

# Imperfect Commitment and the Revelation Principle: the Multi-Agent Case with Transferable Utility

Robert Evans<sup>1</sup> and Sönje Reiche<sup>2</sup>

May 30, 2007

## Abstract

Bester and Strausz (2000) showed that the revelation principle of Bester and Strausz (2001) does not apply in a setting of many agents and no commitment. In their counterexample only one agent has private information. We show that if the parties can make *ex ante* transfers the revelation principle does extend to this setting. However, we show that it does not extend to a setting in which more than one agent has private information.

JEL Classification No.: D23

Keywords: Revelation Principle, Commitment, Asymmetric Information

## 1 Introduction

In many contracting situations, the parties are unable fully to commit to *ex post* actions. For examples of analysis of a variety of such situations, see Laffont and Tirole (1988), Khalil (1997), Crawford and Sobel (1982), Dewatripont (1989). One difficulty with analyzing such models is that the revelation principle (Dasgupta, Hammond and Maskin, 1979) does not apply. Bester and Strausz (2001), however, showed that a variant of the revelation principle does apply in some environments without commitment. More precisely, they showed that in a model with a principal and one agent, for any incentive-efficient equilibrium of an arbitrary mechanism, there is an equilibrium of a direct mechanism in which each type of the agent gets the same payoff and the principal gets the same or more.

In a companion paper (Bester and Strausz, 2000) they showed that this result does not extend to the multi-agent case. The counterexample has two agents, only one of whom (agent 1) has private information. Agent 1 has two possible types. They exhibit an equilibrium, say  $\sigma$ , for a mechanism  $\Gamma$  with three messages for agent 1 such that any equilibrium of any two-message

---

<sup>1</sup>Faculty of Economics and St. John's College, Cambridge, UK. rael@cam.ac.uk

<sup>2</sup>Faculty of Economics, Sidgwick Avenue, Cambridge, CB3 9DD, UK. skr29@cam.ac.uk

mechanism  $\Gamma'$  gives a lower payoff either to the principal or to agent 2. Thus a mechanism with two messages cannot replicate or improve on the outcome of the mechanism with three messages.

However, in this example, one can construct an equilibrium of  $\Gamma'$  in which both types of agent 1 get the same payoff as in  $\sigma$  and the sum of the principal's and agent 2's payoff is higher. Therefore, if the players have quasi-linear utility for money and they can make *ex ante* transfers, it is possible in  $\Gamma'$  for agent 2 and each type of agent 1 to get the same payoff as in  $\sigma$  and for the principal to get more. The case of quasi-linear utility is a leading one in contract theory and it requires negligible, if any, commitment to make *ex ante* transfers.

In this paper we show that the Bester-Strausz result extends to the multi-agent case with quasi-linear utilities and *ex ante* transfers if only one agent has private information. We also examine the case in which more than one agent has private information. We provide an example with two agents, each of whom has two types, and an equilibrium  $\bar{\sigma}$  of a mechanism with three messages for agent 1 and two for agent 2, such that any equilibrium of a mechanism with two messages each has a lower total payoff than  $\bar{\sigma}$  *ex ante* (i.e., evaluated before players learn their types). Therefore it is not possible in such a mechanism to make each type of each player as well off as in  $\bar{\sigma}$ .

The conclusion is that in this setting the revelation principle fails to hold when there are at least two agents, but only if at least two agents have private information.

## 2 Model

There is a principal ( $P$ ) and 2 agents (1 and 2). 1 is privately informed about his type  $\theta_i \in \Theta = \{\theta_1, \dots, \theta_I\}$ . The probability that 1 is of type  $\theta_i$  is denoted by  $\gamma_i$  and we assume that this is common knowledge. 2 has no private information.<sup>3</sup>  $P$  can choose an action  $y \in Y$ . We allow parties to make transfers which enter into payoffs in an additively separable way. Thus, for given action  $y$ , 1's type  $\theta_i$  and transfers  $t = (t^1, t^2)$ , where  $t^j$  is the payment from  $P$  to  $j \in \{1, 2\}$ , we denote  $P$ 's payoff by  $V_i(y) - t^1 - t^2$  and the agents' payoffs by  $U_i^j(y) + t^j$ .<sup>4</sup>

A mechanism  $\Gamma$  specifies a finite message space  $M = \{m_1, \dots, m_H\}$  for

---

<sup>3</sup>The set-up can be extended to  $n$  agents, only one of whom has private information.

<sup>4</sup>We assume that parties' valuation of money is unaffected by agent 1's type. This assumption could be relaxed but it would introduce more notation.

1 and a pair of transfers  $t$ . The mechanism induces a two-stage game of incomplete information, in which 1 chooses, for each type  $\theta_i$ , a probability distribution  $q_i(\cdot)$  over  $M$ . On receipt of message  $m_h$ ,  $P$  updates his beliefs about 1's type to  $p(m_h) = (p_1(m_h), \dots, p_I(m_h))$  and then chooses an action  $y(m_h)$ . Let  $q(\cdot) = (q_1(\cdot), \dots, q_I(\cdot))$  and denote by  $\bar{q}(m_h) = \sum_i \gamma_i q_i(m_h)$  the overall probability that message  $m_h$  is sent. The profile  $(q(\cdot), p(\cdot), y(\cdot))$  constitutes a Perfect Bayesian Equilibrium of the game induced by mechanism  $\Gamma$  if (i) for each  $m_h$ ,  $y(m_h)$  maximizes  $P$ 's expected payoff given his belief  $p_i(m_h)$ , (ii) for each  $\theta_i$ ,  $q(\cdot)$  maximizes 1's expected payoff given  $y(\cdot)$ , and (iii)  $p(\cdot)$  follows Bayes' Rule whenever possible.

### 3 A Revelation Principle

For a given mechanism  $\Gamma = [M, t]$ , is it possible to replicate the same (or higher) payoffs for all parties with a direct mechanism  $\Gamma^d = [\Theta, t^d]$ ? Bester and Strausz (2001) showed that with only one agent, without transfers, the answer to this question is 'yes'. The Bester and Strausz (2000) example, however, showed that this is not necessarily true if there are at least two agents. The example has  $\Theta = \{\theta_1, \theta_2\}$ ,  $Y = \mathfrak{R}$ ,  $V_1(y) = -y^2$ ,  $V_2(y) = -(2-y)^2$ ,  $U_1^1(y) = -(0.5-y)^2$ ,  $U_2^1(y) = -(1.5-y)^2$  and  $U_i^2(y) = -10(1-y)^2$ , for  $i = 1, 2$ . They exhibit an equilibrium for an  $M$  with three messages in which  $P$  gets  $-0.5$ , each type of 1 gets  $-0.25$ , and 2 gets  $-5$  and show that no equilibrium for an  $M$  with two messages can replicate or improve on this outcome. However, with two messages, the pooling equilibrium with equal weight on each message gives  $-1$  to  $P$ ,  $-0.25$  to each type of 1 and 0 to 2, so if transfers were possible each type of each player could be made at least as well off.

In this section we show that the Bester-Strausz revelation principle extends to the set-up of section 2. The argument is closely based on theirs, but we include it here for completeness.

We say that  $\tau = (q, p, y, t|M)$  is *incentive-feasible* if  $(q, p, y)$  is a Perfect Bayesian Equilibrium of the game induced by the mechanism  $[M, t]$ . Slightly abusing notation, we denote the expected payoffs in this equilibrium by  $V(\tau)$ ,  $U_i^1(\tau)$  and  $U^2(\tau)$ .

**Definition 1**  $\tau = (q, p, y, t|M)$  is *incentive-efficient* if it is *incentive-feasible* and there exists no other *incentive-feasible*  $\tau' = (q', p', y', t'|M)$  such that  $V(\tau') > V(\tau)$ ,  $U_i^1(\tau) = U_i^1(\tau')$  for all  $i$  and  $U^2(\tau) = U^2(\tau')$ .

This is a direct extension of the notion of incentive efficiency in Bester

and Strausz (2001). Consider another definition:

**Definition 2**  $\tau = (q, p, y, t|M)$  is incentive-efficient if it is incentive-feasible and there exists no other incentive-feasible  $\tau' = (q', p', y', t'|M)$  such that  $V(\tau') + U^2(\tau') > V(\tau) + U^2(\tau)$  and  $U_i^1(\tau') = U_i^1(\tau)$  for all  $i$ .

It is easy to see that in our setting, with monetary transfers, these two definitions are equivalent. If  $\tau$  fails Definition 2 ( $P$  and agent 2 can jointly be made better off) then for an appropriate choice of transfer  $t'$  it fails Definition 1 ( $P$  can be made better off and agent 2 at least as well off).

We now argue that Definition 2 allows us to apply the same logic as in Bester and Strausz (2001) to prove the revelation principle in our set-up. The argument establishes that for any incentive-efficient  $\tau = (q, p, y, t|M)$  it is possible to find an incentive feasible  $\tau' = (q', p, y, t|M')$  with  $|M'| \leq I$  that gives the same joint payoff to the principal and agent 2 and the same payoff to each type of agent 1.

Let  $q'$  be a  $\lambda$ -perturbation of  $q$  if (a)  $q'_i(m_h) > 0$  implies that  $q_i(m_h) > 0$  and (b)  $p'(\cdot) = p(\cdot)$  where  $p'(\cdot)$  [resp.  $p(\cdot)$ ] is a belief derived from  $q'$  [resp.  $q$ ] via Bayes' rule. In that case if  $(q, p, y, t|M)$  is incentive-feasible then  $(q', p, y, t|M')$  is incentive-feasible for any  $M' \subseteq M$  containing the support of  $\bar{q}'$ . This is because, for any  $m_h$ ,  $p(m_h)$  is the correct belief for  $P$  and  $y(m_h)$  is optimal given  $p(m_h)$ , and each type  $i$  of agent 1 puts positive probability only on messages in the support of  $q_i$  (i.e. optimal messages, given  $y(\cdot)$ ). Furthermore, for each type of agent 1,  $(q, p, y, t|M)$  and  $(q', p, y, t|M')$  give the same expected payoff.

Suppose that  $\tau = (q, p, y, t|M)$  is incentive-efficient according to Definition 2, and  $|M| > I$ , with  $\bar{q}(m_h) > 0$  for each  $m_h \in M$ . Consider the following class of  $\lambda$ -perturbations of  $q$ .  $q'_i(m_h) = \lambda_h q_i(m_h)$  for all  $i$ , where

$$\lambda_h \geq 0 \text{ for all } h, \sum_h \lambda_h q_i(m_h) = 1 \text{ for all } i. \quad (1)$$

By incentive-efficiency no such perturbation can increase  $V + U^2$  and so  $\lambda = (1, \dots, 1)$  maximizes  $\sum_{m_h} \lambda_h \sum_i \gamma_i q_i(m_h) [V_i(y(m_h)) + U_i^2(y(m_h))]$  subject to (1). This implies that, for all  $m_h$  such that  $\bar{q}(m_h) > 0$ ,

$$\sum_i [V_i(y(m_h)) + U_i^2(y(m_h))] p_i(m_h) = \sum_i \tilde{\mu}_i p_i(m_h) \quad (2)$$

for some constants  $\{\tilde{\mu}_i\}_i$  (cf. Lemma 1 in Bester and Strausz (2001)).

Since  $\sum_{m_h} \bar{q}(m_h)p_i(m_h) = \gamma_i$ ,  $\gamma \in \mathfrak{R}^I$  is a convex combination of the  $|M|$  posterior beliefs  $p(m_h)$ . Therefore  $\gamma$  can be expressed as a convex combination of  $I$  of these vectors, corresponding to a subset  $M'$  of the messages with  $|M'| \leq I$ :  $\gamma = \sum_{m_h \in M'} \alpha_h p(m_h)$ . Given  $m_h \in M'$ , let  $q'_i(m_h) = (\gamma_i)^{-1} \alpha_h p_i(m_h)$  and let  $q'_i(m_h) = 0$  if  $m_h \notin M'$ .  $q'$  is a  $\lambda$ -perturbation of  $q$  since, for each  $m_h$ , each type of agent 1 re-scales its probability of  $m_h$  in the same way. Therefore  $\tau' = (q', p, y, t|M')$  is incentive-feasible and  $U_i^1(\tau') = U_i^1(\tau)$  for all  $i$ . We need to show that  $V(\tau') + U^2(\tau') = V(\tau) + U^2(\tau)$ : this will establish that, choosing an appropriate  $t'$ , we can find a PBE which is payoff-equivalent to  $\tau$  and which uses only  $I$  messages. But

$$\begin{aligned} V(\tau) + U^2(\tau) &= \sum_{m_h} \bar{q}(m_h) \sum_i [V_i(y(m_h)) + U_i^2(y(m_h))] p_i(m_h) \\ &= \sum_{m_h} \bar{q}(m_h) \sum_i \tilde{\mu}_i p_i(m_h) = \sum_{m_h} \sum_i \gamma_i q_i(m_h) \tilde{\mu}_i = \sum_i \gamma_i \tilde{\mu}_i, \end{aligned}$$

where the second equality comes from (2) and the third from Bayes' rule. Similarly,

$$V(\tau') + U^2(\tau') = \sum_{m_h} \bar{q}'(m_h) \sum_i [V_i(y(m_h)) + U_i^2(y(m_h))] p_i(m_h) = \sum_i \gamma_i \tilde{\mu}_i.$$

Hence, since we can identify  $M'$  with the set of types  $\Theta$ , we have

**Proposition** *If  $\tau = (q, p, y, t|M)$  is incentive-efficient, then there exists a direct mechanism  $\Gamma^d = (\Theta, t^d)$  and an incentive-feasible  $\tau^d = (q^d, p^d, y^d, t^d|\Theta)$  such that  $\tau$  and  $\tau^d$  are payoff-equivalent.<sup>5</sup>*

## 4 Two privately informed agents

Here we consider an example in which the agents, 1 and 2, are both privately informed about their types. There are two types,  $a$  and  $b$ , and each agent, independently, has probability 0.5 of being each type. Therefore there are four equally likely states of nature,  $aa, ab, ba$  and  $bb$ , where the first letter refers to 1's type and the second to 2's.  $P$ 's action set is  $\{y_{aa}, y_{ab}, y_{ba}, y_{bb}, y_{ca}, y_{cb}\}$  and his payoff as a function of the state of nature and his chosen action are given in the following matrix:

<sup>5</sup>Applying the Marriage Theorem (Weyl (1949)), one can show that  $q_i^d(\theta_i) > 0$ .

	$y_{aa}$	$y_{ab}$	$y_{ba}$	$y_{bb}$	$y_{ca}$	$y_{cb}$
$aa$	1	0	-1	0	$\frac{3}{4}$	0
$ab$	0	1	0	-1	0	$\frac{3}{4}$
$ba$	-1	0	1	0	$\frac{3}{4}$	0
$bb$	0	-1	0	1	0	$\frac{3}{4}$

The matrix below gives, for each type of each agent, the agent's payoff as a function of  $P$ 's action. For example, the top row is the payoff function of  $1a$ , type  $a$  of agent 1.

	$y_{aa}$	$y_{ab}$	$y_{ba}$	$y_{bb}$	$y_{ca}$	$y_{cb}$
$1a$	1	0	0	-1	$\frac{1}{2}$	$\frac{1}{2}$
$1b$	0	-1	1	0	$\frac{1}{2}$	$\frac{1}{2}$
$2a$	0	1	0	1	0	-2
$2b$	0	1	0	1	-2	0

We refer to the beliefs and strategy of  $P$  in any given equilibrium<sup>6</sup> as follows. After 1 has sent message  $m_k$  and 2 has sent message  $m_l$   $P$ 's equilibrium action, if pure, is denoted by  $y(m_k, m_l)$ , his belief (probability) that 1 is type  $a$  is  $p^1(m_k)$  and his belief that 2 is type  $a$  is  $p^2(m_l)$ . Mixed strategies of the agents are denoted by  $q_i^1$  and  $q_j^2$ : thus, for example,  $q_a^1(m)$  is the probability that type  $a$  of agent 1 sends message  $m$ . For an equilibrium  $\tau$ ,  $V(\tau), U_i^1(\tau), U_j^2(\tau)$  refer, respectively, to the expected payoffs of  $P$ , type  $i$  of 1 and type  $j$  of 2 ( $i, j \in \{a, b\}$ ). All proofs are in the Appendix.

First, consider a mechanism with no transfers,  $\Gamma = [M^1, M^2]$ , in which agent 1's message space  $M^1$  contains three messages,  $m_a, m_b$  and  $m_c$ , and agent 2's message space  $M^2$  contains two messages,  $m_a$  and  $m_b$ .

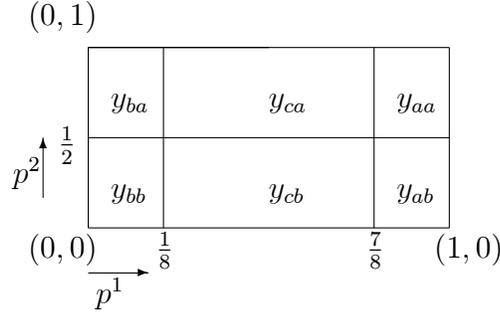
**Claim 1** *The mechanism  $\Gamma$  has an equilibrium in which  $1a$  sends messages  $m_a$  and  $m_c$  with equal probability,  $1b$  sends messages  $m_b$  and  $m_c$  with equal probability, and agent 2 separates -  $2a$  sends message  $m_a$  and  $2b$  sends message  $m_b$ . The total ex ante expected equilibrium payoff is  $\frac{13}{8}$ .*

In the remainder of the section we show that any mechanism  $\Gamma'$  that uses just two messages for each agent must, in any equilibrium, have an ex ante total payoff below  $\frac{13}{8}$ . Suppose that  $M^1 = M^2 = \{m_a, m_b\}$ . Since the

<sup>6</sup>This refers to Perfect Bayesian Equilibrium; see Definition 8.2 in Fudenberg and Tirole (1991).

meaning of messages is arbitrary and the prior is uniform, it is without loss of generality to assume that  $p^i(m_a) \geq \frac{1}{2}$  and  $p^i(m_b) \leq \frac{1}{2}$  for  $i = 1, 2$ .

The diagram below represents  $P$ 's optimal action as a function of his posterior beliefs  $(p^1, p^2)$ . For example,  $y_{aa}$  is optimal if and only if  $p^1 \geq \frac{7}{8}$  and  $p^2 \geq \frac{1}{2}$ .



First, consider equilibria in which the action is in the set  $\{y_{ca}, y_{cb}\}$  with probability 1. There are many such equilibria, with 2 pooling or separating and 1 pooling or partially separating. For example, one is as follows:  $q_a^1(m_a) = q_a^1(m_b) = q_b^1(m_a) = q_b^1(m_b) = \frac{1}{2}$ ,  $q_a^2(m_a) = q_b^2(m_b) = 1$ ,  $y(\cdot, m_i) = y_{ci}$ . In any such equilibrium, 1 gets  $\frac{1}{2}$ , 2 gets at most 0, and  $P$  gets at most  $\frac{3}{4}$ , so the total ex ante expected payoff is at most  $\frac{10}{8}$ .

Next, we look for equilibria with strictly positive probability on the set  $\{y_{aa}, y_{ab}, y_{ba}, y_{bb}\}$ . In that case,  $p^1(m_a) \in (\frac{1}{2}, 1]$  since, if  $p^1(m_a) = \frac{1}{2}$ , either  $p^1(m_b) = \frac{1}{2}$  or both types of 1 send  $m_a$  with probability 1, which would imply  $pr(\{y_{ca}, y_{cb}\}) = 1$ . We distinguish the cases  $p^1(m_a) \in (\frac{1}{2}, 1)$  and  $p^1(m_a) = 1$ .

**Claim 2** *In any equilibrium of  $\Gamma'$  with  $pr(\{y_{aa}, y_{ab}, y_{ba}, y_{bb}\}) > 0$  and  $p^1(m_a) \in (\frac{1}{2}, 1)$ ,*

- (i) *1a sends message  $m_a$  for sure and 1b sends both messages with strictly positive probability,*
- (ii) *if  $p^2(m_a) > \frac{1}{2} > p^2(m_b)$ , 2a sends message  $m_a$  for sure, 2b sends message  $m_b$  for sure and the total payoff is less than  $\frac{13}{8}$ ,*
- (iii) *if  $p^2(m) = \frac{1}{2}$  for any positive probability message  $m$ , the total payoff is less than  $\frac{12}{8}$ .*

Turning to  $p^1(m_a) = 1$ , note that by the symmetry of the game, we have already considered all cases in which  $p^1(m_b) \in (0, \frac{1}{2}]$ . So it remains to consider the case  $p^1(m_a) = 1, p^1(m_b) = 0$ .

**Claim 3** *In any equilibrium of  $\Gamma'$  with  $pr(\{y_{aa}, y_{ab}, y_{ba}, y_{bb}\}) > 0$ , if 1a sends  $m_a$  and 1b sends  $m_b$  for sure,  $p^2(m) = \frac{1}{2}$  for any positive probability message  $m$  and the total payoff is less than  $\frac{12}{8}$ .*

In the equilibrium of the three-message mechanism  $\Gamma$  in Claim 1 agent 2 separates. To achieve this separation  $P$  needs to put positive probability on the set  $\{y_{ca}, y_{cb}\}$  since agent 2's types have the same preferences over the other actions. If  $P$  cannot commit he will not play these actions unless he is unsure about agent 1's type. Hence, 1 must pool to some extent. However, pooling by 1 is costly and the optimal degree of separation requires three messages rather than two.

It is therefore the interplay between the incentives of the two privately informed agents that leads to a failure of the Revelation Principle.

## 5 Appendix

**Proof of Claim 1:** Suppose  $P$ 's beliefs and action rule are:

	$(m_a, m_a)$	$(m_a, m_b)$	$(m_b, m_a)$	$(m_b, m_b)$	$(m_c, m_a)$	$(m_c, m_b)$
$(p^1, p^2)$	(1, 1)	(1, 0)	(0, 1)	(0, 0)	$(\frac{1}{2}, 1)$	$(\frac{1}{2}, 0)$
$y$	$y_{aa}$	$y_{ab}$	$y_{ba}$	$y_{bb}$	$y_{ca}$	$y_{cb}$

Clearly the beliefs satisfy Bayes' rule given the agents' strategies and the action rule is optimal for  $P$  given the beliefs. If 1 sends message  $m_k$ ,  $k = a, b, c$ , he expects  $P$  to choose  $y_{ka}$  or  $y_{kb}$ , with equal probability. Thus, 1a gets expected payoff  $\frac{1}{2}$  from  $m_a$ ,  $-\frac{1}{2}$  from  $m_b$  and  $\frac{1}{2}$  from  $m_c$ , so is willing to randomize between  $m_a$  and  $m_c$ . The argument for 1b is symmetric. If 2 sends message  $m_l$ ,  $l = a, b$ , he expects  $P$  to choose  $y_{al}$  and  $y_{bl}$  each with probability  $\frac{1}{4}$ , and  $y_{cl}$  with probability  $\frac{1}{2}$ . Thus, 2a gets 0 from  $m_a$  and  $-\frac{1}{2}$  from  $m_b$  while 2b gets  $-1$  from  $m_a$  and  $\frac{1}{2}$  from  $m_b$ . Therefore the given strategies form an equilibrium. In each state,  $P$  is equally likely to get 1 or  $\frac{3}{4}$ , so his expected payoff is  $\frac{7}{8}$ . 2a's is 0, 2b's is  $\frac{1}{2}$  and both types of 1 get  $\frac{1}{2}$ . This establishes the claim.

**Proof of Claim 2:**

(i) If neither type of 1 sends  $m_a$  then  $p^1(m_b) = \frac{1}{2}$  and  $pr(\{y_{ca}, y_{cb}\}) = 1$ . So  $pr(1 \text{ sends } m_a) > 0$ . Similarly,  $pr(1 \text{ sends } m_b) > 0$ . If  $q_b^1(m_a) = 0$  then  $p^1(m_a) = 1$ . Hence  $q_b^1(m_a) > 0$ . Also, we know that  $q_b^1(m_b) > 0$  because  $pr(1 \text{ sends } m_b) > 0$  and by assumption  $q_b^1(m_b) \geq q_a^1(m_b)$ . Let  $U_j^i(m)$  be  $ij$ 's expected payoff from sending  $m$ , for  $i = 1, 2$  and  $j = a, b$ .

Since  $pr(\{y_{aa}, y_{ab}\} | 1 \text{ sends } m_b) = pr(\{y_{ba}, y_{bb}\} | 1 \text{ sends } m_a) = 0$ ,

$$U_a^1(m_a) \geq U_b^1(m_a) = U_b^1(m_b) \geq U_a^1(m_b)$$

with at least one inequality strict since  $pr(\{y_{aa}, y_{ab}, y_{ba}, y_{bb}\}) > 0$ . Therefore  $U_a^1(m_a) > U_a^1(m_b)$ . It follows that 1a does not send  $m_b$ .

(ii) Note that in this case  $y(m_b, m_a) = y_{ba}$  and  $y(m_b, m_b) = y_{bb}$ . First, 2b strictly prefers  $y_{bb}$  to  $y_{ba}$  and  $pr(1 \text{ sends } m_b) > 0$ . Second, 2b (weakly) prefers any mixture of  $y_{cb}$  and  $y_{ab}$  to any mixture of  $y_{ca}$  and  $y_{aa}$ . So,  $U_b^2(m_b) > U_b^2(m_a)$ . This shows that 2b only sends  $m_b$ . Second, because 1b is indifferent between  $m_a$  and  $m_b$  we know  $p^1(m_a) \leq \frac{7}{8}$ , otherwise he would prefer  $m_b$ . If  $p^1(m_a) < \frac{7}{8}$ ,  $U_b^1(m_a) = \frac{1}{2}$  and therefore also  $U_b^1(m_b) = \frac{1}{2}$ . This implies that 2a must send  $m_a$  for sure. If  $p^1(m_a) = \frac{7}{8}$ , by Bayes' rule we must have  $q_b^1(m_a) = \frac{1}{7}$  and  $q_b^1(m_b) = \frac{6}{7}$ . Because  $U_a^2(m_a) = 0$  we need  $U_a^2(m_b) = 0$  for 2a to be mixing between  $m_a$  and  $m_b$ . Therefore, after message pair  $(m_a, m_b)$   $P$  must put weights  $\frac{7}{12}$  and  $\frac{5}{12}$  on actions  $y_{cb}$  and  $y_{ab}$  respectively. But this contradicts that 1b is indifferent between  $m_a$  and  $m_b$  because  $U_b^1(m_b) = \frac{1+\gamma}{2}$ , where  $\gamma = q_a^2(m_b)$ , and  $U_b^1(m_a) \leq \frac{1+\gamma}{2}((\frac{7}{12})(\frac{1}{2}) - (\frac{5}{12})) + (\frac{1-\gamma}{2})(\frac{1}{2})$ . This shows that 2a only sends  $m_a$ . It follows that  $U_b^1(m_b) = U_b^1(m_a) = \frac{1}{2}$ , which is possible only if  $p^1(m_a) \leq \frac{7}{8}$ ,  $y(m_a, m_a) = y_{ca}$  and  $y(m_a, m_b) = y_{cb}$ . Let  $\alpha = q_b^1(m_a)$ . By Bayes' rule,  $\alpha \geq \frac{1}{7}$ . So we have  $U_a^2 = 0$ ,  $U_b^2 = \frac{1}{2}(1 - \alpha)$  and  $V = \frac{1}{2}(\frac{3}{4}) + \frac{1}{2}[(1 - \alpha)(1) + \alpha(\frac{3}{4})]$ . The total expected payoff is  $\frac{13}{8} - \frac{3\alpha}{8} < \frac{13}{8}$ .

(iii) Since 2's strategy is pooling,  $P$ 's action is uncorrelated with 2's type, so  $P$  cannot get a payoff of more than  $\frac{1}{2}$ . Let  $pr(y_{bb}|1 \text{ sends } m_b) = p_{bb}$ , so  $pr(y_{ba}|1 \text{ sends } m_b) = 1 - p_{bb}$ . Also, let  $pr(y_{aa}|1 \text{ sends } m_a) = p_{aa}$  and  $pr(y_{ab}|1 \text{ sends } m_a) = p_{ab}$ .

Then  $U_a^1 = \frac{1}{2}(1 - p_{aa} - p_{ab}) + p_{aa}$  and  $U_b^1 = \frac{1}{2}(1 - p_{aa} - p_{ab}) - p_{ab}$  (since 2b sends  $m_a$  with positive probability in equilibrium). Therefore 1's ex ante expected payoff is  $\frac{1}{2} - p_{ab}$ . Conditional on 1 sending  $m_b$ ,  $U_a^2 = U_b^2 = p_{bb}$ . Conditional on 1 sending  $m_a$ ,  $U_a^2 \leq p_{ab}$  and  $U_b^2 \leq p_{ab}$ . Letting  $q_b^1(m_b) = \beta$ , 2's expected payoff is therefore at most  $\frac{1}{2}\beta p_{bb} + [\frac{1}{2}(1 - \beta) + \frac{1}{2}]p_{ab}$ . Thus the sum of 1's and 2's ex ante expected payoffs is at most  $\frac{1}{2} + (\frac{\beta}{2})(p_{bb} - p_{ab}) \leq \frac{1}{2} + \frac{\beta}{2}$ . It follows that the sum of expected payoffs is less than  $\frac{12}{8}$ .

**Proof of Claim 3:** Suppose first that  $p^2(m_a) > \frac{1}{2} > p^2(m_b)$ . Then  $y(m_b, m_a) = y_{ba}$  and  $y(m_b, m_b) = y_{bb}$ . In that case both types of 2 strictly prefer  $m_b$  so  $p^2(m_b) = \frac{1}{2}$  (contradiction). Therefore  $p^2(m) = \frac{1}{2}$  for any positive probability  $m$ . As in the previous proof,  $P$  gets at most  $\frac{1}{2}$ . Using the same notation as in the previous proof, 1a gets  $p_{aa}$ , 1b gets  $p_{ba} = 1 - p_{bb}$ , 2 gets  $\frac{1}{2}p_{bb} + \frac{1}{2}p_{ab} = \frac{1}{2}p_{bb} + \frac{1}{2}(1 - p_{aa})$ . Therefore the total expected payoff is at most  $\frac{3}{2}$ .

## 6 References

Bester, H., Strausz, R., 2000 Imperfect commitment and the revelation princi-

ple: the multi-agent case. *Economics Letters* 69, 165-171.

Bester, H., Strausz, R., 2001 Contracting with imperfect commitment and the revelation principle: the single agent case. *Econometrica* 69, 1077-1098.

Crawford, P., V., Sobel, J., 1982 Strategic information transmission. *Econometrica* 50, 1431-1451.

Dasgupta, P., Hammond, P., Maskin, E., 1979 The implementation of social choice rules. review of *Economic Studies* 46, 185-216.

Dewatripont, M., 1989 Renegotiation and information revelation over time: The case of optimal labor contracts. *Quarterly Journal of Economics* 104, 589-619.

Fudenberg, D., Tirole, J., 1991 *Game Theory*, MIT Press

Khalil, F., 1997 Auditing without commitment. *Rand Journal of Economics* 28, 629-640.

Laffont, J.-J., Tirole, J., 1988 The dynamics of incentive contracts. *Econometrica* 56(5), 1153-1175.

Weyl, H., 1949 Almost periodic invariant vector sets in a metric vector space. *American Journal of Mathematics*, 71, 178-205.