entER Discussion paper

for
omic Research

CIN01544

Center
for
Economic Research

44

No. 9464

GAMES WITH IMPERFECTLY OBSERVABLE COMMITMENT

by Eric van Damme
and Sjaak Hurkens

August 1994

Games with Imperfectly Observable Commitment

Eric van Damme[*]   Sjaak Hurkens[*]

July 1994

Abstract

In a recent paper, Kyle Bagwell claimed that the power to commit oneself to an action does not confer any strategic benefit if this commitment can only be observed imperfectly. In this paper we show that the validity of this claim depends crucially on the restriction to pure strategy equilibria. Specifically, the game analysed by Bagwell always has a mixed equilibrium that is close to the Stackelberg equilibrium of the game in which the commitment is observed perfectly. We introduce a new theory of equilibrium selection that combines elements from the theory of Harsanyi and Selten (1988) with elements from the theory of Harsanyi (1993). When the noise is sufficiently small, this theory selects the Stackelberg equilibrium.

1   Introduction

One of the most important insights in game theory is that the power to commit oneself may confer a strategic advantage: it may be beneficial to constrain one's own behavior in order to induce others to behave in a way that is favorable to oneself. One possibility to commit oneself is to move early: to preempt the others by choosing and communicating the (irreversible) action that one takes before the rivals take their actions. This idea dates back at least to Von Stackelberg (1934) who demonstrated the existence of a "first-mover advantage" in a quantity-setting duopoly. Schelling's (1960) classic *The Strategy of Conflict* generalized Von Stackelberg's initial insight in several dimensions by describing richer commitment tactics as well as illustrating the ubiquity of the phenomenon that in independent decision situations weakness confers strength, that power may result from the power to bind oneself.

Schelling already pointed out that for a commitment to an action to be credible, the commitment must be irreversible, at least reneging should be sufficiently costly. Schelling also stressed that the efficacy of commitment depends on the communication structure of the game. If the opponent is unavailable for messages, or can destroy all communication channels before any communication takes place, being able to commit oneself is of no value. Hence, commitment can be beneficial only if the communication channel is sufficiently reliable. Just how important this latter requirement is has been shown in a recent paper by Kyle Bagwell (1992). Bagwell shows that a precise communication of the commitment is important, that it is vital that there are no ambiguities, that there are no misunderstandings about the action to which the player committed himself. In fact, Bagwell claims that the first-mover advantage is completely eliminated when there is even a *slight* amount of noise associated with the observation of the first-movers action. Specifically, he shows that, if there is some noise, a pure strategy Nash equilibrium outcome of the game in which one of the players can commit must be a Nash equilibrium outcome of the game in which this commitment possibility is absent. This is a counter-intuitive and striking result and it suggests that a reconsideration of the literature that

applies the idea of a "first-mover advantage" might perhaps be required.

The intuition for Bagwell's result can be easily conveyed. Let $g = (A_1, A_2, u_1, u_2)$ be a 2-person normal form game and consider the sequential move game with player 1 moving first. However, assume that player 2 is only imperfectly informed about this commitment. Specifically, if player 1 commits to $a_1 \in A_1$, player 2 receives the signal $a_1' \in A_1$ with probability $p(a_1' \mid a_1) > 0$ where $p(a_1 \mid a_1) \approx 1$. Hence, player 2 is almost perfectly informed about the commitment. The crucial observation, however, is that if player 1 commits to the pure action a_1^*, the signal that player 2 receives is uninformative. Since all information sets of player 2 are reached with positive probability, Bayes' rule dictates that 2 believes that 1 played a_1^* no matter what signal he receives. In equilibrium, player 2 best responds to a_1^* for all possible messages, hence, if 2's best response to a_1^* in g is unique (say it is a_2^*), then 2 will respond with a_2^* no matter what message he receives. However, then, in order to have an equilibrium in the sequential move game, a_1^* should be a best response against a_2^* in g, hence, (a_1^*, a_2^*) must be an equilibrium of g.

As the above paragraph has shown, Bagwell's result is driven by the specific type of imperfection in the communication technology that he assumes. It is not the case that the commitment sometimes is not communicated, it is rather that the opponent with a small probability receives the wrong message. To put it differently, Bagwell's is a model of errors in perception, rather than errors in communication. His result depends on the assumption that if, for example, a seller commits himself to "I do not sell for a price less than \$100", the buyer might interpret this as "I do not sell for less than \$10,000" or as a commitment to "I give the object away for free". We do not want to enter into the debate about whether this is a sensible assumption, although we believe that this specific assumption might explain why Bagwell's result appears counterintuitive at first. It is, however, important to note that the assumption is crucial for the result. If communication errors would take the form as suggested by Schelling (i.e. commitments would not necessarily be communicated to the second mover, but if they would be communicated, they would be communicated without error), then there would not be a lack of robust-

ness of the type that Bagwell notes. The reader can easily verify that in the latter case, as long as the probability that the commitment is received is sufficiently high, a player will commit himself to his Stackelberg strategy. (See Chakravorti and Spiegel (1993)).

As we do not wish to claim that Schelling's modelling of the errors is necessarily better than Bagwell's, we take Bagwell's claim seriously. However, does the theorem that Bagwell proves justify the claim that he makes? Does the result that the pure equilibrium outcomes of the noisy sequential move game coincide with the pure equilibrium outcomes of the simultaneous move game really allow us to conclude that "with even the slightest degree of imperfection in the observability of the first mover's selection (...) the strategic benefit of commitment is totally lost" (Bagwell (1992))? In our opinion such a conclusion would be premature as it would be based on the assumption that only pure strategy Nash equilibria of a game should be taken into consideration. The restriction to pure equilibria, however, is not compelling and the game theory literature has offered no justification for this restriction so far. In fact, the concept of pure strategy Nash equilibrium suffers from the important and well-known drawback of failing to generate a solution for some games. (Existence might be considered the most fundamental property that a solution concept should satisfy.)

In this paper we take the position that there is no a priori reason to discriminate against equilibria that are not in pure strategies. Consequently, we have to take mixed strategy equilibria into account and this raises the question of which outcomes can be obtained by mixed equilibria of the sequential move game with imperfectly observable commitment. We show that Bagwell's noisy game has a "noisy Stackelberg equilibrium", i.e. a mixed equilibrium that generates an outcome that is close to the Stackelberg outcome and that converges to it as the noise vanishes. Furthermore, we show that there may be other equilibria as well. Hence, Bagwell's game raises the issue of equilibrium selection: If the leader's commitment can only be imperfectly observed, will players coordinate on a pure equilibrium of the simultaneous move game (and, if they do, on which one?) or will they coordinate on the noisy Stackelberg equilibrium? We address this

issue in Section 4. We argue that, starting from an original situation in which there is uncertainty about which strategies will be played, players will reason themselves to the noisy Stackelberg equilibrium. The argument in this section is motivated by elements from the equilibrium selection theories of Harsanyi and Selten (1988) and from Harsanyi (1993), but the theory that we develop is different from each of these. As we show in Section 5, neither the theory of Harsanyi and Selten (1988), nor the theory from Harsanyi (1993) selects the noisy Stackelberg equilibrium in general. The comparison of these various theories gives interesting insights in each of them. Hence, although the main message of this paper is that there is no immediate need to reconsider the literature that applies the idea of a "first-mover advantage", the paper may also be read as an exercise in equilibrium selection.

2 The Noisy Commitment Game

Let g be a (finite) 2-person game in strategic form. Since below we will mainly be interested in what happens when the players move sequentially rather than simultaneously, we label the players as L (for leader) and F (for follower). $\mathcal{I}$ (resp. $\mathcal{J}$) denotes the set of pure strategies of player L (resp. player F) in g and u_{ij} (resp. v_{ij}) denotes this player's payoff when the strategy pair (i, j) is played. We write $\mathcal{I} = \{1, \ldots, I\}$ and $\mathcal{J} = \{1, \ldots, J\}$. Throughout this paper we assume that g satisfies the following regularity condition[1]:

$$\text{if } (i, j) \neq (k, l), \text{ then } u_{ij} \neq u_{kl} \text{ and } v_{ij} \neq v_{kl} \qquad (2.1)$$

This assumption implies that F has a unique best response against each pure strategy i of L. This best response will be denoted by b_i and we write

$$u_i = u_{ib_i}. \qquad (2.2)$$

Without further loss of generality we assume that

$$u_1 > \max_{i \neq 1} u_i. \tag{2.3}$$

Hence, in the sequential move game in which L moves before F and in which F is perfectly informed about the pure action that L has chosen, the unique subgame perfect equilibrium is $(1, b)$ with outcome $(1, b_1)$. (We use b to denote the strategy of F in this game that responds to i with $b_i (i \in \mathcal{I})$.)

We focus our attention on the noisy version of the sequential move game in which F is only imperfectly informed about which action has been chosen by L. To that end, let P be a stochastic matrix defined on the state space $\mathcal{I}$. Hence, $P = (p_{ik})_{i,k \in \mathcal{I}}$ with $p_{ik} \geq 0$ and $\sum_k p_{ik} = 1$ for all i. The interpretation is that F receives the signal "L played k" with a probability p_{ik} in case L actually plays i. Emphasis will be on the situation where the noise, i.e. the probability of receiving the "wrong" signal is small but positive. Writing P^0 for the identity matrix on $\mathcal{I}$ (i.e. $p_{ii}^0 = 1$ for all i) we will measure the absolute level of the noise by the distance between P and P^0 and we will write

$$|P| = \max\{|P_{ik} - P_{ik}^0| : i, k \in \mathcal{I}\}. \tag{2.4}$$

We will restrict ourselves to the case where any signal can result from any action, i.e.

$$p_{ik} > 0 \quad \text{for all } i, k \in \mathcal{I}. \tag{2.5}$$

Formally then, we consider the extensive form game g^P given by the following rules:

1. player L chooses an action $i \in \mathcal{I}$,

2. chance chooses $k \in \mathcal{I}$ with probability p_{ik},

3. player F learns k and chooses $j \in \mathcal{J}$,

4. player L receives the payoff u_{ij} and F receives v_{ij}.

This game g^P is referred to as the noisy commitment game. Note that the messages (the signals that F receives) are payoff irrelevant. We will denote a (behavioral) strategy of player L (resp. F) in g^P by s (resp. f) and we write $\sigma = (s, f)$ for a strategy combination. Hence, s is a probability distribution on $\mathcal{I}, s \in \Delta(\mathcal{I})$, and f is a map that assigns a probability distribution on $\mathcal{J}$ to each element of $\mathcal{I}$, i.e. $f \in \Delta(\mathcal{J})^{\mathcal{I}}$. We let s_i denote the probability that L chooses i while f_{kj} is the probability that F chooses j in response to the message k. We write $f_k = j$ if $f_{kj} = 1$ and use similar conventions throughout the text. The outcome of the strategy pair $\sigma = (s, f)$ in g^P is the probability distribution $z^P = z^P(\sigma)$ that σ induces on $I \times J$. Hence, we have that

$$z^P(\sigma)_{ij} = s_i \sum_{k=1}^{I} p_{ik} f_{kj} \tag{2.6}$$

Note that z^P may involve nontrivial correlation of the players' actions. Player L's (expected) payoff in g^P is written as $u^P(\sigma)$ and F's payoff is denoted by $v^P(\sigma)$, hence

$$u^P(\sigma) = \mathcal{E}(u \mid z^P(\sigma)), \quad v^P(\sigma) = \mathcal{E}(v \mid z^P(\sigma)) \tag{2.7}$$

A pair $\sigma = (s, f)$ is a Nash equilibrium of g^P if s is a best reply against f and f is a best reply against s. Note that because of (2.5) there are no unreached information sets in the (extensive form of the) game g^P, hence, any Nash equilibrium is a sequential equilibrium, and in order for f to be a best response against s, it is necessary that f_k is a best response against the posterior beliefs at k induced by s for every message k. By Bayes' rule, this posterior belief that F associates to $i \in I$ after having received the message k is given by

$$\mu_{ik}^{P,s} = p_{ik} s_i / \sum_{\alpha} p_{\alpha k} s_{\alpha}, \tag{2.8}$$

so that, for all s with $s_k > 0$

$$\lim_{|P| \to 0} \mu_{ik}^{P,s} = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases} \tag{2.9}$$

Hence, if the noise is small and F expects L to choose k with positive probability, then he will attach high probability to the event that L actually played k when he receives the message "k". Assumption (2.1) thus implies that F will respond to k with b_k in this case. Lemma 1 proves a slightly stronger statement.

Lemma 1 . *There exists $\varepsilon^* > 0$ such that for all P with $0 < |P| < \varepsilon^*$, all strategy combinations $\tau = (s, f)$ and all $i \in \mathcal{I}$: If $s_i > \sqrt{|P|}$ and f is a best reply against s in g^P, then $f_i = b_i$.*

Proof. The regularity assumption (2.1) implies that there exists $\delta < 1$ such that for all $i \in \mathcal{I}$: If player F assigns at least probability δ to L playing i in g, then b_i is the unique best response of F in g. Let ε^* be such that $(1 + \sqrt{\varepsilon^*})^{-1} \geq \delta$.

Now, let P be such that $0 < |P| = \varepsilon < \varepsilon^*$ and let $s \in \Delta(\mathcal{I})$ and $i \in \mathcal{I}$ be such that $s_i > \sqrt{\varepsilon}$. Then we obtain from (2.8)

$$
\begin{aligned}
\mu_{ii}^{P,s} &\geq \frac{p_{ii}s_i}{\varepsilon(1 - s_i) + p_{ii}s_i} \\
&= [1 + \varepsilon(1 - s_i)/p_{ii}s_i]^{-1} \\
&\geq (1 + \sqrt{\varepsilon})^{-1} \\
&\geq (1 + \sqrt{\varepsilon^*})^{-1}
\end{aligned}
$$

If f is a best reply against s, then f is necessarily a best reply against the posterior beliefs $\mu_{\cdot i}$ for all i. It, hence, follows from the above inequalities, and the choice of ε^*, that $f_i = b_i$. $\qquad\square$

3 Equilibria in the Noisy Commitment Game

For the sake of completeness we start by stating (and proving) Bagwell's main result.

Proposition 1 *(Bagwell (1992)). The set of pure strategy equilibrium outcomes of g and g^P coincide.*

Proof. Assume (i,j) is a pure strategy Nash equilibrium in g. Then $j = b_i$ and if f is the strategy of F in g^P defined by $f_k = b_i (k \in \mathcal{I})$, then (i, f) is an equilibrium of g^P. It obviously produces the same outcome as (i, j) does. Assume (i, f) is a pure strategy Nash equilibrium in g^P. Since $\mu_{ik}^{P,i} = 1$ for all k, we must have $f_k = b_i$ for all k. Hence, i is a best reply against b_i in g and (i, b_i) is an equilibrium of g with the same outcome as (i, f). $\square$

Proposition 1 gives a sufficient condition for an outcome to be an equilibrium outcome of the game g^P. We now give a necessary condition for the case where the noise is small. Write

$$\mathcal{N} = \{(i, b_i) : \quad u_i \geq \max_k \min_j u_{kj}\} \tag{3.1}$$

for the set of Nash equilibrium outcomes of the game in which player L's commitment is perfectly observed by F. (Note that because of (2.1) any Nash outcome has to be pure.) We have that the Nash equilibrium outcome correspondence of g^P is upper hemi continuous at $P = p^0$.

Proposition 2 *Let z^P be an equilibrium outcome of g^P. If $z = \lim_{|P| \to 0} z^P$ exists, then $z \in \mathcal{N}$.*

Proof. The proof follows from regularity assumption (2.1) and Lemma 1. Let ε^* be as in Lemma 1 and for P with $0 < |P| < \varepsilon^*$, let (s^P, f^P) be an equilibrium of g^P with outcome z^P. Assume the limit outcome z to exist. If $i \neq k, s_i^P > \sqrt{|P|}$ and $s_k^P > \sqrt{|P|}$, then $f_i^P = b_i$ and $f_k^P = b_k$, hence

$$\lim_{|P| \to 0} u^P(i, f^P) = u_i, \quad \lim_{|P| \to 0} u^P(k, f^P) = u_k.$$

But (2.1) implies that $u_i \neq u_k$, hence that $s_i^P s_k^P = 0$ for $|P|$ sufficienty small. The contra-

diction shows that, for $|P|$ sufficiently small there is at most one $i \in \mathcal{I}$ with $s_i^P > \sqrt{|P|}$. Consequently, we have that $z = (i, b_i)$ for this particular value of i. It is obvious that the inequalities in (3.1) must be satisfied. If there would exist $k \neq i$ with $u_i < \min_j u_{kj}$, then L would strictly prefer choosing k above choosing i in g^P for sufficiently small $|P|$. $\square$

Proposition 2 implies that, when the noise is small, any equilibrium outcome of g^P is almost pure. This in turn implies that, if g has only mixed equilibria, the equilibrium outcomes of g are disjoint from the limit equilibrium outcomes of the noisy commitment game when the noise vanishes. This shows that a result similar to Proposition 1 cannot be proved for a "satisfactory" solution concept, i.e. there does not exist a refinement of the Nash equilibrium concept that generates a nonempty set of solutions for every game for which the equilibrium outcomes of the simultaneous move game coincide with those of the noisy commitment game when the noise vanishes.

It is not true that any Nash equilibrium outcome of the commitment game with perfect observability can be approximated by Nash equilibrium outcomes of games with slight noise: the Nash equilibrium correspondence is not lower hemi continuous. In the game of Figure 1, (B, W) is a Nash outcome of the non-noisy game: It is optimal for L to commit to B if F responds to T with E. However, noise forces F to choose W in response to any signal since W is a dominant strategy. Consequently, only (T, W) can be approximated by equilibrium outcomes of noisy games. (More generally, it follows that g^P has a unique equilibrium outcome in case F has a dominant strategy in g.)

	W	E
T	3,3	0,0
B	2,2	1,1

Figure 1.

In this paper we take the position that there is no a priori reason to discriminate against equilibria that are not in pure strategies. Consequently, we have to take mixed strategy equilibria into account and Proposition 2 raises the question of which outcomes of the base game g can be approximated by equilibrium outcomes of the game g^P when the noise vanishes. Proposition 3 gives part of the answer: If the noise P is small, there is always an equilibrium that produces an outcome that is close to the Stackelberg outcome, i.e. that is close to the subgame perfect equilibrium outcome of the sequential move game without noise. We will refer to such an equilibrium as a noisy Stackelberg equilibrium.

Proposition 3 . *The game g^P has an equilibrium $\sigma^P = (s^P, f^P)$ with an outcome z^P that converges to $(1, b_1)$ as $|P| \to 0$.*

Proof. Consider the reduced strategic form $\bar{g}^P$ that results from the strategic form of g^P by eliminating all pure strategies of F that do not prescribe to play b_1 after the signal "1". In this reduced game, player L's expected payoff resulting from playing "1" is approximately u_1 if the noise is small, no matter what F plays. Let $\sigma^P = (s^P, f^P)$ be an equilibrium of $\bar{g}^P$. If $s_i^P > \sqrt{|P|}$ for some $i \neq 1$, then Lemma 1 guarantees that $f_i^P = b_i$ provided that $|P| < \varepsilon^*$. However, in this case L's payoff resulting from "i" is approximately u_i, hence, $u^P(i, f^P) < u^P(1, f^P)$, so that player L wants to choose i with probability zero. The contradiction shows that, if $|P|$ is sufficiently small

$$s_i^P \leq \sqrt{|P|} \quad \text{for all } i \neq 1. \tag{3.2}$$

The inequalities (3.2) in turn imply that $s_1^P \to 1$ as $|P| \to 0$, hence, (by Lemma 1) that at the signal "1" only b_1 is a best response of player F. This shows that σ^P is an equilibrium of g^P if $|P|$ is small. Obviously, the outcome z^P of σ^P converges to $(1, b_1)$ as $|P| \to 0$. $\qquad\qquad\square$

We have seen two sufficient conditions for limit equilibrium outcomes (Propositions 1 and 3) and one necessary condition (Proposition 2). The necessary condition is not sufficient (Figure 1) and the sufficient conditions are not necessary: Also outcomes that are not pure Nash equilibria, nor Stackelberg equilibria of g may be approximated. Consider the game of Figure 2 in which L has M as a dominant strategy, so that (M, C) is the unique Nash equilibrium. The Stackelberg equilibrium is (T, W). Consider the noisy commitment game with uniform noise, i.e. $p_{ij} = \varepsilon$ if $i \neq j$ and $p_{ii} = 1 - 2\varepsilon$. It is easily seen that the following strategy combination is an equilibrium of this game: Player L commits to M with probability $\frac{3\varepsilon}{1+\varepsilon}$ and to B with the remaining probability $\frac{1-2\varepsilon}{1+\varepsilon}$; player F responds to signals T and B with E, after signal M he plays C with probability $\frac{2}{4-11\varepsilon}$ and E with probability $\frac{2-11\varepsilon}{4-11\varepsilon}$. The corresponding limit outcome is (B, E).

	W	C	E
T	4,4	0,0	0,0
M	5,0	1,1	5,0
B	0,0	0,0	3,3

Figure 2.

We will not attempt to describe exactly which outcomes can be obtained as limits of equilibrium outcomes of the noisy game as the noise tends to zero. Rather we conclude from the Propositions 1 and 3 that typically there exist multiple limits and, hence, that there exists an equilibrium selection problem. We will attempt to address this selection problem directly and we will propose an argument (an equilibrium selection theory) that actually selects a noisy Stackelberg equilibrium. Our theory incorporates elements from the theory proposed by Harsanyi and Selten (1988) as well as elements from the theory proposed in Harsanyi (1993), however, it differs from these and it may select different outcomes. In particular, neither the theory of Harsanyi and Selten (1988) nor that of Harsanyi (1993) need to select a noisy Stackelberg equilibrium. The next section describes our theory and proves the main result of this paper, while Section 5 discusses the theories of Harsanyi/Selten and Harsanyi.

4 Equilibrium Selection

The strategy b of player F that prescribes to play the best response b_k against action $k \in \mathcal{I}$ for any signal k is a (weakly) dominant strategy in the (strategic form of) the game where L's commitment is observed perfectly. If there is a slight amount of noise (i.e. $P \neq P^0$), then b is no longer dominant, however, as long as the noise is small, it is quite likely that b is a best response. Specifically, as Lemma 1 has shown, if $|P| < \varepsilon^*$ and $s_k > \sqrt{|P|}$ for all k, then b is the unique best response against s in g^P. To put it differently, b is a best response to a set of mixed strategies of player L in g^P that converges to the set of all strategies as $|P| \to 0$. On the basis of these considerations it would seem that L should assign a large (prior) probability to F playing b and, hence, he will be tempted to play his Stackelberg leader strategy "1". However, if $P \neq P^0$ and b_1 is not a dominant strategy in g, then $(1, b)$ is not an equilibrium of g^P, so that a theory that tells player L to play "1" and that tells F to play b is self-destroying. The simple point we make in this section is that, if the players' reasoning process corresponds to the tracing procedure (Harsanyi (1975), Harsanyi and Selten (1988)), then players will finally coordinate on a noisy Stackelberg equilibrium if they start from a prior that assigns sufficient weight to F playing b.

The tracing procedure is a process that gradually adjusts players' plans and expectations until they are in equilibrium. We only describe the mechanics of this procedure, for the motivation and heuristic description of the process we refer to the original sources. Let $\sigma^0 = (s^0, f^0)$ be a mixed strategy combination[2] in g^P. We interpret σ^0 a the players' prior expectations, hence, a priori player F believes that L will play i with probability s_i^0 while L believes that F will play the pure strategy f with probability $f^0(f)$. For $t \in [0, 1]$ consider the strategic form g^{P,t,σ^0} defined by

$$u^{P,t,\sigma^0}(i, f) = t u^P(i, f) + (1 - t) u^P(i, f^0) \tag{4.1}$$

$$v^{P,t,\sigma^0}(i, f) = t v^P(i, f) + (1 - t) v^P(s^0, f) \tag{4.2}$$

Hence, for $t = 1$ this game coincides with g^P, while for $t = 0$ we have a trivial game

in which each player's payoff depends only on his prior expectations. Write $\Gamma^P(\sigma^0)$ for the graph of the equilibrium correspondence, i.e.

$$\Gamma^P(\sigma^0) = \{(t; s, f) : t \in [0, 1], (s, f) \text{ is equilibrium of } g^{P,t,\sigma^0}\} \tag{4.3}$$

It can be shown that in nondegenerate cases this graph $\Gamma^P(\sigma^0)$ contains a unique distinguished curve that connects the unique equilibrium of $g^{P,0,\sigma^0}$ to an equilibrium (s^1, f^1) of g^P. (See Schanuel et al. (1991) for details.) The (linear) tracing procedure consists of following this curve until its endpoint, and the endpoint $T^P(\sigma^0) = (s^1, f^1)$ is called the linear trace of σ^0 in g^P. The interpretation is that players eventually reason themselves to the equilibrium $T^P(\sigma^0)$ if they start from the prior σ^0. Write $z^P(\sigma^0)$ for the outcome of this linear trace $T^P(\sigma^0)$ in g^P. We have that this outcome is close to the Stackelberg outcome $(1, b_1)$ of g if $|P|$ is small. Formally

Lemma 2 . *If the prior $\sigma^0 = (s^0, f^0)$ is such that $f^0(b)$ is sufficiently close to one, then* $\lim_{|P| \to 0} z^P(\sigma^0) = (1, b_1)$.

Proof. Let $f^0(b)$ be large enough such that

$$u^{P^0}(i, f^0) < u^{P^0}(1, f^0), \tag{4.4}$$

i.e. if player L's commitment is perfectly observed by F, then L strictly prefers to play "1" when F responds with f^0. Note that the regularity condition (2.1) implies that (4.4) holds whenever $f^0(b)$ is sufficiently close to 1. Condition (4.4) in turn implies that there exist $\varepsilon > 0$ and $t^* > 0$ such that

$$\text{"1" is strictly dominant for } L \text{ in } g^{P,t,\sigma^0} \text{ if } t < t^* \text{ and } |P| < \varepsilon \tag{4.5}$$

Furthermore, by choosing ε sufficiently small we can guarantee that for all $i \neq 1$:

if $|P| < \varepsilon$, then $u^P(i, f) < u^P(1, f)$ for all f with $f_i = b_i$ and $f_1 = b_1$ (4.6)

We will restrict ourselves to stochastic matrices P with

$$\sqrt{|P|} \leq t^*/2I. \tag{4.7}$$

Finally, with ε^* as in Lemma 1, we assume that

$$|P| < \varepsilon^*. \tag{4.8}$$

Let P be such that (4.5) - (4.8) hold and denote by $\sigma^{P,t} = (s^{P,t}, f^{P,t})$ an equilibrium on the distinguished curve in $\Gamma^P(\sigma^0)$ that connects the unique equilibrium of $g^{P,0,\sigma^0}$ with $T^P(\sigma^0)$. We claim that

$$s_i^{P,t} < 1/2I \quad \text{for all } i \neq 1 \text{ and all } t. \tag{4.9}$$

Assume, to the contrary, that there exist some $i \neq 1$ and t such that $s_i^{P,t} \geq 1/2I$ and let τ be the smallest t for which an equilibrium of this type can be found. Then $\tau \geq t^*$ in view of (4.5). Hence, at $t = \tau$, the total probability that F assigns to L playing i in g^{P,t,σ^0} is at least $t^*/2I$, so that (4.7), (4.8) and Lemma 1 guarantee that $f_i^{P,\tau} = b_i$. At the same time we have that

$$s_1^{P,\tau} = 1 - \sum_{i \neq 1} s_i^{P,\tau} > 1 - I/2I = 1/2 \geq \sqrt{|P|}$$

so that $f_1^{P,\tau} = b_1$ by the same argument. But now (4.5) and (4.6) imply that

$$u^{P,\tau,\sigma^0}(i, f^{P,\tau}) < u^{P,\tau,\sigma^0}(1, f^{P,\tau}),$$

hence, $s_i^{P,\tau} = 0$. The contradiction shows that (4.9) holds. In particular, we have that $s_1^{P,1} > 1/2$, hence $f_1^{P,1} = b_1$ in view of Lemma 1. Applying Lemma 1 and (4.6) once more we see that, therefore, $s_i^{P,1} \leq \sqrt{|P|}$ for all $i \neq 1$, hence, that

$$\lim_{|P| \to 0} s_1^{P,1} = 1.$$

This completes the proof. □

To complete our argument that players will (or should) coordinate on a noisy Stackelberg equilibrium, we have to give an argument why player L should attach a high prior probability to F playing b. We will borrow such an idea from Harsanyi (1993). Harsanyi proposes that the prior should be based on (should be proportional to) the structural incentive that a player has to use this strategy and he suggests to measure this structural incentive by the size of the stability set.

Formally, Harsanyi proceeds as follows. Let $g = < A_1, A_2, u_1, u_2 >$ be a 2-person game and let $S_i = \Delta(A_i)$ be player i's set of mixed strategies. The stability set of $s_i \in S_i$ is the set $S_j(s_i)$ of all mixed strategies of player j against which s_i is a best response. At first it seems natural to measure the structural incentives of a pure strategy a_i by the Lebesgue measure of $S_j(a_i)$, but Harsanyi (1993) shows that this definition would violate certain desirable properties. To circumvent these, Harsanyi first transforms the strategy simplex S_j by the so-called inversion mapping ω_j and he then takes the Lebesgue measure of the transformed set. Formally, ω_j is the mapping from the interior of S_j to the interior of S_j that maps s_j into $\bar{s}_j$ defined by

$$\bar{s}_j(a_j) = s_j^{-1}(a_j) / \sum_{a \in A_j} s_j^{-1}(a). \tag{4.10}$$

Hence, Harsanyi measures the structural incentives of player i to use the pure strategy

$a_i \in \Lambda_i$ by number $\rho(a_i) = \lambda(\omega_j(S_j(a_i)))$ where λ denotes Lebesgue measure. The prior probability that player j then assigns to i playing a_i is proportional to these incentives, hence,

$$p_j(a_i) = \rho(a_i)/ \sum_{a \in \Lambda_i} \rho(a). \tag{4.11}$$

In the special case of our noisy commitment game, we have that the stability set of the strategy b of player F converges to the entire strategy simplex of player L as $|P| \to 0$ and, hence, that the stability set of any other pure strategy converges to a set of measure zero. It follows that the prior of player L, as constructed by using (4.10) and (4.11) puts almost all weight on the strategy b of player F when $|P|$ is small. Hence, from Lemma 2 we can conclude that players will end up in the Stackelberg equilibrium in the limit. Formally, we have proved

Proposition 4 . *If players construct their prior beliefs by using Harsanyi's (1993) theory of structural incentives and if they update their priors by using the tracing procedure of Harsanyi (1975) to obtain an equilibrium, then, in the limit when the noise vanishes, they will play the Stackelberg equilibrium.*

5 Alternative Methods of Equilibrium Selection

5.1 Evolutionary and Eductive Theories

In this section we show that the theories proposed in Harsanyi and Selten (1988) and in Harsanyi (1993) do not necessarily select a noisy Stackelberg equilibrium as the solution of the game with imperfectly observable commitment. The basic reason is that these theories do not consider all equilibria of a game to be eligible as solution candidates. Both theories start by eliminating certain Nash equilibria as candidates. Specifically, equilibria that are considered to have poor stability properties are eliminated. Harsanyi and Selten (1988) first eliminate all primitive equilibria, i.e. equilibria that do not belong to a primitive formation. A formation is a set of strategy pairs that is closed with

respect to taking best responses and a formation is said to be primitive if it does not properly contain any other formation. Harsanyi (1993) only considers equilibria that are both proper (Myerson (1978)) and persistent (Kalai and Samet (1984)) as eligible. For generic 2-person games every Nash equilibrium is proper and an equilibrium is persistent if and only if it belongs to a primitive formation.[3] Hence, for generic 2-person games, both theories start from the same set of initial candidates.

The game displayed in Figure 3 may show that the restriction to primitive (persistent) equilibria may eliminate any noisy Stackelberg equilibrium. The game g^P has three equilibria: one corresponds to Proposition 1 (with outcome (1,1)), another corresponds to Proposition 3 (with outcome close to (3,3)), and there is a third mixed strategy equilibrium. Action B (i.e. the dominant strategy of L in g) is used with positive probability in all three equilibria and the unique best response of player F against B in g^P is to always respond with E. Consequently, $\{(B, EE)\}$ is the unique primitive formation in g^P, hence (B, EE) is the unique persistent equilibrium of this game. Therefore, the theories of Harsanyi/Selten and Harsanyi select the pure equilibrium of g as the solution of g^P. These theories confirm Bagwell's claim that slight noise eliminates the commitment power.

	W	E
T	3,3	0,2
B	4,0	1,1

Figure 3.

The argument used in the above example can be generalised. If one accepts persistency as a selection criterion, one is led to the conclusion that in any game in which the leader has a dominant strategy, slight noise eliminates the benefits of the leader being able to commit himself:

Proposition 5 . *If player L has a dominant strategy in g, then g^P has a unique primitive formation (resp. persistent retract), viz. the singleton set in which L plays this*

dominant strategy i and in which F responds with b_i to any signal k.

Proof. Let $R = R_L \times R_F$ be a persistent retract (resp. primitive formation) and let j be a pure strategy of player L in R_L. Then F has a unique best response against j in g^P, viz. the strategy f with $f_k = b_j$ for all k. Hence, $f \in R_F$. The unique best response of L against f is to play his dominant strategy i from g, hence $i \in R_L$. Let $\bar{f}_k = b_i$ for all k. Then the strict equilibrium $(i, \bar{f})$ belongs to R. Consequently, if R is primitive (persistent), then $R = \{(i, \bar{f})\}$. □

The basic reason why Harsanyi and Selten eliminate equilibria that are not primitive is that such equilibria may have very poor stability properties (cf. Harsanyi and Selten (1988, p. 201) and Harsanyi (1993, footnote 12)). Requiring persistency favors the selection of equilibria that have similar stability properties as strict equilibria, hence, the solution theories of Harsanyi and Selten are biased in favor of the selection of pure equilibria. However, one may very well wonder whether such a bias is justified: The stability property captured by persistency may be relevant in an evolutionary context — where the game is repeatedly played by a large population of players who receive feedback about evolution of the play during the game (see, for example, Hurkens (1994)) — but it is not clear that it has any relevance in the case where the game is played only once and players rely exclusively on deductive personal reflection in order to figure out what to play. At the same time, the theories of Harsanyi/Selten and Harsanyi rely strongly on arguments (such as the tracing procedure) that seem to be particularly relevant in this latter case and that seem irrelevant in the former. Hence, these theories may be criticized for the fact that they mix arguments that are relevant in an evolutionary context with arguments that are relevant in an eductive context. In the following subsections we return to the purely deductive perspective.

5.2 Harsanyi's (1993) Theory

In this subsection we show that, even in the case where the Stackelberg equilibrium is a strict equilibrium of g^P and, hence, satisfies all of Harsanyi's (1993) eligibility criteria, Harsanyi's theory need not select this Stackelberg equilibrium. The reason is that Harsanyi's theory does not invoke the tracing procedure. Rather, Harsanyi proposes to select as the solution of the game that equilibrium that has the highest prior probability. With the prior probability of a pure strategy as in (4.11), the prior probability of a pure strategy pair a is simply given by

$$p(a) = p_F(a_L)p_L(a_F) \tag{5.1}$$

and in the case where only pure equilibria are eligible, Harsanyi selects that equilibrium a^* for which $p(a^*)$ is largest. (At least this is the solution in case the argmax is unique.) The game from Figure 4 (in which K is some real positive number) may show that this procedure need not select the Stackelberg equilibrium.

	W	E
T	2,1	0,0
B	3,0	1,K

	WW	WE	EE
T	2,1	$2(1-\varepsilon),1-\varepsilon$	0,0
B	0,0	$1-\varepsilon,K(1-\varepsilon)$	1,K

Figure 4.

The game g from the left panel of Figure 4 is a unanimity game with Stackelberg outcome $(2,1)$. The panel on the right displays (a reduced form of) the game g^P where P involves uniform noise ($p_{ij} = \varepsilon$ if $i \neq j$). We have eliminated the strategy EW for player F in g^P (i.e. the strategy in which F responds to T by E and to B by W) since this is a dominated strategy. Harsanyi indeed suggests to eliminate all dominated strategies before computing the players' structural incentives. The game g^P has three equilibria (T,WW), (B,EE) and a mixed equilibrium. Only the former two satisfy Harsanyi's eligibility criteria, hence, to compute the Harsanyi solution of the game, we have to compare the prior probabilities of these equilibria. Note that although player

L's prior assigns almost all weight to the strategy WE of player F, this prior probability plays no role in this comparison.

Note that the structural incentives for player L to use any of his pure strategies are independent of K: These incentives only depend on player L's own payoff matrix. Furthermore, note that both the prior of T and the prior of B remain bounded away from zero as ε tends to zero. Turning now to the structural incentives of player F, we note that the calculations are simple since, in the 1-dimensional case, the inversion mapping is measure preserving. Hence, the prior probability of a strategy is just the Lebesgue measure of the stability set of that strategy. Straightforward computations show that

$$p_L^\varepsilon(WW) = \varepsilon/(K - K\varepsilon + \varepsilon) \tag{5.2}$$

and

$$p_L^\varepsilon(EE) = K\varepsilon/(K\varepsilon + 1 - \varepsilon), \tag{5.3}$$

hence

$$\lim_{\varepsilon \downarrow 0} p_L^\varepsilon(EE)/p_1^\varepsilon(WW) = K^2. \tag{5.4}$$

It follows that, if K is sufficiently large

$$\lim_{\varepsilon \downarrow 0} p^\varepsilon(T, WW) < \lim_{\varepsilon \downarrow 0} p^\varepsilon(B, EE) \tag{5.5}$$

and, hence, that Harsanyi's theory selects the equilibrium (B, EE) in that case. For large values of K, Harsanyi's theory does not select the Stackelberg equilibrium.

5.3 Risk Dominance and the Harsanyi/Selten Theory

An essential ingredient in the equilibrium selection theory from Harsanyi and Selten (1988) is the notion of risk dominance. An equilibrium s is said to risk dominate an equilibrium s' if the tracing procedure, when started at a certain (bicentric) prior $p(s, s')$ ends up at the equilibrium s. (Below we describe how this bicentric prior has to be computed.) Starting from an initial candidate set, Harsanyi and Selten repeatedly eliminate equilibria that are either payoff dominated or risk dominated until finally only one candidate — the solution — remains. We have already seen that the Stackelberg equilibrium need not belong to the initial candidate set, hence, the Harsanyi/Selten theory need not select it. However, in Section 5.1 we argued that this elimination step is not convincing. Hence, the question remains whether the Stackelberg equilibrium can be eliminated by considerations of payoff dominance or risk dominance.

Proposition 2 implies that the noisy Stackelberg equilibrium cannot be payoff dominated when the noise is small. Any Nash equilibrium outcome of the noisy game converges to a Nash outcome of the game in which the commitment is observed perfectly and among the latter the Stackelberg equilibrium is most preferred by player L. Consequently, it remains to address the question of whether the Stackelberg equilibrium can be risk dominated. We have not been able to resolve the issue in its complete generality, however, for two important subclasses of games — 2×2 games and unanimity games — we can show that the (noisy) Stackelberg equilibrium risk dominates any other equilibrium of g^P when the noise P is small.

To formally define the risk dominance relation we have to describe how the bicentric prior $p(s, s')$ should be computed at which to start the tracing procedure. Harsanyi and Selten have the situation in mind where it is common knowledge among the players that either s or s' is the solution of the game. Each player i will initially assume that his opponent j already knows which of the two is the solution. Player i will assign a subjective probability z_i to the solution being s (and, hence, to j playing s_j) and he will assign the complementary probability $z_i' = 1 - z_i$ to j playing s_j'. After having

constructed these beliefs, i will play a best response $b_i(z_i)$ against $z_i s_j + z_i' s_j'$. Player j does not know i's beliefs z_i and, according to the principle of insufficient reason, j will assume that z_i is uniformly distributed on the interval $[0, 1]$. Hence, j will expect i to play the strategy

$$p_j(s, s') = \int_0^1 b_i(z_i) dz_i. \tag{5.6}$$

The mixed strategy of player i defined by (5.6) describes player j's a priori beliefs which are used to determine the risk dominance relation between s and s'.

Before being able to state the main result of this section, one more definition is needed. We say that $g = < \mathcal{I}, \mathcal{J}, u, v >$ is a unanimity game if (a) $\mathcal{I} = \mathcal{J}$, (b) $u_{ij} = v_{ij} = 0$ for all $i \neq j$, and (c) $u_{ii} > 0$ and $v_{ii} > 0$ for all i. We simplify notation by writing $u_i = u_{ii}$ and $v_i = v_{ii}$ and recall from (2.3) that $u_1 > u_i$ for $i \neq 1$. We also write "$\underline{i}$" for the strategy of player F in g^P that prescribes to respond to any signal $k \in \mathcal{I}$ by playing $i \in \mathcal{I}$.

Proposition 6 . *Let g be a unanimity game. Then the Stackelberg equilibrium $(1, \underline{1})$ risk dominates any other equilibrium of g^P when the noise P is small.*

Proof. We first show that $(1, \underline{1})$ risk dominates any other pure Nash equilibrium of g^P when $|P|$ is small. It suffices to show that $(1, \underline{1})$ risk dominates $(2, \underline{2})$. We first compute the bicentric prior that is used in the risk dominance comparison. Let us first compute the prior p_F of player F. If F plays $z.\underline{1} + (1 - z).\underline{2}$ then the best response of L is

$$b_L^P(z) = \begin{cases} 1 & \text{if } z > \quad u_2/(u_1 + u_2) \\ 2 & \text{if } z < \quad u_2/(u_1 + u_2) \end{cases} \tag{5.7}$$

hence, the prior of F is given by

$$p_F^P(i) = \begin{cases} u_1/(u_1 + u_2) & \text{if } i = 1 \\ u_2/(u_1 + u_2) & \text{if } i = 2 \end{cases} \tag{5.8}$$

Next we compute the prior of player L. If L plays $z.1 + (1-z).2$, then the best response of F depends on the message that F receives and on the size of the noise. However, since the posterior of F puts positive weight only on the actions 1 and 2 of player L, F will respond with either 1 or 2 at each possible message. Furthermore, if the noise is small, then F will respond to the message $i = 1$ (resp. $i = 2$) with the action 1 (resp. 2) for most values of z. Hence, without doing any computations, we may state that player L's prior P_L^P corresponds to a behavioral strategy f^0 of player F that is of the following form:

$$f_{ik}^0 = \begin{cases} \approx 1 & \text{if } i = 1 \quad \text{and } k = 1 \\ \approx 1 & \text{if } i = 2 \quad \text{and } k = 2 \\ 0 & \text{if } i \notin \{1,2\} \quad \text{and } k \notin \{1,2\} \end{cases} \tag{5.9}$$

(f_{ik}^0 is the probability that F responds to signal i with action k.) Now, let the prior $\sigma^0 = (p_F^P, p_L^P) = (p_F^P, f^0)$ be as in (5.8), (5.9) and let the game g^{P,t,σ^0} be as in (4.1), (4.2). If t is sufficiently small, then the unique equilibrium $(s^{P,t}, f^{P,t})$ of this game is the best reply against the prior, hence

$$f_{ik}^{P,t} = \begin{cases} 1 & \text{if } i = 1 \quad \text{and } k = 1 \\ 1 & \text{if } i = 2 \quad \text{and } k = 2 \\ 0 & \text{if } i \notin \{1,2\} \quad \text{and } k \notin \{1,2\} \end{cases} \tag{5.10}$$

and, provided that $|P|$ is sufficiently small,

$$s_i^{P,t}(i) = 1 \quad \text{if } i = 1. \tag{5.11}$$

Hence, in particular, player L chooses the Stackelberg strategy with probability 1 for small t. We claim that, if we move along the distinguished curve in $\Gamma^P(\sigma^0)$ by increasing t, then player F has to switch his strategy before player L does. The argument is simply that, if F does not switch from a strategy as in (5.10), then L is facing a convex combination of strategies of type (5.9) and (5.10), hence, this is just a strategy of type (5.9), against which the strategy from (5.11) is the unique best response. Hence, as t increases, player F's posterior beliefs put more and more weight on L playing "$\underline{1}$" and gradually F switches to respond with "1" at more and more messages. Such changes in behavior of F however, do not necessitate a change in behavior of L: The strategy from (5.11) remains a best response. Consequently, if no equilibrium is reached yet, F will have to change again. Eventually (when t gets close to 1), F's posterior after the message "2" will put so much weight on L playing "1" that F will respond to that message by playing "1" as well. At that point in time we have obtained the equilibrium $(1,\underline{1})$ from g^P and no further adjustments are necessary. Hence, starting at the prior (5.8) – (5.9), the tracing procedure converges to $(1, \underline{1})$, so that $(1,\underline{1})$ risk dominates $(2,\underline{2})$. Hence, the Stackelberg equilibrium risk dominates any pure equilibrium of g^P.

Next, let s' be a mixed strategy equilibrium of g^P. Proposition 2 implies that, if the noise is small, there exists an action $i \in \mathcal{I}$ such that player L plays i with a probability very close to one. If $i = 1$, then $(1,\underline{1})$ is the unique equilibrium of g^{P,t,σ^0} for all t. If $i \neq 1$, then the proof follows exactly the same line as above: Player L plays "1" for each value of t and player F switches several times until he finally responds to all messages by playing "1". $\qquad\square$

Our final result is

Proposition 7 . *If g is 2×2 game and $|P|$ is small, then g^P has one equilibrium that risk dominates all other equilibria and the outcome generated by this risk dominant equilibrium converges to the Stackelberg outcome $(1, b_1)$ as $|P| \to 0$.*

Proof. The result follows from Proposition 3 in case player F has a dominant strategy in g (g^P has only one equilibrium in this case). Hence, assume that F does not have a dominant strategy. Without loss of generality assume $b_1 = 1$ and $b_2 = 2$. In case g does not have any pure equilibria, the result again follows from Proposition 3 since g^P has a unique equilibrium in this case. (The unique best response of F against strategy i of player L is to respond with i to any message, but then L's best response is to play $j \neq i$.) There are three cases left to consider:

   (i) (1,1) is the unique pure equilibrium of g.

   (ii) (2,2) is the unique pure equilibrium of g.

   (iii) both (1,1) and (2,2) are pure equilibria in g.

The first case is easy: It can be resolved by iterative elimination of strictly dominated strategies. (It should be obvious from the description of risk dominance on the preceding pages that strategies that are iteratively strictly dominated cannot influence the risk dominance relationship.) The strategy "21" of player F (play $k \neq i$ in response to i for $i = 1, 2$) is strictly dominated and once this strategy has been eliminated, the strategy 1 becomes strictly dominant for player L. (Note that action 1 is dominant for L in g in case (i).) The third case is very much like the case considered in Proposition 6 and the proof proceeds along the same lines. We leave the details to the reader. In case (ii), g^P has three equilibria, viz. a mixed equilibrium with outcome close to (1,1), a mixed equilibrium with outcome close to (2,2), and the pure equilibrium (2,2). We have to show that the first equilibrium risk dominates the latter two. The proof follows from Lemma 2. Namely, consider the bicentric prior p_L^P of player L in game g^P relevant for the comparison between the noisy Stackelberg equilibrium and the pure equilibrium (2,2). The reader easily verifies that

$$\lim_{|P| \to 0} p_L^P(b) = 1, \tag{5.12}$$

since the strategy b of player F (with $b_i = i$ all i is a best response to the noisy Stackelberg equilibrium and is "almost" a best response to the pure equilibrium. Hence, it

follows from Lemma 2 that the noisy Stackelberg equilibrium risk dominates the pure Nash equilibrium. To show that this equilibrium also risk dominates the third mixed equilibrium, we note that the strategy b of player F is the unique best response against a strict convex combination of the two mixed equilibrium strategies of player L in g^{F}. Hence, in this case the prior satisfies $p_{L}^{F}(12) = 1$ and the conclusion again follows from Lemma 2. □

Although we conjecture that the result from the Propositions 6 and 7 can be generalized to other classes of games, we have to admit that we have not been able to find a general proof. (We do not have a counterexample either.) However, we note that applying the tracing procedure can be rather complex, so that a multilateral procedure as that in Section 4 – in which the tracing procedure is applied only once – might be preferable to a theory in which one is forced to make a rather large number of bilateral comparisons. Furthermore, in order to apply the Harsanyi/Selten theory one has to first compute all (primitive) equilibria of the game. We were able to prove Propositon 4 without knowing this set of all equilibria.

6   Conclusion

From the fact that any pure Nash equilibrium of a 2-person simultaneous move game is also a pure Nash equilibrium outcome of the sequential move game in which the follower can only observe imperfectly the action to which the leader committed himself (Proposition 1 in this paper), Kyle Bagwell concluded in his 1992 paper that slight noise eliminates any first mover advantages. In the concluding section of his paper, Bagwell writes

"For applied theorists, the key message of the paper is that the many predictions derived from models with commitment may require reconsideration. Apparently these predictions are valid only for settings in which the committed action is in fact *perfectly* observed by subsequent players. This requirement is quite stringent, and it would seem to be violated in a number

of real-world settings to which popular commitment models are thought to apply" (Bagwell (1992, p. 9) emphasis in original).

While we agree with the observation that the assumption of perfect observability is stringent, we disagree with the statement that this assumption is crucial. In fact, we would claim that this paper shows that the assumption is inessential. Not only have we shown that the noisy game analyzed by Bagwell has always an equilibrium outcome that is close to the subgame perfect equilibrium of the game in which the commitment can be observed perfectly (Proposition 3), we have also given several arguments for why players should coordinate on this particular equilibrium (Propositions 4, 6 and 7). In addition, we have remarked that the structure of the noise as assumed by Bagwell is somewhat peculiar and that other specifications, which are, perhaps, more natural and which are closer to Schelling's original ideas (Schelling (1960, p. 149)) also allow the conclusion that the assumption of perfect observability is inessential. Hence, we do not see any need to reconsider the fundamental game theoretic insight that the power to commit oneself may be beneficial.

References

Bagwell, K. (1992). "Commitment and observability in games". Northwestern University DP 1014.

Balkenborg, D. (1992). "The Properties of Persistent Retracts and Related Concepts". Ph.D. Thesis, University of Bonn.

Chakravorti, B. and Y. Spiegel (1993). "Commitment Under Imperfect Observability: Why it is better to have followers who know that they don't know rather than those who don't know that they don't know". Bellcore Econommics DP 104.

Harsanyi, J. (1975). "The Tracing Procedure: A Bayesian approach to defining a solution for n-person noncooperative games". *International Journal of Game Theory* **4**, 61-94.

Harsanyi, J. and R. Selten (1988). *A general theory of equilibrium selection in games.* Cambridge, Mass.: MIT Press.

Harsanyi, J. (1993). "A New Theory of Equilibrium Selection for Games with Complete Information". Mimeo University of California at Berkeley.

Hurkens, S. (1994). "Learning by Forgetful Players: From primitive formations to persistent retracts". CentER DP 9437.

Kalai, E. and D. Samet (1984). "Persistent equilibria in strategic games". *International Journal of Game Theory* **13**, 129-144.

Myerson, R.B. (1978). "Refinements of the Nash Equilibrium Concept". *International Journal of Game Theory* **7**, 73-80.

Schanuel, S.H., L.K. Simon and W.R. Zame (1991). "The Algebraic Geometry of Games and the Tracing Procedure", in: R. Selten (ed.), *Game Equilibrium Models, Vol. 2: Methods, Morals and Markets*, Springer Verlag, 9-43.

Schelling, T. (1960). *The Strategy of Conflict.* Harvard University Press, Cambridge, Mass.

Van Damme, E. (1987). *Stability and Perfection of Nash Equilibria.* Springer Verlag, Berlin. Second edition 1991.

Von Stackelberg, H. (1934). *Marktform und Gleichgewicht.* Springer Verlag, Vienna and Berlin.

Endnotes

1. Bagwell (1992) restricts himself to the case where player F has a unique best response to any pure action i of player L. He writes that the basic results are most easily reported in this case, from which the reader might be tempted to conclude that his result (Proposition 1 in our paper) is also valid for games that do not satisfy this condition. That conclusion, however, is unwarranted as the following

example shows

	W	E
T	2,4	0,4
B	3,0	1,2

Figure A1.

The unique pure equilibrium of the game g of Figure A1 is (B, E), however, (T, WE) is also a pure Nash equilibrium of g^P and this equilibrium results in the outcome (T, W). The reader might object that the latter equilibrium is not credible since it is not perfect (although it certainly is sequential). This deficiency is easily eliminated by adding a third (dominated) strategy of player L to the game against which W is the unique best response of player F.

2. We could equivalently work with behavioral strategies, cf. also (5.9).

3. For a proof of the first statement, see Van Damme (1987, Thm 2.6.2). The second statement follows from the observation that in generic games, a pure strategy that is a best response is a unique best response against an open set of strategies in the neighborhood. See Balkenborg (1992) for further details about the proof.

Discussion Paper Series, CentER, Tilburg University, The Netherlands:

(For previous papers please consult previous discussion papers.)

No.	Author(s)	Title
9355	Z. Yang	A Simplicial Algorithm for Computing Robust Stationary Points of a Continuous Function on the Unit Simplex
9356	E. van Damme and S. Hurkens	Commitment Robust Equilibria and Endogenous Timing
9357	W. Güth and B. Peleg	On Ring Formation In Auctions
9358	V. Bhaskar	Neutral Stability In Asymmetric Evolutionary Games
9359	F. Vella and M. Verbeek	Estimating and Testing Simultaneous Equation Panel Data Models with Censored Endogenous Variables
9360	W.B. van den Hout and J.P.C. Blanc	The Power-Series Algorithm Extended to the $BMAP/PH/1$ Queue
9361	R. Heuts and J. de Klein	An (s,q) Inventory Model with Stochastic and Interrelated Lead Times
9362	K.-E. Wärneryd	A Closer Look at Economic Psychology
9363	P.J.-J. Herings	On the Connectedness of the Set of Constrained Equilibria
9364	P.J.-J. Herings	A Note on "Macroeconomic Policy in a Two-Party System as a Repeated Game"
9365	F. van der Ploeg and A. L. Bovenberg	Direct Crowding Out, Optimal Taxation and Pollution Abatement
9366	M. Pradhan	Sector Participation in Labour Supply Models: Preferences or Rationing?
9367	H.G. Bloemen and A. Kapteyn	The Estimation of Utility Consistent Labor Supply Models by Means of Simulated Scores
9368	M.R. Baye, D. Kovenock and C.G. de Vries	The Solution to the Tullock Rent-Seeking Game When $R > 2$: Mixed-Strategy Equilibria and Mean Dissipation Rates
9369	T. van de Klundert and S. Smulders	The Welfare Consequences of Different Regimes of Oligopolistic Competition in a Growing Economy with Firm-Specific Knowledge
9370	G. van der Laan and D. Talman	Intersection Theorems on the Simplotope
9371	S. Muto	Alternating-Move Preplays and $vN - M$ Stable Sets in Two Person Strategic Form Games
9372	S. Muto	Voters' Power in Indirect Voting Systems with Political Parties: the Square Root Effect

No.	Author(s)	Title
9373	S. Smulders and R. Gradus	Pollution Abatement and Long-term Growth
9374	C. Fernandez, J. Osiewalski and M.F.J. Steel	Marginal Equivalence in v-Spherical Models
9375	E. van Damme	Evolutionary Game Theory
9376	P.M. Kort	Pollution Control and the Dynamics of the Firm: the Effects of Market Based Instruments on Optimal Firm Investments
9377	A. L. Bovenberg and F. van der Ploeg	Optimal Taxation, Public Goods and Environmental Policy with Involuntary Unemployment
9378	F. Thuijsman, B. Peleg, M. Amitai & A. Shmida	Automata, Matching and Foraging Behavior of Bees
9379	A. Lejour and H. Verbon	Capital Mobility and Social Insurance in an Integrated Market
9380	C. Fernandez, J. Osiewalski and M. Steel	The Continuous Multivariate Location-Scale Model Revisited: A Tale of Robustness
9381	F. de Jong	Specification, Solution and Estimation of a Discrete Time Target Zone Model of EMS Exchange Rates
9401	J.P.C. Kleijnen and R.Y. Rubinstein	Monte Carlo Sampling and Variance Reduction Techniques
9402	F.C. Drost and B.J.M. Werker	Closing the Garch Gap: Continuous Time Garch Modeling
9403	A. Kapteyn	The Measurement of Household Cost Functions: Revealed Preference Versus Subjective Measures
9404	H.G. Bloemen	Job Search, Search Intensity and Labour Market Transitions: An Empirical Exercise
9405	P.W.J. De Bijl	Moral Hazard and Noisy Information Disclosure
9406	A. De Waegenaere	Redistribution of Risk Through Incomplete Markets with Trading Constraints
9407	A. van den Nouweland, P. Borm, W. van Golstein Brouwers, R. Groot Bruinderink, and S. Tijs	A Game Theoretic Approach to Problems in Telecommunication
9408	A.L. Bovenberg and F. van der Ploeg	Consequences of Environmental Tax Reform for Involuntary Unemployment and Welfare

No.	Author(s)	Title
9409	P. Smit	Arnoldi Type Methods for Eigenvalue Calculation: Theory and Experiments
9410	J. Eichberger and D. Kelsey	Non-additive Beliefs and Game Theory
9411	N. Dagan, R. Serrano and O. Volij	A Non-cooperative View of Consistent Bankruptcy Rules
9412	H. Bester and E. Petrakis	Coupons and Oligopolistic Price Discrimination
9413	G. Koop, J. Osiewalski and M.F.J. Steel	Bayesian Efficiency Analysis with a Flexible Form: The AIM Cost Function
9414	C. Kilby	World Bank-Borrower Relations and Project Supervision
9415	H. Bester	A Bargaining Model of Financial Intermediation
9416	J.J.G. Lemmen and S.C.W. Eijffinger	The Price Approach to Financial Integration: Decomposing European Money Market Interest Rate Differentials
9417	J. de la Horra and C. Fernandez	Sensitivity to Prior Independence via Farlie-Gumbel -Morgenstern Model
9418	D. Talman and Z. Yang	A Simplicial Algorithm for Computing Proper Nash Equilibria of Finite Games
9419	H.J. Bierens	Nonparametric Cointegration Tests
9420	G. van der Laan, D. Talman and Z. Yang	Intersection Theorems on Polytopes
9421	R. van den Brink and R.P. Gilles	Ranking the Nodes in Directed and Weighted Directed Graphs
9422	A. van Soest	Youth Minimum Wage Rates: The Dutch Experience
9423	N. Dagan and O. Volij	Bilateral Comparisons and Consistent Fair Division Rules in the Context of Bankruptcy Problems
9424	R. van den Brink and P. Borm	Digraph Competitions and Cooperative Games
9425	P.H.M. Ruys and R.P. Gilles	The Interdependence between Production and Allocation
9426	T. Callan and A. van Soest	Family Labour Supply and Taxes in Ireland
9427	R.M.W.J. Beetsma and F. van der Ploeg	Macroeconomic Stabilisation and Intervention Policy under an Exchange Rate Band

No.	Author(s)	Title
9447	G. Koop, J. Osiewalski and M.F.J. Steel	Hospital Efficiency Analysis Through Individual Effects: A Bayesian Approach
9448	H. Hamers, J. Suijs, S. Tijs and P. Borm	The Split Core for Sequencing Games
9449	G.-J. Otten, H. Peters, and O. Volij	Two Characterizations of the Uniform Rule for Division Problems with Single-Peaked Preferences
9450	A.L. Bovenberg and S.A. Smulders	Transitional Impacts of Environmental Policy in an Endogenous Growth Model
9451	F. Verboven	International Price Discrimination in the European Car Market: An Econometric Model of Oligopoly Behavior with Product Differentiation
9452	P.J.-J. Herings	A Globally and Universally Stable Price Adjustment Process
9453	D. Diamantaras, R.P. Gilles and S. Scotchmer	A Note on the Decentralization of Pareto Optima in Economies with Public Projects and Nonessential Private Goods
9454	F. de Jong, T. Nijman and A. Röell	Price Effects of Trading and Components of the Bid-ask Spread on the Paris Bourse
9455	F. Vella and M. Verbeek	Two-Step Estimation of Simultaneous Equation Panel Data Models with Censored Endogenous Variables
9456	H.A. Keuzenkamp and M. McAleer	Simplicity, Scientific Inference and Econometric Modelling
9457	K. Chatterjee and B. Dutta	Rubinstein Auctions: On Competition for Bargaining Partners
9458	A. van den Nouweland, B. Peleg and S. Tijs	Axiomatic Characterizations of the Walras Correspondence for Generalized Economies
9459	T. ten Raa and E.N. Wolff	Outsourcing of Services and Productivity Growth in Goods Industries
9460	G.J. Almekinders	A Positive Theory of Central Bank Intervention
9461	J.P. Choi	Standardization and Experimentation: Ex Ante Versus Ex Post Standardization
9462	J.P. Choi	Herd Behavior, the "Penguin Effect", and the Suppression of Informational Diffusion: An Analysis of Informational Externalities and Payoff Interdependency
9463	R.H. Gordon and A.L. Bovenberg	Why is Capital so Immobile Internationally?: Possible Explanations and Implications for Capital Income Taxation
9464	E. van Damme and S. Hurkens	Games with Imperfectly Observable Commitment