Supplemental Appendix

In this supplement I provide a formal argument for Lemma 1, which stated that non-uniqueness of the SPNE outcome is endemic to any mechanism for which we require both (i) implementation of the first-best outcome and (ii) no collusion in equilibrium (where by 'collusion' we mean the outcome described in Definition 1). Let $\mathcal{E}$ denote the economic environment, i.e. two agents, two effort choices, two output values, and the public signals being the value of output and the messages that the agents send to the principal. Let $\mathcal{C}(\mathcal{E})$ denote a generic wage contract for this environment (satisfying symmetry and limited liability). Normalize effort costs so that $c(e_L) = 0$ and let ϵ denote the (arbitrarily small) liability bound.

Lemma 1. *For any contract $\mathcal{C}(\mathcal{E})$, let $\mathcal{G}$ denote the game form induced by the contract and let $\Sigma(\mathcal{G})$ be the set of SPNE. If the efficient outcome (i.e. high effort and full insurance) is an outcome of some equilibrium in $\Sigma(\mathcal{G})$ and no equilibrium outcomes exhibit collusion, then there must be multiple equilibrium outcomes.*

Proof. There are two physical points in time. First, the effort choice time, say t_1. Second, the time where output is realized, time t_2. Assume that $t_1 < t_2$. In our setting, I will say that an (extensive) game form, $\mathcal{G}$, is comprised of the following:

- A finite set of messages, $\mathcal{M} := \{m_1, \ldots, m_k\}$.
- A wage contract for each agent i, $w_i : X \times \prod_i \mathcal{M}_i \to \mathbf{R}$, i.e. a contract is defined by assigning a number to each realized output value and profile of messages.
- The wage contract is symmetric: $w_i(\cdot, (m^1, m^2)) = w_{-i}(\cdot, (m^2, m^1))$.
- Agents have two (pure) choice sets, at two separate points in time:
  - One of these is the effort choice, $e \in \{e_H, e_L\}$, at time time t_1.
  - The second (pure) choice set is the message space and is determined by a timing structure of when messages are sent. Let $t_{\mathcal{M}}$ denote this time and assume that we have one of the possibilities, (i) $t_{\mathcal{M}} \in (0, t_1)$, (ii) $t_{\mathcal{M}} \in (t_1, t_2]$, (iii) $t_{\mathcal{M}} \in (t_2, \infty)$.
- Agents have perfect information, so that each agent's information set at time $t_{\mathcal{M}}$ is a singleton node.

This completes the description of the abstract game form $\mathcal{G}$.[1] Let $\Sigma(\mathcal{G})$ denote the set of SPNE in the extensive form game induced by $\mathcal{G}$. For $\sigma \in \Sigma(\mathcal{M})$ let m_σ denote the profile of messages induced by σ. Abusing notation, let it also denote a distribution over profiles if σ is mixed. We wish to find $\mathcal{G}$ satisfying two desiderata:

[1]Note that we have left the message protocol, e.g. sequential, simultaneous, etc., unspecified. The result of the lemma does not depend on the message protocol of the game form.

- There is some $\sigma_{FB} \in \Sigma(\mathcal{G})$ such that at the time t_1 all agents choose e_H and at time $t_{\mathcal{M}}$ agents choose some profile $\{m^i\}_i$ such that $w(\cdot, \{m^i\}_i)$ is constant and equals (in utility space, for simplicity) $u_0 + c(e_H) + \epsilon'$, for some small ϵ'.

- There is no $\sigma \in \Sigma(\mathcal{M})$ such that $\sigma_i \neq e_H$ for some i and $u_j(\sigma) > u_0$ for some j, i.e. no collusion in equilibrium.

The claim of the lemma is that any game form $\mathcal{G}$ that satisfies these two properties must admit multiple (SPNE) equilibria. Consider game forms where $t_{\mathcal{M}} \in (t_2, \infty)$ or $t_{\mathcal{M}} \in (0, t_1]$. First consider the former case. By hypothesis, there is a message pair (m^1, m^2) where agents (in the subgame following the effort choices $e^1 = e_H, e^2 = e_H$ and $x \in \{x_H, x_L\}$) receive a constant wage $w(\cdot, (m^1_{\sigma_{FB}}, m^2_{\sigma_{FB}}))$ and do not deviate from this profile of messages. Since there is no collusion in equilibrium, if agent 1, say, selects $e^1 = e_L$, then agent 2 has a strictly profitable deviation to, say, m'. If $w_2(\cdot, (m^1, m'))$ is constant, then agent 2 would deviate from the message profile (m^1, m^2) – contradicting that this is an equilibrium. Hence, $w_2(\cdot, (m^1, m'))$ is non-constant. Since switching to m' is a strictly profitable deviation, it must be the case that for some realization $x \in \{x_L, x_H\}$, 2 gets more than $w_2(\cdot, (m^1, m^2))$. This means that he deviates from m^2 following a realization of x even when both players select high effort, again contradicting that this is an equilibrium.

Now consider $\mathcal{M}$ where $t_{\mathcal{M}} \in (0, t_1]$. First consider $t_{\mathcal{M}} \in (0, t_1)$. The fact that $\sigma_{FB} \in \Sigma(\mathcal{G})$ implies that, letting $m_{\sigma_{FB}}$ be the message profile component of σ_{FB}, the pair $e^1 = e_H = e^2$ is a N.E. in the subgame following any realized message pair (m^1, m^2) in the support of $m_{\sigma_{FB}}$. But since $w(\cdot, m_{\sigma_{FB}})$ is constant[2], in the subgame following (m^1, m^2) both players will deviate from high effort. If $t_{\mathcal{M}} = t_1$, then if player 2, say, anticipates that player 1 plays $e^1 = e_H$ and sends message m^1, where m^1 is player 1's message in the putative efficient SPNE, he will deviate to $e^2 = e_L$ and send m^2. Hence, we cannot have $t_{\mathcal{M}} \in (0, t_1]$ and maintain both desiderata.

Finally, consider $t_{\mathcal{M}} \in (t_1, t_2]$. The analysis is the same for times $t_{\mathcal{M}} \in (t_1, t_2)$ as for $t_{\mathcal{M}} = t_2$, since in the latter case the agent evaluates expected payoffs as he would if output is realized just after messages are chosen. Assume $w(\cdot, \cdot)$ (weakly) implements the efficient outcome (at first-best cost) and prevents collusion. We show the existence of another (inefficient) SPNE. For the subgames induced by (e_H, e_L) and (e_L, e_H) consider the following two equilibria. Let $\vec{m}_{(e_H, e_L)}$ denote the worst equilibrium for player 1 and let $\vec{m}_{(e_L, e_H)}$ denote the worst equilibrium for player 2. By symmetry of $w(\cdot, \cdot)$ player 2's (gross) payoff is weakly higher than player 1's in equilibrium $\vec{m}_{(e_H, e_L)}$ and, similarly, player 1's gross payoff is higher than player 2's in the equilibrium $\vec{m}_{(e_L, e_H)}$.[3] Let $\vec{m}_{(e_H, e_H)}$ denote the message profile in the first-best SPNE and let $\vec{m}_{(e_L, e_L)}$ denote any equilibrium in the subgame (e_L, e_L). Now

[2] By definition $w(\cdot, (m^1, m^2))$ is constant for any (m^1, m^2) in the support of $m_{\sigma_{FB}}$.

[3] Notice that with a bivariate signal (as in Ma (1988)) we would never use such a contract, since

use these message profiles to construct the following extensive form profile: Each player i chooses $e_i = e_L$ and sends message $\vec{m}^i_{(e_1,e_2)}$ (player i's component of $\vec{m}_{(e_1,e_2)}$) when (e_1, e_2) is observed. Denote this profile as $\{\sigma_1(e_L, e_L), \sigma_2(e_L, e_L)\}$. Similarly use the message profiles to construct extensive form profiles $\{\sigma_1(e_H, e_L), \sigma_2(e_H, e_L)\}$ and $\{\sigma_1(e_L, e_H), \sigma_2(e_L, e_H)\}$.

I claim that $\{\sigma_1(e_L, e_L), \sigma_2(e_L, e_L)\}$ is an SPNE. Towards contradiction, if it is not, then player 1, say, wishes to deviate to $e_1 = e_H$. Let $u^1_{(e_1,e_2)}$ denote his gross utility payoff under this profile when choices (e_1, e_2) are observed. Plug in (e_H, e_L) to get: (note that $u^i_{(e_1,e_2)} \geq u_0 - \epsilon$, where ϵ is the liability bound)

$$u^1_{(e_H,e_L)} - c(e_H) > u^1_{(e_L,e_L)}$$

Note that in the subgame (e_H, e_L) player 2 obtains $u^2_{(e_H,e_L)}$. Since $u^2_{(e_H,e_L)} \geq u^1_{(e_H,e_L)}$ and

$$u^1_{(e_H,e_L)} > u^1_{(e_L,e_L)} + c(e_H) \geq \min(u^1_{(e_L,e_L)}, u^2_{(e_L,e_L)}) + c(e_H)$$

we obtain that

$$u^2_{(e_H,e_L)} \geq u_0 - \epsilon + c(e_H)$$

Hence, consider the profile $\{\sigma_1(e_H, e_L), \sigma_2(e_H, e_L))\}$. I claim this must then be a SPNE. We have just argued that player 1 does not deviate to $e_1 = e_L$. Moreover, player 2 obtains (net) payoff of at least $u_0 - \epsilon + c(e_H)$ if he plays $e_2 = e_L$. If he switches to e_H his payoff net of $c(e_H)$ is $u_0 + \epsilon'$. Hence, $\{\sigma_1(e_H, e_L), \sigma_2(e_H, e_L)\}$ is an inefficient SPNE where collusion occurs – contradicting the hypothesis that the initial contract is collusion proof. Similarly, if player 2 has a deviation from $\{\sigma_1(e_L, e_L), \sigma_2(e_L, e_L)\}$, then we argue that $\{\sigma_1(e_L, e_H), \sigma_2(e_L, e_H)\}$ is an SPNE with collusion. It follows that $\{(\sigma_1(e_L, e_L), \sigma_2(e_L, e_L)\}$ is an SPNE, with both players obtaining u_0 in equilibrium by individual rationality. □

Let us make three remarks about the preceding argument. First, note that the argument does not use the FOSD assumption. This is because we are assuming the existence of a collusion-proof contract with an efficient SPNE, and taking this as given we prove multiplicity of SPNE. The FOSD assumption is required to actually construct such a contract. Second, the only place we use the simplicity of the $2 \times 2 \times 2$ model is in the last step. In particular, with more than two effort choices the set of possible deviations is larger – so that a distinct, but otherwise straightforward, argument is required. Third, we have not proved that the symmetry assumption is necessary for the result, but it is (in our view) a normatively reasonable feature since there is only a single public signal and conditional distributions satisfy symmetry, e.g.. $f(\cdot|(e_H, e_L)) = f(\cdot|(e_L, e_H))$.

it allows an equilibrium where the shirker does better than the 'worker', i.e. the agent who chooses e_H. We have such an outcome in our setting because, with a sole public signal, the principal cannot statistically distinguish the worker from the shirker.