Reprinted from JOURNAL OF ECONOMIC THEORY
All Rights Reserved by Academic Press, New York and London

Vol. 62, No. 1, February 1994
Printed in Belgium

Destroying the "Pretending" Equilibria in the Demski–Sappington–Spiller Model*

RUDOLF KERSCHBAMER

Department of Economics, University of Vienna,
A-1010 Vienna, Austria

Received April 17, 1991; revised January 25, 1993

Demski, Sappington, and Spiller [*J. Econ. Theory* **44** (1988), 156–167] explore the effects of bankruptcy constraints on incentive schemes when two risk-neutral agents operate in correlated environments. For a setting in which the agents receive their private information before contracts are agreed upon they claim that the truthtelling equilibrium induced by a pair of optimal direct schemes is necessarily subgame dominated. The present paper shows how the reward functions in the Demski–Sappington–Spiller revelation schemes can costlessly be modified so as to destroy all "pretending" equilibria which dominate the truth-telling one. *Journal of Economic Literature* Classification Numbers: C72, D74, D82.

1. INTRODUCTION

In [2] Demski, Sappington, and Spiller (henceforth D-S-S) examine a situation in which bankruptcy considerations (or liability constraints) limit the penalty a principal can impose on two risk-neutral agents working in correlated environments. Their main concern is the problem of multiple equilibria in the subgame defined by a pair of optimal direct (revelation) schemes.

D-S-S's work offers two major conclusions. First, if the principal and the agents agree on *contracts after the agents have received their private information* the truthtelling equilibrium induced by a pair of second best revelation schemes is necessarily subgame dominated; that is, there is another, non-truthful equilibrium that provides a higher type-contingent expected utility to both agents under these contracts. This finding is

* I am particularly grateful to an associate editor and to an anonymous referee for useful comments on an earlier draft. I also thank A. Van der Bellen, G. Clemenz, E. Dierker, and N. Maderner for helpful discussions and the Austrian Scientific Research Fund for general support. Of course, the usual disclaimer applies.

0022-0531/94 $6.00

qualitatively identical to an earlier finding of Demski and Sappington [1] for the case where the agents are risk averse. Second, if the principal and the agents agree on *contracts before agents receive their private information* truthtelling can be implemented via direct schemes as a subgame equilibrium that is undominated.[1] This finding stands in contrast to earlier findings for the case where the agents are averse to risk.

The purpose of the present paper is to modify the conclusion for the setting with asymmetric precontractual information. We show that by rearranging the reward functions in the contracts constructed by D-S-S it is always possible to implement the desired strategies in an undominated equilibrium while preserving each agent's interim utility and the principal's expected surplus for the supposed subgame behavior.

The assumption of risk neutrality is central for our finding. It implies that putting more structure on the wagelottery intended for a good type agent causes no welfare loss in terms of risk bearing. This does not hold for a setting with (strictly) risk-averse agents. There, if the good type agent is asked to bear risk, he requires a higher expected income.

D-S-S's contention for the setting with symmetric precontractual information—that different qualitative conclusions regarding the issue of multiple equilibria emerge according to whether risk aversion or limited liability restrictions prevent the attainment of the first best solution—therefore generalizes to the case in which agents receive their private information before contracting with the principal.

The D-S-S Model

D-S-S consider the following agency problem between a principal and two agents A and B. The principal owns two units of a productive technology. Technology i ($= A, B$) requires as an input the effort ($a^i \in \mathbb{R}^+$) of agent i. Each agent's effort, together with the realization of a random

[1] It should be noted that this conclusion (D-S-S's Proposition 3) is based on a characterization result for second best revelation schemes (their Proposition 2) whose range of application is vacuous. To see this, compare the bound on the maximal punishment M in D-S-S's Proposition 1(b) (the fulfillment of this restriction ensures the attainment of the first best solution) with the assumption D-S-S make for M. It can be demonstrated, however, that the D-S-S result holds for a slightly modified scenario: If one replaces their restriction on M (i.e., $M^i < \bar{U}^i$) by the condition $M^i < \bar{U}^i + D^i(x_1^*, \theta_1)$, the ideal outcome for the principal is no longer feasible in every instance; with binding bankruptcy constraints second best revelation schemes exhibit the features listed in D-S-S's Proposition 2, and Proposition 3 has a sound basis.

variable ($\theta^i \in \Theta^i = \{\theta^i_1, \theta^i_2\}$) determines output ($x^i \in \mathbb{R}^+$) according to the known relationship $x^i = X^i(a^i, \theta^i)$. Each agent's output is observable by all parties. Effort input in technology i is privately observed by agent i ($= A, B$).

Random variables θ^A and θ^B are drawn from the known distribution $\phi(\cdot)$ on Θ, where $\Theta = \Theta^A \times \Theta^B$. θ^A and θ^B are positively correlated. That is, defining $p^i_k \equiv \text{Prob}(\theta^j_1 \mid \theta^i_k)$ $(i, j \in \{A, B\};\ i \neq j;\ k \in \{1, 2\})$, it is assumed that $1 > p^i_1 > p^i_2 > 0$ for $i = A, B$. Before a contract is signed, agent i (alone) learns the actual realization of θ^i.

Each agent's reservation utility level $\bar{U}^i$ is common knowledge, as is his utility function $U^i = R^i - V^i(a^i)$, where R^i is the reward from the principal to agent i. There is a minimum reward M^i ($M^i < \bar{U}^i$) that agent i must receive, provided he abides by the terms of his contract.

Given knowledge of θ^i, an agent's effortchoice is equivalent to a choice of x^i. The agent's utility functions can therefore be rewritten as $U^i = R^i - D^i(x^i, \theta^i)$, where the last term is the disutility incurred by agent i in producing x^i when his technology is characterized by θ^i. It is assumed that $D^{i'}(\cdot, \theta^i) > 0$ and $D^{i''}(\cdot, \theta^i) > 0$ for all θ^i and all $x^i > 0$ and that $D^i(x^i, \theta^i_1) > D^i(x^i, \theta^i_2)$ and $(D(x^i, \theta^i_1)/D(x^i, \theta^i_2))' \geq 0$ for all $x^i \geq 0$, where primes denote partial derivatives with respect to x^i.[2] The principal maximizes expected output of the agents net of payments made to them, $U^P = x^A + x^B - R^A - R^B$.

3. Truthtelling as a Subgame-Undominated Equilibrium

D-S-S in their Proposition 4 partially characterize the solution to the principal's maximization problem

$$\underset{R, x}{\text{Max}} \sum_{k=1}^{2} \sum_{l=1}^{2} \phi_{kl} \cdot (x^A_k + x^B_l - R^A_{kl} - R^B_{lk})$$

subject to the constraints of Bayesian incentive compatibility

$$p^i_k R^i_{k1} + (1 - p^i_k) R^i_{k2} - D^i(x^i_k, \theta^i_k)$$
$$\geq p^i_k R^i_{l1} + (1 - p^i_k) R^i_{l2} - D(x^i_l, \theta^i_k) \qquad \forall k, l \in \{1, 2\};\ k \neq l;\ \forall i \in \{A, B\},$$

[2] The properties of optimal (second best) revelation schemes that D-S-S record in their Proposition 4 do not necessarily follow from their own set of assumptions. (See our discussion in footnote 4). To preserve the D-S-S characterization result we have replaced their (standard) "single crossing" condition $[D^{i'}(x^i, \theta^i_1) > D^{i'}(x^i, \theta^i_2)\ \forall x^i > 0]$ by the somewhat stronger condition $(D^i(x^i, \theta^i_1)/D^i(x^i, \theta^i_2)) \geq 0\ \forall x^i \geq 0$.

individual rationality

$$p_k^i R_{k1}^i + (1-p_k^i) R_{k2}^i - D(x_k^i, \theta_k^i) \geqslant \bar{U} \qquad \forall k \in \{1,2\},\ \forall i \in \{A,B\}$$

and limited liability

$$R_{kl}^i \geqslant M \qquad \forall k, l \in \{1,2\}\ \forall i \in \{A,B\}.^3$$

A solution consists of a pair of revelation schemes (one for each agent) each defining a *decision function* (specifying an output level for each type of the agent, $x^i: \theta^i \to \mathbb{R}^+$) and a *reward function* (fixing the transfers from the principal to the agent for each vector of observed outputs, $R^i: X \to \mathbb{R}$).

In Proposition 5 D-S-S claim that given a pair of optimal revelation schemes there is necessarily another pair of equilibrium strategies for the agents beyond that which the principal wants to implement. The associated payoffs leave each type of each agent strictly better off and the principal strictly worse off relative to the truthtelling outcome.

In their arguments leading to Proposition 5 D-S-S focused exclusively on reward structures where an agent is "penalized" for being the only one to report low productivity ($R_{12}^i < R_{11}^i$) but receives a reward independent of the report by the rival when reporting high productivity ($R_{21}^i = R_{22}^i$). While such wage structures are *uniquely optimal* when the agents are strictly income *risk averse* (see Demski and Sappington [1]) they are *not* when the agents are *risk neutral but protected by limited liability*. In this case the principal's maximization problem yields a uniquely defined decision function for each agent ($x_1^i < x_1^{*i} < x_2^{*i} = x_2^i$ for $i = A, B$; the asterisk refers to first best policy)[4] but it does not give reward functions. It only specifies conditions reward functions have to satisfy in order to be in the maximal set.

[3] Here x_k^i is the output level required from agent i if he announces type θ_k^i; R_{kl}^i is the reward for agent i when $x^i = x_k^i$ and $x^j = x_l^j (i, j \in \{A,B\};\ i \neq j;\ k, l \in \{1,2\})$; and $\phi_{kl} = \phi(\theta^A = \theta_k^A, \theta^B = \theta_l^B)$.

[4] Note that our modified single crossing condition is needed to produce this result. If this assumption is violated it is quite possible to get $x_1^i \geqslant x_1^{*i}$. This surprising possibility arises from the fact that in the D-S-S model two analytically different effects come into play. On one side a decrease in production from the first best level reduces the rents to good type agents and thereby produces a first-order gain for the principal. In the neighborhood of x_1^{*i}, the cost of such a deviation in terms of reduced total surplus is of second-order importance. This is the well-known "incentive compatibility effect." A second effect—we call it the "penalty effect"—arises because an increase in production (with an offsetting increase in the payments R_{1k}) relaxes the binding bankruptcy constraint, enabling the principal to design "better" lotteries. Even though the incentive compatibility effect causes production to decline, the "perverse" penalty effect may be sufficiently strong to make the total effect positive. (See Kerschbamer [3] for a discussion.)

These conditions are (we focus on the reward function for one agent and delete the i superscript):[5]

$$\text{(i)} \quad R_{11} = \frac{\bar{U} + D(x_1, \theta_1) - (1 - p_1)M}{p_1}; \qquad R_{12} = M; \qquad R_{21}, R_{22} \geqslant M;$$

$$\text{(ii)} \quad p_2(R_{21} - R_{11}) + (1 - p_2)(R_{22} - R_{12}) = D(x_2, \theta_2) - D(x_1, \theta_2);$$

$$\text{(iii)} \quad p_1(R_{21} - R_{11}) + (1 - p_1)(R_{22} - R_{12}) \leqslant D(x_2, \theta_1) - D(x_1, \theta_1).$$

It is straightforward to verify that second best incentive schemes with the property $R_{21} \leqslant R_{11} + D(x_2, \theta_2) - D(x_1, \theta_2)$ have truthtelling as a subgame-dominated equilibrium. Among the jeopardizing equilibria is the one discussed by D-S-S in which both agents always claim to be unproductive. However, by putting a more sophisticated structure on the agents' rewards the principal is always able to destroy these equilibria without changing the payoffs in the truthtelling equilibrium for any of the parties involved:

PROPOSITION. *For any truthtelling equilibrium of an optimal direct mechanism which is subgame dominated from the agents' perspective there exist modified reward functions which produce truthtelling as a subgame-undominated equilibrium. The modification in reward functions leaves the expected payoffs in the truthtelling equilibrium for the principal and for the agents unaffected.*

Proof. To prevent the jeopardizing equilibria we set $R_{21} = R_{11} + D(x_2, \theta_2) - D(x_1, \theta_2) + \varepsilon$ with $\varepsilon > 0$. Solving for R_{22} yields $R_{22} = R_{12} + D(x_2, \theta_2) - D(x_1, \theta_2) - (p_2/(1 - p_2))\varepsilon$. Of course, R_{21} should not be too large, otherwise other restrictions are violated. Specifically, we assume that ε is chosen sufficiently small to satisfy $R_{21} < R_{11} + D(x_2, \theta_1) - D(x_1, \theta_1)$ and $R_{22} > M$. The first of these upper bounds yields $\varepsilon < [(D(x_2, \theta_1) - D(x_1, \theta_1)) - (D(x_2, \theta_2) - D(x_1, \theta_2))]$, the second $\varepsilon < ((1 - p_2)/p_2)[D(x_2, \theta_2) - D(x_1, \theta_2)]$. Since both terms in brackets are strictly positive an ε consistent with these requirements exists. Such an ε satisfies also the restriction $p_1(R_{21} - R_{11}) + (1 - p_1)(R_{22} - R_{12}) \leqslant D(x_2, \theta_1) - D(x_1, \theta_1)$. The resulting

[5] It is straightforward to verify that reward functions exhibiting these features produce participation and truthful reporting as an equilibrium. It is also obvious that if both agents use their truthful strategies these schemes all generate the same expected payoff for the principal (and the same interim utility for each agent). The proof for the claim that reward schemes with these properties maximize the principal's objective function is available from the author upon request.

reward matrix is therefore a solution to the principal's maximization problem.[6]

We are now in the position to prove that for such a reward matrix truthtelling is a subgame-undominated equilibrium.

An agent participating in the game has the following strategies: (a) honestly revealing his type; (b) always lying; (c) always claiming to be unproductive; (d) always claiming to be productive; (e) some convex combination of strategies (a)–(d). We make the following observations:

- *For an agent observing θ_1 lying is strictly dominated.* This follows immediately from the fact that the proposed incentive schemes have $R_{21} - R_{11} < D(x_2, \theta_1) - D(x_1, \theta_1)$ and $R_{22} - R_{12} < D(x_2, \theta_2) - D(x_1, \theta_2)$ and from $[D(x_2, \theta_1) - D(x_1, \theta_1)] > [D(x_2, \theta_2) - D(x_1, \theta_2)]$ $\forall (x_1, x_2)/x_2 > x_1$.

- *In each equilibrium, at least one agent plays the pure strategy a.* After elimination of stategies b and d there remain the two pure strategies a and c and a convex combination between a and c. Suppose one agent plays strategy c with some strictly positive probability. Then the probability for the second agent to get payment R_{21} when he produces x_2 (we denote this probability by p_2^0) is increased ($p_2^0 > p_2$). Since the incentive schemes in question have $p_2 R_{21} + (1 - p_2) R_{22} - D(x_2, \theta_2) = p_2 \quad R_{11} + (1 - p_2) R_{12} - D(x_1, \theta_2)$ and $R_{21} - D(x_2, \theta_2) > R_{11} - D(x_1, \theta_2)$ it follows that $p_2^0 R_{21} + (1 - p_2^0) R_{22} - D(x_2, \theta_2) > p_2^0 R_{11} + (1 - p_2^0) R_{12} - D(x_1, \theta_2)$. Hence, the second agent's best response to strategy c and to a convex combination of a and c is to truthfully reveal his type.

- *If in each equilibrium at least one agent uses his truthful strategy, then the truthtelling equilibrium cannot be subgame dominated.* This follows from the definition of a (Bayesian) equilibrium strategy and from the definition of a subgame-dominated equilibrium. ▮

The *intuition underlying our result* is fairly straightforward. For the principal the setting in which agents are income-risk neutral but protected by limited liability is equivalent to a situation in which agents' utility exhibit an artificial concavity of the following form: agent i's utility is $-\infty$ if $R^i < M^i$; it is linear in income if $R^i \geqslant M^i$. The linearity in the upper tail

[6] It is interesting to note that we could also modify the reward functions so that truthtelling becomes a dominant strategy for each agent without changing the expected payments the principal must make. The resulting reward structure ($R_{21} = R_{11} + D(x_2, \theta_2) - D(x_1, \theta_2)$; $R_{22} = R_{12} + D(x_2, \theta_2) - D(x_1, \theta_2)$) has, however, again the property that aside from honest reporting there is the untruthful equilibrium where both agents always claim to be unproductive. As before, the associated outcome leaves each type of each agent strictly better off and the principal strictly worse off relative to the truthful equilibrium outcome. The explanation for this seemingly paradoxical result is the inherent interest conflict between the principal and the agents and the fact that the dominant strategies the principal wishes to implement form only a weak equilibrium point. (See Kerschbamer [3] for a discussion on this point.)

of the utility functions allows the principal to rearrange the reward lottery intended for a good type agent without altering its expected value for the supposed equilibrium behavior. Roughly speaking our rearrangement works such that each agent is not only "penalized" for being the only one to reveal the "bad" type ($R_{12} < R_{11}$) but also "rewarded" for being the only to report the "good" productivity parameter ($R_{22} < R_{21}$). The reward for "exceptional" performance is chosen so that a good type agent who expects that his rival will underreport his productivity with (strictly) positive probability has (strictly) positive incentives to truthfully reveal his type ($R_{21} > R_{11} + D(x_2, \theta_2) - D(x_1, \theta_2)$). This eliminates the attraction to the agents of jointly adopting strategies other than truthtelling.

4. Related Work

Ma, Moore, and Turnbull [4] (henceforth M-M-T) have shown that by considering non-revelation (or indirect) schemes the multiple equilibria problem in the risk-averse agents scenario of Demski and Sappington [1] can be solved.

The basic idea underlying the M-M-T construction is to use one agent as a "sneak." The sneak receives a set of additional "non-type" messages. The rewards for using these messages are structured in a way so that their use is profitable if (and only if) the second agent chooses undesired reporting strategies. If a non-type message is observed the second agent is penalized. Although non-type messages are never used in equilibrium their presence suffices to "stop agents from cheating".

In one respect the technique introduced by M-M-T resembles the one applied here: Both use carefully designed payoffs for "out of equilibrium" strategies to reduce the set of equilibria in the subgame played among agents. There is one key difference: In the M-M-T setting with risk-averse agents the principal has to choose between two costly solutions to the multiplicity problem. He can incur the explicit cost of a higher wagebill in the context of direct schemes or the implicit cost of an augmented message space. In the risk-neutral agent case considered here there is no need to accept any additional costs: fine-tuned wage lotteries in optimal direct schemes implement the second best as a subgame-undominated equilibrium without welfare loss.

5. Conclusion

It is by now well known that binding liability restrictions and risk aversion on the part of agents cause qualitatively *similar results concerning*

the principal's trade-off in optimal direct schemes. The present paper shows that there is definitely a *contrast* between these settings *regarding the issue of multiple equilibria* in these schemes: While the reward structure required to destroy undesired pretending equilibria causes a welfare loss in terms of risk bearing when agents are averse to risk, there is no such welfare loss if agents are risk neutral but protected by limited liability.

REFERENCES

1. J. DEMSKI AND D. SAPPINGTON, Optimal incentive contracts with multiple agents, *J. Econ. Theory* **33** (1984), 152–171.
2. J. DEMSKI, D. SAPPINGTON, AND P. SPILLER, Incentive schemes with multiple agents and bankruptcy constraints, *J. Econ. Theory* **44** (1988), 156–167.
3. R. KERSCHBAMER, "Multiple Equilibria Problems in Incentive Contracts with Many Agents and Bankruptcy Constraints," Working paper No. 9101, Department of Economics, University of Vienna, March 1991.
4. C. MA, J. MOORE, AND S. TURNBULL, Stopping agents from "cheating," *J. Econ. Theory* **46** (1988), 355–372.

Printed by Catherine Press, Ltd., Tempelhof 41, B-8000 Brugge, Belgium