Web Appendix for “Optimal Contracting with Endogenous Social Norms”

Paul Fischer and Steven Huddart

This appendix provides proofs omitted from the body of the paper.

Proof of Lemma 2: From (8), note that \( (\partial^2 z / \partial a_i^2) (\partial^2 z / \partial u_i^2) > (\partial^2 z / \partial a_i \partial u_i)^2 \) for all \( a_i \geq 0 \) and \( u_i \geq 0 \), which implies that the objective is concave. Analogous to the one-action case, therefore, the first-order conditions with respect to \( a_i \) and \( u_i \) of (8),

\[
\begin{align*}
\text{(A1)} & \quad b_i h'(a_i + u_i) - f'(a_i - (1 - \alpha_i)A_i - \alpha_i S_a) \leq 0 \quad \text{and} \\
\text{(A2)} & \quad b_i h'(a_i + u_i) - f'(u_i + (1 - \mu_i)U_i + \mu_i S_u) \leq 0
\end{align*}
\]

characterize agent \( i \)'s action choices, where condition (A1) is an equality if \( a_i > 0 \) and condition (A2) is an equality if \( u_i > 0 \).

The optimal choices of \( a_i \) and \( u_i \) implied by (A1) and (A2) must be finite because \( h'(x) \) approaches 0 as \( x \) approaches infinity while \( f \) increases at an increasing rate. Furthermore, \( a_i \) and \( u_i \) cannot both be zero because \( b_i > 0 \), which implies that agent \( i \) supplies a positive amount of at least one action. Otherwise, inequalities (A1) and (A2) are violated because \( h'(x) \) approaches infinity as \( x \) approaches 0.

Let \( a_i(S_a, S_u) \) and \( u_i(S_a, S_u) \) denote the optimal action choices for agent \( i \) as functions of the social norm parameters. Suppose \( a_i \) and \( u_i \) are interior for \( b_i > 0 \). Inspection of (A1) and (A2) implies that \( a_i - (1 - \alpha_i)A_i - \alpha_i S_a = u_i + (1 - \mu_i)U_i + \mu_i S_u \) at any interior optimum. By applying the implicit function rule to the first-order conditions (A1) and (A2), (i) \( a_i \) and \( u_i \) can be written as implicit functions of \( S_a \) and \( S_u \) and (ii) the derivatives of \( a_i \) and \( u_i \) with respect to \( S_a \) and \( S_u \) can be computed.

\[
\frac{\partial a_i(S_a, S_u)}{\partial S_a} = \frac{\alpha_i f''[-b_i h'' + f'']}{f''[-b_i h'' + f''] - b_i h'' f''} = \frac{\alpha_i G(i)}{1 + G(i)} \in [0, \alpha_i),
\]

(3)
where $G(i) \equiv 1 - f''(a_i - (1 - \alpha_i)A_i - \alpha_i S_a)/b_i h''(a_i + u_i) \geq 1$ because $f'' > 0$, $h'' < 0$, and $b_i > 0$ by assumption. In like fashion,

\begin{align}
(A4) \quad \frac{\partial a_i(S_a, S_u)}{\partial S_u} &= \frac{\mu_i}{1 + G(i)} \in [0, \mu_i/2), \\
(A5) \quad \frac{\partial u_i(S_a, S_u)}{\partial S_u} &= -\frac{\mu_i G(i)}{1 + G(i)} \in (-\mu_i, 0], \quad \text{and} \\
(A6) \quad \frac{\partial u_i(S_a, S_u)}{\partial S_a} &= -\frac{\alpha_i}{1 + G(i)} \in (-\alpha_i/2, 0].
\end{align}

These results establish that, when $a_i$ and $u_i$ are interior, $a_i$ is increasing in $S_a$ and $S_u$ while $u_i$ is decreasing in $S_a$ and $S_u$.

We next assess how $a_i$ changes in $S_a$ and $S_u$ on the boundaries, i.e., when $a_i = 0$ or $u_i = 0$. There are four cases to consider:

**Case 1:** Suppose $a_i > 0$, $u_i = 0$, and the inequality in (A2) is strict. In this case, a change in $S_a$, although it affects $a_i$, does not affect $u_i$ because $u_i = 0$ so

\begin{equation}
(A7) \quad \frac{\partial a_i(S_a, S_u)}{\partial S_a} = \frac{\alpha_i f''(a_i - N_{a_i})}{-b_i h''(a_i) + f''(a_i - N_{a_i})} \in (0, \alpha_i).
\end{equation}

**Case 2:** Suppose $a_i > 0$, $u_i = 0$, and (A2) is an equality. In this case, an increase in $S_a$, although it affects $a_i$, does not affect $u_i$ because inequality (A2) is strict at $u_i = 0$. A decrease in $S_a$ causes an increase in $u_i$ above zero because (A2) holds with equality at $u_i = 0$. Hence, the right-hand derivative is given by (A7) and the left-hand derivative is given by (A3).

**Case 3:** Suppose $a_i = 0$, $u_i > 0$, and the inequality in (A1) is strict. In this case, a change in $S_a$, does not affect $a_i$ because inequality (A1) is strict at $a_i = 0$.

**Case 4:** Suppose $a_i = 0$, $u_i > 0$, and (A1) is an equality. In this case, an increase in $S_a$ causes an increase in $a_i$ because (A1) is an equality. A decrease in $S_a$ does not affect $a_i$ because $a_i$ is constrained to be non-negative.

Combined, these cases imply that for $S_a$ above a threshold value, $a_i$ is increasing in $S_a$ and for $S_a$ below that threshold, $a_i = 0$.  

2
Now consider how $a_i$ changes in $S_u$ on the boundaries. Similar reasoning to that above implies that if $a_i(S_a, S_u) = 0$ for some $S_u$, then $a_i(S_a, X) = 0$ for all $X \leq S_u$. If $u_i = 0$ for some $S_u$, then $a_i(S_a, S_u) > 0$ and $a_i(S_a, X) = a_i(S_a, S_u)$ for all $X > S_u$. Thus, for $S_u$ below a threshold value, $a_i = 0$. Above that threshold value, (A4) implies $a_i$ in increasing in $S_u$, until a second threshold is reached at which $u_i = 0$. Above this threshold, $a_i$ does not change in $S_u$.

We next assess how $u_i$ changes in $S_a$ and $S_u$ on the boundaries. Parallel arguments to those above establish that $u_i(S_a, S_u) = 0$ implies $u_i(S_a, X) = 0$ for all $X \geq S_u$ and $u_i(S_a, S_u) > 0$ implies $u_i(S_a, S_u)$ is strictly decreasing in $S_u$. Thus, for $S_u$ below a threshold value, $u_i$ is interior and decreasing in $S_u$ and for $S_u$ above that threshold, $u_i = 0$. Furthermore, for $S_a$ above a threshold, $u_i$ is zero. Below that threshold value, (A6) implies $u_i$ is decreasing in $S_a$, until a second threshold is reached at which $a_i = 0$. Below this threshold, $u_i$ does not change in $S_a$.

In summary, $a_i$ is weakly increasing in $S_a$ and $S_u$ while $u_i$ is weakly decreasing in $S_a$ and $S_u$ for all $a_i \geq 0$ and $u_i \geq 0$. Because $a_i(S_a, S_u)$ and $a_i(S_a, S_u)$ (although monotone and continuous) are kinked, the derivatives below are right-hand derivatives.

For there to exist a unique equilibrium, if must be the case that there is a unique $S_a$ and $S_u$ that satisfy (3) and (10). Rewrite (10) as the equation

$$
(A8) \quad S_u + \frac{\int_I u_i(S_a, S_u)di}{\int_I di} = 0
$$

and use this to characterize $S_u$ as an implicit function of $S_a$. The definition of $S_u$ and the restrictions on the domain of the $u_i$ imply that $S_u \in (-\infty, 0]$. Because $u_i \geq 0$ for all $i \in I$, it follows that the left-hand side of equation (A8) is weakly positive at $S_u = 0$. Furthermore, the left-hand side of (A8) (i) decreases continuously and monotonically as
$S_u$ decreases and (ii) approaches $-\infty$ as $S_u$ approaches $-\infty$ because (A6) implies that

$$
\frac{\partial}{\partial S_u} \left( S_u + \int_{\mathcal{I}} u_i(S_a, S_u) \, di \right) \bigg|_{S_u}
= 1 + \frac{\int_{\mathcal{I}} \frac{\partial u_i(S_a, S_u)}{\partial S_u} \, di}{\int_{\mathcal{I}} \, di}
> 1 - \bar{\mu}
> 0.
$$

It follows from these observations that for any $S_a$ there is a unique value of $S_u$ satisfying (A8).

Let $S_u(S_a)$ denote the unique value of $S_u$ that satisfies (A8). Applying the implicit function rule to (A8), we have

$$
\frac{\partial S_u(S_a)}{\partial S_a} = -\frac{\int_{\mathcal{I}} \frac{\partial u_i(S_a, S_u)}{\partial S_u} \, di}{\int_{\mathcal{I}} \, di(1 + \frac{\partial u_i(S_a, S_u)}{\partial S_u} \, di)}.
$$

Now rewrite (3) as the equation

(A9) \quad $F \equiv S_a - \frac{\int_{\mathcal{I}} a_i(S_a, S_u(S_a)) \, di}{\int_{\mathcal{I}} \, di} = 0$

and totally differentiate with respect to $S_a$ to obtain

$$
\frac{dF}{dS_a} = \int_{\mathcal{I}} \left( 1 - \frac{\partial a_i}{\partial S_a} - \frac{\partial a_i}{\partial S_u} \frac{\partial S_u}{\partial S_a} \right) dS_a \bigg/ \int_{\mathcal{I}} \, di
\propto \int_{\mathcal{I}_1} \left( 1 - \frac{\partial a_i}{\partial S_a} - \frac{\partial a_i}{\partial S_u} \frac{\partial S_u}{\partial S_a} \right) dS_a \bigg/ \int_{\mathcal{I}_1} \, di
+ \int_{\mathcal{I}_a} \left( 1 - \frac{\partial a_i}{\partial S_a} - \frac{\partial a_i}{\partial S_u} \frac{\partial S_u}{\partial S_a} \right) dS_a \bigg/ \int_{\mathcal{I}_a} \, di
+ \int_{\mathcal{I}_u} \left( 1 - \frac{\partial a_i}{\partial S_a} - \frac{\partial a_i}{\partial S_u} \frac{\partial S_u}{\partial S_a} \right) dS_a \bigg/ \int_{\mathcal{I}_u} \, di
= \int_{\mathcal{I}_1} \left( 1 - \frac{\partial a_i}{\partial S_a} - \frac{\partial a_i}{\partial S_u} \frac{\partial S_u}{\partial S_a} \right) dS_a + \int_{\mathcal{I}_a} \left( 1 - \frac{\partial a_i}{\partial S_a} + \int_{\mathcal{I}_u} \, di\right)
$$

where $\mathcal{I}_i$, $\mathcal{I}_a$, and $\mathcal{I}_u$ are a partition of $\mathcal{I}$ such that $j \in \mathcal{I}_i$ iff (A1) and (A2) are equalities, $j \in \mathcal{I}_a$ iff (A1) is an equality and (A2) is a strict inequality, and $j \in \mathcal{I}_u$ iff (A2) is an equality but (A1) is a strict inequality. Provided the integration is over a set of positive
measure, we show that each of these integrals is positive. Obviously, \( \int_{I_a} 1 - \frac{\partial a_i}{\partial S_a} \, di \geq (1 - \bar{\alpha}) \int_{I_a} di > 0. \) From (A7),

\[
\int_{I_a} 1 - \frac{\partial a_i}{\partial S_a} \, di \geq (1 - \bar{\alpha}) \int_{I_a} di \]

Finally,

\[
\int_{I_a} 1 - \frac{\partial a_i}{\partial S_a} \, di = 1 - \int_{I_a} \frac{\partial a_i}{\partial S_a} \, di + \int_{I_a} \frac{\partial a_i}{\partial S_u} \, di - \int_{I_a} \frac{\partial a_i}{\partial S_u} \, di \int_{I_a} 1 + \frac{\partial u_i}{\partial S_u} di \int_{I_a} \frac{\partial a_i}{\partial S_a} \, di \\
\int_{I_a} 1 - \alpha_i G(i) \, di \left(1 - \int_{I_i} 1 + G(i) \, di\right) - \int_{I_i} \mu_i G(i) \, di - \int_{I_i} \frac{\alpha_i}{1 + G(i)} \, di \int_{I_i} \frac{\mu_i}{1 + G(i)} \, di \\
> (1 - \bar{\alpha} \int_{I_i} 1 + G(i) \, di) \left(1 - \bar{\mu} \int_{I_i} \frac{G(i)}{1 + G(i)} \, di\right) - \bar{\alpha} \bar{\mu} \left(\int_{I_i} 1 + G(i) \, di\right)^2 \\
= \left(\int_{I_i} 1 + G(i) - \bar{\alpha} G(i) \, di\right) \left(\int_{I_i} 1 + G(i) - \bar{\mu} G(i) \, di\right) - \bar{\alpha} \bar{\mu} \left(\int_{I_i} 1 + G(i) \, di\right)^2 \\
= \left(\int_{I_i} \frac{1}{1 + G(i)} \, di + \int_{I_i} \frac{(1 - \bar{\alpha}) G(i)}{1 + G(i)} \, di\right) \left(\int_{I_i} \frac{1}{1 + G(i)} \, di + \int_{I_i} \frac{(1 - \bar{\mu}) G(i)}{1 + G(i)} \, di\right) \\
- \bar{\alpha} \bar{\mu} \left(\int_{I_i} 1 + G(i) \, di\right)^2 \geq (1 - \bar{\alpha} \bar{\mu}) \left(\int_{I_i} 1 + G(i) \, di\right)^2 > 0,
\]

Hence, \( dF/dS_a > 0 \) and so

\[
S_a = \frac{\int_{I_a} a_i(S_a, S_u(S_a)) \, di}{\int_{I_a} di} = 0
\]

must have a unique solution value for \( S_a \) because the left-hand side is (i) non-positive at \( S_a = 0 \), (ii) strictly increasing in \( S_a \), and (iii) approaches infinity as \( S_a \) approaches infinity.
Proof of Corollary 2: For interior $a$ and $u$, (1) and (2) are equalities. The definitions (3) and (10) imply $P \equiv -a + S_a = 0$ and $Q \equiv u + S_u = 0$. These four equations (i) define the equilibrium values of the endogenous variables $a$, $u$, $S_a$, and $S_u$ in terms of the exogenous variables $\alpha$, $\mu$, $A$, $U$, and $b$ and (ii) have continuous partial derivatives with respect to each of the endogenous and exogenous variables. The corresponding Jacobian determinant is

$$|J| = \left| \frac{\partial(z_a, z_u, P, Q)}{\partial(a, u, S_a, S_u)} \right|$$

$$= \begin{vmatrix}
  z_{aa} & z_{au} & \frac{\partial z_a}{\partial S_a} & \frac{\partial z_a}{\partial S_u} \\
  z_{ua} & z_{uu} & \frac{\partial P}{\partial S_a} & \frac{\partial P}{\partial S_u} \\
  \frac{\partial P}{\partial a} & \frac{\partial P}{\partial u} & \frac{\partial P}{\partial S_a} & \frac{\partial P}{\partial S_u} \\
  \frac{\partial Q}{\partial a} & \frac{\partial Q}{\partial u} & \frac{\partial Q}{\partial S_a} & \frac{\partial Q}{\partial S_u} \\
\end{vmatrix}$$

$$= \begin{vmatrix}
  bh'' - f'' & bh'' & \alpha f'' & 0 \\
  bh'' & bh'' - f'' & 0 & -\mu f'' \\
  -1 & 0 & 1 & 0 \\
  0 & 1 & 0 & 1 \\
\end{vmatrix}$$

$$= [(1 - \alpha)(1 - \mu) f'' - (2 - \alpha - \mu) bh''] f''$$

$$> 0,$$

where subscripts on $z$ denote partial differentiation with respect to the subscripted variable, $f''$ is evaluated at either $a - N_a$ or $u + N_u$ (and, in equilibrium, $a - N_a = u + N_u$), and $h''$ is evaluated at $a + u$. Therefore, by the implicit function rule, for any change in an exogenous variable, $y$,

(A10) \[ z_{aa} \frac{da}{dy} + z_{au} \frac{du}{dy} + \frac{\partial z_a}{\partial S_a} \frac{dS_a}{dy} + \frac{\partial z_a}{\partial S_u} \frac{dS_u}{dy} = -\frac{\partial z_a}{\partial y}, \]

(A11) \[ z_{au} \frac{da}{dy} + z_{uu} \frac{du}{dy} + \frac{\partial z_u}{\partial S_a} \frac{dS_a}{dy} + \frac{\partial z_u}{\partial S_u} \frac{dS_u}{dy} = -\frac{\partial z_u}{\partial y}, \]

\[ -\frac{da}{dy} + \frac{dS_a}{dy} = \frac{\partial P}{\partial y}, \quad \text{and} \]

\[ \frac{dS_u}{dy} + \frac{du}{dy} = \frac{\partial Q}{\partial y}, \]

where $y$ is any one of the exogenous variables. Observe that $\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial y}$ =
\( \partial z_u / \partial S_a = \partial z_u / \partial S_u = 0. \) Hence, (A10) and (A11) reduce to

\[
\left( z_{aa} + \frac{\partial z_a}{\partial S_a} \right) \frac{da}{dy} + z_{au} \frac{du}{dy} = -\frac{\partial z_a}{\partial y} \quad \text{and}
\]

\[
z_{au} \frac{da}{dy} + \left( z_{uu} - \frac{\partial z_u}{\partial S_u} \right) \frac{du}{dy} = -\frac{\partial z_u}{\partial y}.
\]

Solving this last system of equations for \( da/dy \) and \( du/dy \) yields

\[
\frac{da}{dy} = \left\{ [-b h'' + (1 - \mu) f''] \frac{\partial z_a}{\partial y} + b h'' \frac{\partial z_u}{\partial y} \right\} \left/ \left| J \right| \right. \quad \text{and}
\]

\[
\frac{du}{dy} = \left\{ [-b h'' + (1 - \alpha) f''] \frac{\partial z_u}{\partial y} + b h'' \frac{\partial z_a}{\partial y} \right\} \left/ \left| J \right| \right.
\]

The comparative-static derivatives in the statement of the corollary follow directly from these last two expressions upon observing that \( 0 < -b h'', \ 0 < (1 - \alpha) f'', \) and \( 0 < (1 - \mu) f'' \).

**Proof of Proposition 1:** Because \( N_{a} = (1 - \alpha)A + \alpha [\delta a_h + (1 - \delta) a_l] \) and \( a_i - N_{a_i} = u_i + N_{u_i} \) for any agent \( i \), it is possible to write the difference in costs between single and separate organizations as

\[
C_1 - C_2 = 2 \left[ \delta f ((1 - \alpha) (a_h - A) + (1 - \delta) x) \\
+ (1 - \delta) f ((1 - \alpha) (a_l - A) - \delta x) \\
- \delta f ((1 - \alpha) (a_h - A)) \\
- (1 - \delta) f ((1 - \alpha) (a_l - A)) \\
+ \frac{\delta(1 - \delta)(a_h - a_l)(\alpha - \mu)(k_h - k_l)}{1 - \mu}, \right]
\]

(A12)

where \( x = \alpha(a_h - a_l) \). We first show that the term in square brackets is always non-negative. Note that the equation (A12) equals 0 when \( x = 0 \) and that \( x \) must be non-negative. Furthermore, the first derivative (A12) with respect to \( x \) yields

\[
2\delta (1 - \delta) [f' ((1 - \alpha) (a_h - A) + (1 - \delta) x) - f' ((1 - \alpha) (a_l - A) - \delta x)] > 0
\]
because \( f'' > 0 \) and \( (1 - \alpha)(a_h - A) + (1 - \delta)x > (1 - \alpha)(a_l - A) - \delta x \).

Because the term in square brackets is non-negative, the costs associated with a single organization are greater than with separate organizations if and only if
\[
[\delta(1 - \delta)(a_h - a_l)(\alpha - \mu)(k_h - k_l)]/(1 - \mu) \text{ is not too small (negative), which occurs when }
(\alpha - \mu)(k_h - k_l) < 0.
\]

**Proof of Proposition 2:** Differentiating the cost (16) with respect to the parameters determining the sensitivity to the social norms yields
\[
\frac{dC(a)}{d\alpha} = -2(a - A)f'(1 - \alpha)(a - A)) - \frac{(a - A)}{1 - \mu}k \propto (A - a) \quad \text{and}
\frac{dC(a)}{d\mu} = k\frac{(a - A)(1 - \alpha)}{(1 - \mu)^2} \propto (a - A).
\]

**Proof of Observation 1:** Let \( \alpha_h \) denote the parameter \( \alpha \) for the agents who are most sensitive to the social norm for the desired action and \( \alpha_l < \alpha_h \) denote the parameter \( \alpha \) for the agents who are least sensitive to that norm. Consider first the case where the level of desired action \( a \) exceeds the common personal norm \( A \). Proposition 2 implies that the \( \alpha_h \) agents are preferred. Under the contract that induces \( a \) from all \( \alpha_h \) agents when type is observable, each of these agents attains the reservation level of utility \( v \) the and faces a norm for the desired action \( N_{a_h} = \alpha a + (1 - \alpha)A \). Assume that same contract is offered for all agents willing to join the organization when type is not observed. The proof for this case is completed by showing that, if all agents conjecture that only the \( \alpha_h \) types will accept the offer, only the \( \alpha_h \) types will weakly prefer to accept the offer and the \( \alpha_l \) types will strictly prefer to not accept the offer. An \( \alpha_h \) agent weakly prefers to accept the offer because he anticipates attaining \( v \). An \( \alpha_l \) type that accepts the offer faces a lower social norm for the desired action than an \( \alpha_h \) type because \( \alpha_l < \alpha_h \) and
\( a > A \) implies \( N_{a_l} < N_{a_h} \). As a consequence an \( \alpha_l \) type attains a level of expected utility less than \( v \) if he joins the organization. Hence, an \( \alpha_l \) type strictly prefers to not accept the offer. The proof for the case where \( a < A \) is analogous. Finally, the principal is indifferent between types when \( a = A \).
Proof of Observation 2: Let $\mu_h$ denote the parameter $\mu$ for the agents who are most sensitive to the social norm for the undesirable action and $\mu_l < \mu_h$ denote the parameter $\mu$ for the agents who are least sensitive to that norm. Consider first the case where the level of desired action $a$ exceeds the common personal norm $A$. Proposition 2 implies that the $\mu_l$ agents are preferred. The proof for this case is by contradiction. Conjecture a contract offer that induces each $\mu_l$ agent to accept the contract and take action $a$ and that induces each $\mu_h$ to not accept the offer. Let $v_l$ denote the level of expected utility attained by a $\mu_l$ type that accepts the contract, where $v_l$ must weakly exceed $v$. It must be the case that the social norm for the undesirable action faced by any individual agent accepting the contract must be

$$S_u = -u = U - \frac{(a - A)(1 - \alpha)}{1 - \mu_l} < U,$$

where $\alpha$ is the common sensitivity to the social norm for the desired action and $U$ is the common personal norm for the undesirable action. Because $S_u < U$, it follows that the norm for the undesirable action faced an agent of type $\mu_h$ who joins the organization must be strictly less than that for an agent of type $\mu_l$. Hence, the expected utility of an agent of type $\mu_h$ who joins the organization, $v_h$, must be strictly greater than that for an agent of type $\mu_l$, $v_l$. The contradiction follows from the observation that a type $\mu_h$ agent must strictly prefer to accept the contract and join the organization because $v_h > v_l \geq v$. The proof for the case where $a < A$ is analogous.