, 0, 0)$	$(0, 0, 0, 6)$	s_3^1	$(1, -5, 0, 0)$	$(1, -5, 0, 0)$
s_3^2	$(0, 0, 0, 0)$	$(0, 0, 6, 0)$	s_3^2	$(1, -5, 0, 0)$	$(1, -5, 0, 0)$
u^{21}	s_4^1	s_4^2	u^{22}	s_4^1	s_4^2
s_3^1	$(-5, 0, 1, 0)$	$(-5, 0, 0, 5)$	s_3^1	$(0, 15, 6, 0)$	$(0, 15, 0, 0)$
s_3^2	$(1, 0, 0, 1)$	$(1, 0, 5, 0)$	s_3^2	$(0, -1, 0, 6)$	$(0, -1, 0, 0)$

of the latter may lie anywhere close to the essential component of Γ_G, not necessarily close to the equilibria that induce subgame perfect equilibrium outcomes of G.

EXAMPLE 3. Let Γ be the four-player game in Table III, where again $S_i = \{s_i^1, s_i^2\}$, $\forall\, i \in N = \{1, 2, 3, 4\}$, $u = (u_1, u_2, u_3, u_4)$, and we abbreviate $u^{kh} = u(s_1^k, s_2^h, \cdot)$, $\forall\, k, h \in \{1, 2\}$.

Define G by letting players 1 and 2 be the leaders, $I = \{1, 2\}$, and players 3 and 4 be the followers, $II = \{3, 4\}$. Since the spirit of this example is negative as far as subgame perfection is concerned, we wish to avoid a failure of accessibility merely because of a lack of correlation between the followers' signals. Therefore, we assume for this example that followers receive perfectly correlated signals, i.e., all followers observe the *same* signal (as they do in G). Formally, $T_3 = T_4 = \{1, 2, 3, 4\}$ and $\text{supp}(\lambda) \subset \text{diag}(T) = \{t \in T \mid t_3 = t_4\}$. To save on notation, write $t \in \{1, 2, 3, 4\}$ for $(t_3, t_4) = (t, t) \in \text{diag}(T)$ and $\lambda(t \mid s)$ for $\lambda(t_3, t_4 \mid s)$, whenever $t_3 = t_4 = t$, $\forall\, s \in S^I$. Of course, signals are still noisy, $\lambda \in \text{int}(\Lambda)$, as long as $\text{supp}(\lambda) = \text{diag}(T)$.

The game G thus defined has a few features of extensive form interaction. In particular, in the subgame starting after (s_1^1, s_2^1) follower $i = 4$ may choose an "outside option" (s_4^1) or give the move to follower $i = 3$. Similarly, in the subgame starting after (s_1^1, s_2^2) no follower gets to move.

The game G has a single connected set of subgame perfect equilibria where player 1 randomizes between her pure strategies with probability between $\frac{3}{4}$ and 1 on s_1^1 and player 2 chooses s_2^1 with certainty, $(\sigma_1(s_1^1), \sigma_2(s_2^1)) \in [\frac{3}{4}, 1] \times \{1\}$. The highest expected payoff that follower $i = 4$ can obtain in any of the subgame perfect equilibria of G is $\frac{5}{24}$. This set is part of a larger connected component of Nash equilibria of G. Denote $b_3(s_3^1 \mid t) = x_t$ and $b_4(s_4^1 \mid t) = y_t$, $\forall\, t \in \{1, 2, 3, 4\}$, $\sigma_1(s_1^1) = \sigma_1$ and $\sigma_2(s_2^1) = \sigma_2$, and $\lambda(t \mid s_1^k, s_2^h) = \lambda_{t,\, 2k+h-2}$, $\forall\, k, h \in \{1, 2\}$. The followers' conditional payoffs given signal t are proportional to

$$V_3^\lambda(\sigma^I, x, y \mid t) \propto x_t\big[6y_t(\sigma_1\sigma_2\lambda_{t1} + (1 - \sigma_1)(\sigma_2\lambda_{t3} + (1 - \sigma_2)\lambda_{t4})) - \sigma_2(6\sigma_1\lambda_{t1} + 5(1 - \sigma_1)\lambda_{t3})\big] + (1 - y_t)\sigma_2(6\sigma_1\lambda_{t1} + 5(1 - \sigma_1)\lambda_{t3})$$

$$V_4^\lambda(\sigma^I, x, y \mid t) \propto y_t\big[(1-\sigma_1)(\sigma_2\,\lambda_{t3} + 6(1-\sigma_2)\lambda_{t4}) \\ - 6\,x_t(\sigma_1\,\sigma_2\,\lambda_{t1} + (1-\sigma_1)(\sigma_2\,\lambda_{t3} + (1-\sigma_2)\lambda_{t4}))\big] \\ + x_t\,\sigma_2(6\,\sigma_1\,\lambda_{t1} + 5(1-\sigma_1)\lambda_{t3}).$$

Inspection of these conditional payoffs reveals that against any strategy combination of the leaders which happens to satisfy $(\sigma_1, \sigma_2) \in (0,1]^2$, except if $\sigma_1 = 1$, followers play a game with a unique completely mixed equilibrium. If $\sigma_1 = 1$ and $\sigma_2 > 0$, then $x_t = 0$ and $y_t \in [0,1]$ holds in equilibrium. In any case $(\sigma_1, \sigma_2) \in (0,1]^2$, therefore, implies in equilibrium

$$x_t = \frac{(1-\sigma_1)[\sigma_2\,\lambda_{t3} + 6(1-\sigma_2)\lambda_{t4}]}{6[\sigma_1\,\sigma_2\,\lambda_{t1} + (1-\sigma_1)(\sigma_2\,\lambda_{t3} + (1-\sigma_2)\lambda_{t4})]} \qquad \forall\, t = 1, \ldots, 4.$$

Denoting $X_j = \sum_{t\in T} \lambda_{tj}\, x_t$, $\forall\, j = 1, \ldots, 4$, the leaders' payoffs can be written as

$$V_1^\lambda(\sigma) = \sigma_1[1 - 2\,\sigma_2(1 - 3\,X_3)] + \sigma_2(1 - 6\,X_3)$$

$$V_2^\lambda(\sigma) = \sigma_2[4\,\sigma_1(1 + 4\,X_4) + 1 - 16\,X_4] - 1 - 4\,\sigma_1 + 16(1-\sigma_1)X_4.$$

If in equilibrium $\sigma_1 > 0$ and $X_4 \leq \frac{1}{16}$ hold, then $\sigma_2 = 1$ is the only optimal choice for player 2, which would imply $x_t < \frac{1}{6}$, $\forall\, t = 1, \ldots, 4$, from the above explicit formula for x_t; but then $2\,\sigma_2(1 - 3\,X_3) > 2(1 - \frac{3}{6}) = 1$ implies that only $\sigma_1 = 0$ is optimal for player 1 — a contradiction. Similarly, if in equilibrium $X_3 \geq \frac{1}{3}$ holds, then $\sigma_1 = 1$ is player 1's unique best choice, which implies that $\sigma_2 = 1$ is player 2's unique best choice, which in turn implies $x_t = 0$, $\forall\, t = 1, \ldots, 4$ — again a contradiction. Thus in all equilibria with $(\sigma_1, \sigma_2) \in (0,1]^2$ one must have $X_3 < \frac{1}{3}$ and $X_4 > \frac{1}{16}$. But this implies, from the leaders' payoffs, that leaders are also playing a game with a unique completely mixed equilibrium.

Finally, if $\sigma_2 = 0$, then followers in equilibrium must play $x_t = 1$ and $y_t \in [0,1]$, $\forall\, t = 1, \ldots, 4$, implying that $\sigma_1 = 1$ is the unique best choice for player 1 and, therefore, $\sigma_2 = 1$ is player 2's unique best choice, contradicting the assumption $\sigma_2 = 0$. Similarly, $\sigma_1 = 0$ in equilibrium implies $x_t \geq \frac{1}{6}$, $\forall\, t = 1, \ldots, 4$, and, therefore, $X_j \geq \frac{1}{6}$, $\forall\, j = 1, \ldots, 4$, which implies from $1 - 2\,\sigma_2(1 - 3\,X_3) \geq 1 - \sigma_2$ that $\sigma_2 = 1$ if $\sigma_1 = 0$ is an optimal choice; but at $\sigma_1 = 0$ one has that $1 - 16\,X_4 \leq -\frac{5}{3}$ implies $\sigma_2 = 0$, a contradiction. It follows that *all* equilibria of the games $G(\lambda)$ with noisy signals have a completely mixed equilibrium with respect to leaders. All equilibria must, therefore, be solutions to the following system of six equations:

$$6\,x_t[\sigma_1\,\sigma_2\,\lambda_{t1} + (1-\sigma_1)(\sigma_2\,\lambda_{t3} + (1-\sigma_2)\lambda_{t4})] \\ - (1-\sigma_1)[\sigma_2\,\lambda_{t3} + 6(1-\sigma_2)\lambda_{t4}] = 0 \qquad \forall\, t = 1, \ldots, 4$$

$$4\,\sigma_1\left[1+4\sum_{t\in T}\lambda_{t4}\,x_t\right]+1-16\sum_{t\in T}\lambda_{t4}\,x_t=0$$

$$1-2\,\sigma_2\left[1-3\sum_{t\in T}\lambda_{t3}\,x_t\right]=0.$$

The equation for σ_1 implies that in any equilibrium $\sigma_1 \le \frac{3}{4}$. Now consider the signal distribution (for signals in the support of λ)

$$\lambda=(\lambda_{tj})_{(t,\,j)\in\{1,\dots,4\}^2}=\begin{pmatrix}1-7\,\varepsilon & \varepsilon & \varepsilon^2 & \varepsilon\\ \varepsilon & 1-3\,\varepsilon & \varepsilon^2 & \varepsilon\\ \varepsilon & \varepsilon & 1-3\,\varepsilon^2 & \varepsilon\\ 5\,\varepsilon & \varepsilon & \varepsilon^2 & 1-3\,\varepsilon\end{pmatrix}$$

with $\varepsilon > 0$ sufficiently small. With this λ the solutions to the above six equations can be regarded as functions of ε. For the outcome induced by a sequence $\{(\sigma_1(\varepsilon), \sigma_2(\varepsilon), (x_t(\varepsilon))_{t\in T})\}_{\varepsilon\searrow 0}$ to converge to an outcome supported by a subgame perfect equilibrium of G it is necessary that $\sigma_1(0)=\frac{3}{4}$ and $\sigma_2(0)=1$. Clearly,

$$x_3(\varepsilon)=\frac{(1-\sigma_1)[(1-3\,\varepsilon^2)\sigma_2+6\,\varepsilon(1-\sigma_2)]}{6[\varepsilon\,\sigma_1\sigma_2+(1-3\,\varepsilon^2)(1-\sigma_1)\sigma_2+\varepsilon(1-\sigma_1)(1-\sigma_2)]}\longrightarrow_{\varepsilon\searrow 0}\frac{1}{6}$$

holds, where $\sigma_i=\sigma_i(\varepsilon)$, $\forall\, i\in I$. But the limit of $x_4(\varepsilon)$ as $\varepsilon\searrow 0$ is indeterminate. Now observe that the reduced system of (the three) equations for $(x_3, \sigma_1, \sigma_2)$ at fixed values $(x_1, x_2, x_4)=(\bar{x}_1, \bar{x}_2, \bar{x}_4)\in[0,1]^3$ has a non-singular Jacobian matrix with determinant $72(1-\sigma_1(0))(1-4\bar{x}_4)(4x_3(0)-1)=-24(1-\sigma_1(0))(1-4\bar{x}_4)\neq 0$ at $\varepsilon=0$, whenever $\bar{x}_4\neq\frac{1}{4}$, and, therefore, has a non-singular Jacobian matrix at all sufficiently small $\varepsilon>0$ if $\bar{x}_4\neq\frac{1}{4}$. Hence, one can invoke l'Hospital's rule to determine the limit of $x_4(\varepsilon)$ as $\varepsilon\searrow 0$. So again with $\sigma_i=\sigma_i(\varepsilon)$, $\forall\, i\in I$:

$$\begin{aligned}x_4(0)&=\frac{1}{6}\lim_{\varepsilon\searrow 0}\frac{(1-\sigma_1)[\varepsilon^2\,\sigma_2+6(1-3\,\varepsilon)(1-\sigma_2)]}{5\varepsilon\sigma_1\sigma_2+\varepsilon^2(1-\sigma_1)\sigma_2+(1-3\,\varepsilon)(1-\sigma_1)(1-\sigma_2)}\\&=\frac{1-\sigma_1(0)}{6}\,\frac{-6\,\sigma_2'(0)}{5\sigma_1(0)-(1-\sigma_1(0))\sigma_2'(0)}\\&=\frac{(1-\sigma_1(0))\sigma_2'(0)}{(1-\sigma_1(0))\sigma_2'(0)-5\sigma_1(0)}.\end{aligned}$$

Since the derivative at $\varepsilon=0$ of $\sigma_2(\varepsilon)$ is given by $\sigma_2'(0)=6\,x_3'(0)$ and $x_3'(0)=-\sigma_1(0)/[6(1-\sigma_1(0))]$, this yields

$$\sigma_2'(0)=-\frac{\sigma_1(0)}{1-\sigma_1(0)}\quad\Rightarrow\quad x_4(0)=\frac{1}{6}.$$

But then from the equation for $\sigma_1(\varepsilon)$ it follows that

$$\sigma_1(0) = \frac{16\,x_4(0) - 1}{16\,x_4(0) + 4} \quad \Rightarrow \quad \sigma_1(0) = \frac{1}{4}.$$

Since this violates $\sigma_1(0) = \frac{3}{4}$, the only equilibrium of $G(\lambda)$ with the above noise structure does not induce an outcome close to any of the outcomes supported by subgame perfect equilibria of G.

In fact, the unique equilibrium of $G(\lambda)$ with the above noise structure does not even induce *payoffs* close to some subgame perfect equilibrium payoff. As ε approaches zero, the payoff to follower $i = 4$ approaches $\frac{15}{24} = \frac{5}{8}$, which exceeds the highest payoff to follower 4 in any of the subgame perfect equilibria of G.

Finally, we turn to the question whether *all* subgame perfect equilibrium outcomes of G for a *generic* game Γ might be accessible. The example below shows that this is *not* the case, even if there is only a single leader in G. The reason is that not all subgame perfect equilibrium outcomes of games with several followers (and perfect observability) must be induced by (equilibria in) an essential component of the associated normal form game. This shows that Theorem 1 is an existence theorem in a strict sense.

EXAMPLE 4. Consider the three-player game in Table IV, where once again $S_i = \{s_i^1, s_i^2\}$, $\forall\, i \in N$, and $u = (u_1, u_2, u_3)$.

The associated game G, where player 1 is the leader, $I = \{1\}$, and players 2 and 3 are the followers, $II = \{2, 3\}$, is a generic game in the sense of Theorem 1. It has two subgame perfect equilibrium outcomes, (s_1^1, s_2^1, s_3^1) and (s_1^2, s_2^2, s_3^1). (The payoff to the leader from the mixed equilibrium in the subgame after s_1^2 is $-\frac{5}{14} < 0$ and thus cannot be supported by (s_2^1, s_3^1) after s_1^1.)

Let $T_i = \{1, 2\}$, $\forall\, i = 2, 3$, and $T = T_2 \times T_3$. To abbreviate notation let $b_2(s_2^1 \mid t) = x_t$, $b_3(s_3^1 \mid t) = y_t$, $\forall\, t \in \{1, 2\}$, $\sigma = \sigma_1(s_1^1)$, and $\lambda(h, k \mid s_1^j) = \lambda_{hk}^j$, $\forall\,(h, k, j) \in \{1, 2\}^3$. For this example only weak assumptions on the signal distribution are required. First, assume that $\lambda \in \text{int}(\Lambda)$ with $\lambda_{jj}^j \approx 1$, $\forall\, j = 1, 2$, to ensure that $d(\lambda)$ is small. Second, a sufficient condition for the arguments below is $3\lambda_{12}^1 + \lambda_{22}^1 > 2\lambda_{21}^1$. Now define

$$\bar{y}(\sigma, \lambda) = \frac{1}{4} - \frac{3\lambda_{21}^2}{4\lambda_{22}^2} + \frac{\sigma(\lambda_{21}^1 + \lambda_{22}^1)}{2(1-\sigma)\lambda_{22}^2}$$

TABLE IV

$u(s_1^1, \cdot)$	s_3^1	s_3^2	$u(s_1^2, \cdot)$	s_3^1	s_3^2
s_2^1	$(0, 2, 2)$	$(4, -1, 1)$	s_2^1	$\left(\frac{1}{6}, \frac{3}{2}, -\frac{1}{4}\right)$	$\left(-\frac{1}{2}, 0, 0\right)$
s_2^2	$(2, 1, -1)$	$(-2, -2, -2)$	s_2^2	$(1, 3, 3)$	$\left(-1, -\frac{1}{2}, \frac{3}{2}\right)$

and

$$\bar{x}(\sigma,\lambda) = \frac{6}{7} - \frac{\lambda_{12}^2}{7\lambda_{22}^2} + \frac{4\sigma(\lambda_{12}^1+\lambda_{22}^1)}{7(1-\sigma)\lambda_{22}^2}.$$

Clearly, $\bar{y}(\sigma,\lambda) \le 1$ if and only if $\sigma \le z_2(\lambda) := 3(\lambda_{21}^2+\lambda_{22}^2)/[2(\lambda_{21}^1+\lambda_{22}^1)+3(\lambda_{21}^2+\lambda_{22}^2)]$ and $\bar{x}(\sigma,\lambda) \le 1$ if and only if $\sigma \le z_3(\lambda) := (\lambda_{12}^2+\lambda_{22}^2)/[4(\lambda_{12}^1+\lambda_{22}^1)+\lambda_{12}^2+\lambda_{22}^2]$, where $z_2(\lambda) > z_3(\lambda)$ is equivalent to $6(\lambda_{12}^1+\lambda_{22}^1)/(\lambda_{21}^1+\lambda_{22}^1) > (\lambda_{12}^2+\lambda_{22}^2)/(\lambda_{21}^2+\lambda_{22}^2)$. Since the above assumption $3\lambda_{12}^1+\lambda_{22}^1 > 2\lambda_{21}^1$ is equivalent to $6(\lambda_{12}^1+\lambda_{22}^1)/(\lambda_{21}^1+\lambda_{22}^1) > 4$ and $(\lambda_{12}^2+\lambda_{22}^2)/(\lambda_{21}^2+\lambda_{22}^2) \approx 1$, the inequality $z_2(\lambda) > z_3(\lambda)$ holds true under the present assumptions.

The followers' payoffs in $G(\lambda)$ are given by

$$\begin{aligned} V_2^\lambda(\sigma,x,y) = \sum_{h=1}^{2} x_h\Big[&\sigma(\lambda_{h1}^1+\lambda_{h2}^1) \\ &+ (1-\sigma)\big(\tfrac{1}{2}(\lambda_{h1}^2+\lambda_{h2}^2) - 2(\lambda_{h1}^2\, y_1 + \lambda_{h2}^2\, y_2)\big)\Big] \\ + \sigma&\left[3\sum_{k=1}^{2}(\lambda_{1k}^1+\lambda_{2k}^1)y_k - 2\right] \\ + \tfrac{1}{2}(1-\sigma)&\left[7\sum_{k=1}^{2}(\lambda_{1k}^2+\lambda_{2k}^2)y_k - 1\right] \end{aligned}$$

for follower $i = 2$ and for follower $i = 3$ by

$$\begin{aligned} V_3^\lambda(\sigma,x,y) = \sum_{k=1}^{2} y_k\Big[&\sigma(\lambda_{1k}^1+\lambda_{2k}^1) \\ &+ (1-\sigma)\big(\tfrac{3}{2}(\lambda_{1k}^2+\lambda_{2k}^2) - \tfrac{7}{4}(\lambda_{1k}^2\, x_1 + \lambda_{2k}^2\, x_2)\big)\Big] \\ + \sigma&\left[3\sum_{h=1}^{2}(\lambda_{h1}^1+\lambda_{h2}^1)x_h - 2\right] \\ + \tfrac{3}{2}(1-\sigma)&\left[1 - \sum_{h=1}^{2}(\lambda_{h1}^2+\lambda_{h2}^2)x_h\right]. \end{aligned}$$

If (s_1^1, s_2^1, s_3^1) is accessible and $\lambda_{jj}^j \approx 1$, $\forall\, j = 1,2$, then certainly $\sigma > 0$ and $x_1 = y_1 = 1$ must hold for all $\lambda \in \mathrm{int}(\Lambda)$ with $d(\lambda)$ sufficiently small. If $\sigma = 1$, then $x_1 = x_2 = y_1 = y_2 = 1$ implies a payoff of $(1-\sigma)/6$ for the leader which is maximal for $\sigma = 0$. Hence, $0 < \sigma < 1$ must hold.

At $\sigma \in (0,1)$ and $x_1 = y_1 = 1$ the equilibrium values of (x_2, y_2) for $z_2(\lambda) > z_3(\lambda)$ are as follows: If $\sigma > z_2(\lambda)$, then $(x_2, y_2) = (1,1)$;

if $\sigma = z_2(\lambda)$, then $(x_2, y_2) \in [0,1] \times \{1\}$; if $z_2(\lambda) > \sigma > z_3(\lambda)$, then $(x_2, y_2) = (0,1)$; if $\sigma = z_3(\lambda)$, then $(x_2, y_2) = (0,1)$ or $(x_2, y_2) \in \{1\} \times [0, \bar{y}(z_3(\lambda), \lambda)]$; and, finally, if $z_3(\lambda) > \sigma$, then $(x_2, y_2) = (0,1)$, or $(x_2, y_2) = (1,0)$, or $(x_2, y_2) = (\bar{x}(\sigma, \lambda), \bar{y}(\sigma, \lambda))$. In any case, $y_2 < 1$ implies $y_2 \leq \bar{y}(z_3(\lambda), \lambda)$, because $\bar{y}$ is an increasing function of σ. Under the assumption $3\lambda_{12}^1 + \lambda_{22}^1 > 2\lambda_{21}^1$ then $y_2 < 1$ implies

$$y_2 \leq \bar{y}(z_3(\lambda), \lambda) = \frac{1}{4} - \frac{3\lambda_{21}^2}{4\lambda_{22}^2} + \frac{1}{8}\left(\frac{\lambda_{21}^1 + \lambda_{22}^1}{\lambda_{12}^1 + \lambda_{22}^1}\right)\frac{\lambda_{12}^2 + \lambda_{22}^2}{\lambda_{22}^2}$$
$$< \frac{1}{4} - \frac{3\lambda_{21}^2}{4\lambda_{22}^2} + \frac{3(\lambda_{12}^2 + \lambda_{22}^2)}{16\,\lambda_{22}^2} \approx \frac{7}{16} < \frac{1}{2}.$$

The leader's payoff in $G(\lambda)$ is given by

$$V_1^\lambda(\sigma, x, y) = \sigma\Bigg[\sum_{(h,k)\in T} \lambda_{hk}^1(6x_h + 4y_k - 8x_h y_k) - \sum_{(h,k)\in T} \lambda_{hk}^2\left(\tfrac{1}{2}x_h + 2y_k - \tfrac{4}{3}x_h y_k\right) - 1\Bigg] + \sum_{(h,k)\in T} \lambda_{hk}^2\left(\tfrac{1}{2}x_h + 2y_k - \tfrac{4}{3}x_h y_k\right) - 1,$$

taking into account that $\sum_{(h,k)\in T} \lambda_{hk}^j = 1$, $\forall\, j = 1, 2$. As noted earlier, if (s_1^1, s_2^1, s_3^1) is accessible, then $0 < \sigma < 1$, which implies at $x_1 = y_1 = 1$ that

$$y_2 = \left[4\lambda_{12}^1 - 4\lambda_{22}^1 + \frac{2}{3}\lambda_{12}^2 + 2\lambda_{22}^2 - \left(\frac{4}{3}\lambda_{22}^2 - 8\lambda_{22}^1\right)x_2\right]^{-1}$$
$$\times \Bigg[2\lambda_{11}^1 + 6\lambda_{12}^1 + 4\lambda_{21}^1 - \frac{7}{6}\lambda_{11}^2 - \frac{1}{2}\lambda_{12}^2 - 2\lambda_{21}^2 - 1 - \left(2\lambda_{21}^1 - 6\lambda_{22}^1 - \frac{5}{6}\lambda_{21}^2 + \frac{1}{2}\lambda_{22}^2\right)x_2\Bigg]$$
$$\approx \frac{6 - 3x_2}{12 - 8x_2} \in \left[\frac{1}{2}, \frac{3}{4}\right],$$

which is inconsistent with $y_2 \in [0, 7/16] \cup \{1\}$ if λ_{jj}^j is sufficiently close to 1, $\forall\, j = 1, 2$. It follows that for arbitrarily small noise in the signals there is no equilibrium with an outcome close to (s_1^1, s_2^1, s_3^1). In fact the *only* accessible outcome in this example is (s_1^2, s_2^2, s_3^1).

One might ask what the properties of subgame perfect equilibrium outcomes are that make them accessible or not accessible. A partial answer can be deduced from the proof of Theorem 1 and Example 4. The proof of Theorem 1 rests on the existence of an essential component of Nash equilibria for the normal form Γ_G. The accessible outcome in Example 4 is induced by the equilibria in the essential component, while the subgame perfect equilibrium outcome of G in Example 4 which is not accessible is induced by the equilibria in a component which is *not* essential. So accessibility in general seems to be driven by essential components for the normal form Γ_G and not by any extensive form considerations. That the normal form games associated with games $G(\lambda)$ are indeed payoff perturbations of Γ_G can be seen from Eq. (1).

This also suggests why accessibility takes us outside the framework considered by Kohlberg and Mertens (1986). First, for the original definition of *strategically stable sets* it is trivial that such sets for Γ_G may not induce an accessible outcome (unless they coincide with hyperstable sets), because strategically stable sets need not induce sequential equilibrium outcomes for an associated extensive form (see the example by Faruk Gul in Kohlberg and Mertens, 1986, Fig. 11). Second, one could take limits, as noisy signals become more and more informative, of (outcomes induced by) strategically stable sets for (the normal form of) $G(\lambda)$. Van Damme and Hurkens (1997), by employing an *equilibrium selection* theory, provide a successful example of an approach in that spirit which, however, goes beyond strategic stability. For more standard refinement concepts, like strategic stability, even Bagwell's (1995) example serves to show that this procedure may generate *all* Nash equilibrium outcomes of G, not just those that are subgame perfect in G.

Strategically stable sets are set-valued versions of strict perfection (Okada, 1981) and thus rely on *strategy perturbations*. Strategy perturbations indeed correspond to specific payoff perturbations, but are more special than the latter. In particular, strategy perturbations need not correspond to the payoff perturbations generated by introducing noisy signals into G. But it is the robustness against the payoff perturbations generated by noisy signals which underlies accessibility. And strategy perturbations may not cover such payoff perturbations.

5. CONCLUSIONS

This paper has shown that in general n-player games imperfect observability of earlier moves in an extensive form game need not necessarily eliminate a precommitment effect: For almost all games there is an accessible outcome, independently of how the player set is split into leaders and followers. But this conclusion does depend on generic payoffs. Moreover, the main theorem of this paper is strictly an existence result in the sense that not all subgame perfect equilibrium outcomes of the unperturbed game may qualify as accessible. The latter is due to the fact that some subgame perfect equilibrium outcomes of the game with perfect observability may not be robust to the payoff perturbations introduced by noisy signals.

REFERENCES

Bagwell, K. (1995). "Commitment and Observability in Games," *Games Econ. Behav.* **8**, 271–280.

Harsanyi, J. C., and Selten, R. (1988). *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: MIT Press.

Kohlberg, E., and Mertens, J.-F. (1986). "On the Strategic Stability of Equilibria," *Econometrica* **54**, 1003–1037.

Kreps, D. M., and Wilson, R. (1982). "Sequential Equilibrium," *Econometrica* **50**, 863–894.

Myerson, R. B. (1978). "Refinement of the Nash Equilibrium Concept," *Int. J. Game Theory* **7**, 73–80.

Okada, A. (1981). "On Stability of Perfect Equilibrium Points," *Int. J. Game Theory* **10**, 67–73.

Ritzberger, K. (1994). "The Theory of Normal Form Games from the Differentiable Viewpoint," *Int. J. Game Theory* **23**, 207–236.

Selten, R. (1965). "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit," *Zeitschrift für die gesamte Staatswissenschaft* **12**, 301–324, 667–689.

van Damme, E. (1984). "A Relation between Perfect Equilibria in Extensive Form Games and Proper Equilibria in Normal Form Games," *Int. J. Game Theory* **13**, 1–13.

van Damme, E. (1987). *Stability and Perfection of Nash Equilibria*. Berlin/Heidelberg/New York: Springer-Verlag.

van Damme, E., and Hurkens, S. (1997). "Games with Imperfectly Observable Commitment," *Games Econ. Behav.*, **21** 1/2, 282–308.