Optimal Contracting with
Endogenous Social Norms

By Steven Huddart and Paul Fischer*

Research in sociology and ethics suggests that individuals adhere to social norms of behavior established by their peers. Within an agency framework, we model an endogenous social norm by assuming each agent’s disutility for an action decreases when other agents in the same organization also take that action. We show how endogenous social norms: alter the effectiveness of monetary incentives, determine whether it is optimal to group agents in a single or two separate organizations, and give rise to a costly adverse selection problem when agents’ sensitivity to social norms is unobservable. (JEL C70 D63 D70 J33 Z13)

In most economic models of organization, individual behavior is guided by contractual, legal, and reputational considerations. With respect to many choices that individuals face, a desire to adhere to some standard of conduct or social norm also guides their behavior. In this paper, we explore the role of social norms as a determinant of organization design.

We study a principal-agent model in which a principal contracts with a continuum of agents within an organization. Each agent can favorably influence the performance measure upon which his compensation depends by taking an action that serves the principal’s interests, which we call the desirable action, as well as taking an action that harms the principal’s interests, which we call the undesirable action. Because both actions favorably affect the performance measure used for contracting, the principal

* Fischer: Smeal College of Business, Pennsylvania State University, Box 1912, University Park, PA 16802-1912 (e-mail: pef5@psu.edu); Huddart: Smeal College of Business, Pennsylvania State University, Box 1912, University Park, PA 16802-1912 (e-mail: huddart@psu.edu). We thank Douglas Bernheim, Kalyan Chatterjee, Michel Clement, Mark Dirsmith, Don Hambrick, Michael Kirschenheiter, Richard Sansing, Steven Schwartz, Shiva Sivaramakrishnan, Linda Treviño, Amir Ziv, two anonymous reviewers, and seminar participants at the 2004 meetings of the American Accounting Association, Carnegie Mellon University, Columbia University, the University of Houston, the Ohio State University, and SUNY–Binghamton for helpful discussions.
must incur the cost of the undesirable action to motivate an agent to take the desirable action.

We incorporate a social norm into the model in a manner consistent with findings in social psychology suggesting that individuals conform to norms of behavior established by their peers’ actions. Specifically, the psychic cost of an action to an agent is determined in part by an endogenous social norm, which is a function of the actions of all agents in the organization. The modeling of the norm and psychic cost imply that an agent incurs a lower cost for an action when other agents in the organization undertake more of that action. In our primary analysis, we suppose the social norm affects the psychic cost (e.g., feelings of guilt) incurred by an agent for taking the undesirable action. Later, we consider a norm that affects the psychic cost (e.g., cognitive fatigue) of the desirable action.

Besides introducing an endogenous social norm, our agency model differs from the standard multi-task framework of Bengt Holmström and Paul Milgrom (1991) because the undesirable action is detrimental to the principal even though it favorably influences the performance measure. The undesirable action is included in the model because individuals commonly take actions to improve a performance measure, say earnings, that give rise to economic costs or foregone benefits not reflected in the performance measure. Two classes of such undesirable actions are accounting manipulation and real manipulation.

Accounting manipulation opportunities arise in part because accounting requires estimates, which determine various components of earnings. For example, estimates of the value of consideration received in non-cash transactions affect reported revenue; estimates of uncollectible receivables affect bad debt expense, estimates of ending inventory values affect cost of goods sold; and estimates of future compensation rates, interest rates, life expectancy, turnover, and rates of return on pension plan assets affect defined benefit pension plan and health care expense. Reported earnings, which is revenues net of expenses, therefore depends on each of these estimates. Since the estimates are inherently unverifiable, managers have latitude to manipulate the estimates and distort
earnings to further their interests. In addition, accounting rules often provide managers with discretion regarding how to account for a class of transactions. For example, managers have discretion to choose revenue recognition, inventory accounting and depreciation policies; they may use that discretion to distort earnings.\(^3\)

Real manipulation involves altering operating activities to improve a reported performance measure, even though such alterations result in economic costs that are not reflected in the measure. For example, managers may forego expenditures on research and development, advertising, and maintenance because financial accounting rules dictate that such expenditures are expensed immediately, while the benefits are recognized in future earnings. As another example, managers may engage in channel stuffing (i.e., actions to encourage or coerce intermediary buyers to make purchases that would otherwise be deferred until later dates) near the end of a period to boost revenues for the current period. Channel stuffing is costly because it is achieved by offering excessive discounts, cannibalizing future periods’ sales, and incurring the costs of future returns. As a final example, managers may set up off-balance sheet vehicles to create immediate revenue flows or defer expense flows. Such vehicles, however, are costly to create and may expose shareholders to uncompensated risks.\(^4\)

We first discuss how a social norm alters the impact of monetary incentives on action choices. The assumption that each agent’s action cost is influenced by other agents’ action choices via a social norm implies that increasing the power of monetary incentives across an organization has a direct effect on agents’ actions, as well as an indirect effect through the endogenous social norm. When the norm affects the cost of the undesirable action, the indirect effect enhances the effect of increased monetary incentives on undesirable actions and mitigates the effect on desirable actions. In contrast, when the norm pertains to the desirable action, the indirect effect enhances the effect of increased monetary incentives on desirable actions and mitigates the effect on undesirable actions.

Next, we explore how endogenous social norms can influence the optimal design of an organization. Because the choices of one group of employees affect the social norm for the entire peer group, monetary incentives offered to some agents within a peer
group have spill-over affects on the action choices of others in that peer group. If the boundaries of an organization define the peer group to which a social norm applies, the spill-over effects from incentive contracts can determine the optimal boundaries of the organization. Our analysis provides conditions under which it is better to split up a diverse organization employing high-powered incentives for one group of agents and low-powered incentives for another.

Finally, because individuals differ with respect to how sensitive their behavior is to social norms, we analyze issues related to the matching of agent types with tasks. In particular, we consider a setting in which some agents are more sensitive to a social norm than others, and assess the relation between the level of desirable action implemented by the principal and the types of agents preferred by the principal. We also assess whether the principal will naturally attract the preferred agents when agent type is not observable.

Other papers have considered how non-monetary incentives, such as social norms, have consequences for the returns to financial incentives. In a setting with a single agent, Roland Bénabou and Jean Tirole (2003) formalize the notion that monetary incentives can crowd out incentives provided by intrinsic factors. Relatedly, experimental and field study work in a non-business setting by Uri Gneezy and Aldo Rustichini (2000a, 2000b), supports the notion that intrinsic factors can undermine the effectiveness of financial incentives. In particular, they find that introducing financial incentives in a non-business setting can lead to less of a desirable action (correct answers to test questions) and more of an undesirable action (the tardy collection of children from a day care center), respectively. Their evidence of important interactions between extrinsic (i.e., monetary) and intrinsic incentives (e.g., a social norm or peer pressure) suggests that a norm mechanism, in addition to a standard compensation mechanism, merits study.

Some recent working papers also consider the role of norms in an agency context. In a study of team production incorporating a norm, Steffen Huck, Dorothea Kübler, and Jörgen Weibull (2003) show that strong monetary incentives tied to aggregate team
production can rule out Pareto-preferred equilibria that are attainable if incentives are weak and norms are strong. Dirk Sliwka (2003) shows that firms may forego high-powered incentive contracts because they attract agents who respond solely to the incentives which, in turn, undermines the behavior of agents employed by that firm who conform to the behavior of others. These papers do not consider how norms determine organizational boundaries or how agents are matched to organizations.

Finally, some tax and welfare policy research considers the role of norms. The tax compliance studies of Joel Slemrod (2004), and Jon Davis, Gary Hecht, and Jon Perkins (2003) suggest that social norms are an important determinant of tax compliance behavior. In a related vein, N. Soren Blomquist (1993) and Assar Lindbeck, Sten Nyberg, and Jörgen Weibull (1996) consider the effects of tax and welfare policy, respectively, on labor supply when norms influence behavior.6

The paper proceeds as follows. Section I presents the primary model in which norms influence the cost of undesirable actions. Section II considers how the design of organizations is influenced by the social norm. Section III analyzes how implementation costs vary with behavioral traits, and discusses whether firms can offer contracts that attract the agents with the desired traits. Section IV revisits the results derived from our primary model in a setting in which norms influence the cost of the desirable, as opposed to undesirable, actions. Section V concludes.

I. Norms for Undesirable Actions

A risk-neutral principal employs a continuum of risk-neutral agents in one or more organizations. The agents are indexed on [0, 1] and an organization of agents is defined to be a set $J \subseteq [0, 1]$, where $J$ has strictly positive measure, such that an agent with index $i$ is in organization $J$ if $i \in J$. An agent is employed in only one organization.

Each agent $i$ makes two unobservable action choices that affect the principal’s welfare, $a_i \in [a, \infty)$ and $u_i \in [u, \infty)$, where $a$ and $u$ may be negative, zero, or positive. We restrict the analysis to the design of contracts that induce each agent $i$ to take an action $a_i$, which we call the desirable action. Agent $i$’s other action choice $u_i$, which we
call the undesirable action, imposes a cost on the principal of \( k_i(u_i - u) \) where \( k_i > 0 \).

Agent \( i \)'s compensation is a function of a (possibly stochastic) report, \( r_i \), where the mean of \( r_i \) is \( h(a_i + u_i) \), \( h' > 0 \), and \( h'' \leq 0 \). We further restrict our analysis to linear contracts of the form \( w_i + b_i r_i \), with \( w_i \in \mathbb{R} \) and \( b_i \in \mathbb{R}^+ \) because these contracts make intuitive the interplay between the intensity of incentives, represented by the \( b_i \)'s, and social norms. Furthermore, when the density function for \( r_i \) as a function of \( a_i \) and \( u_i \) can be written in the form \( g(r_i, a_i + u_i) \), restricting attention to linear contracts is without loss of generality because linear contracts are as efficient as any other contract that induces any interior set of desirable actions by the agents of an organization.

A compensation scheme is represented by \( \{ W, B \} \), where \( W = \{ w_j \}_{j \in [0,1]} \) and \( B = \{ b_j \}_{j \in [0,1]} \). Given contract \( \{ w_i, b_i \} \), agent \( i \) chooses actions \( \{ a_i, u_i \} \) to maximize

\[
(1) \quad w_i + b_i h(a_i + u_i) - f(a_i) - f(u_i + N_i),
\]

where \( f \) is a twice-differentiable cost function satisfying: \( f(x) = 0 \) and \( f'(x) = 0 \) for \( x \leq a \); \( f'(x) > 0 \) and \( f''(x) > 0 \) for \( x > a \); and \( f'(x) \to \infty \) as \( x \to \infty \). The cost functions for \( a_i \) and \( u_i \) differ in that a norm parameter, \( N_i \), affects the total and marginal costs of the undesirable action. The agents have a common reservation level of expected utility, \( v \).

Introducing a norm \( N_i \) into the cost function for the undesirable action is consistent with the idea that, in addition to any physical or cognitive cost, undesirable actions have an associated psychic cost (e.g., a guilty conscience), which is determined by an individual’s norm. Consistent with the social psychology literature, we assume that individual \( i \)'s norm is a function of his personal norm and a common social norm. The norm for agent \( i \) is a weighted average of agent \( i \)'s personal norm \( P_i \), which is an element of the finitely bounded set \([P, \overline{P}]\), and a common social norm \( S \), which is a function of the per-capita level of the undesirable action by agents in \( i \)'s organization. Formally, the norm for agent \( i \) in an organization \( J \), is

\[
(2) \quad N_i \equiv (1 - \lambda_i) P_i + \lambda_i S, \quad \text{where}
\]

\[
(3) \quad S \equiv -\frac{\int_J u_j \, dj}{\int_J \, dj}.
\]
Parameter $\lambda_i$ represents the extent to which agent $i$ is influenced by the behavior of others in his organization. We assume that each $\lambda_i$ lies in $(0, \bar{\lambda})$, with $\bar{\lambda} \in (0,1)$.

The definition of $S$ implies that marginal cost of $u_i$ to agent $i$ decreases in the average undesirable action taken by others in his organization, which formalizes the idea that individuals are more willing to engage in undesirable actions when other members of their peer group undertake such actions.\footnote{Note that we leave unspecified the payoff that the principal derives from a given supply of the desired action. We do this to focus on the minimum cost of implementing the desired action, which includes the cost of any undesirable action undertaken by the agents. We emphasize that the principal uses the report as a source of information (i.e., the report has no intrinsic value), and that he cares about the desirable action because it generates some valuable noncontractible payoff. In addition, analogous to models of atomistic agents where each agent is a price taker, agents are “norm takers.” In particular, changing $u_i$ for any one agent has no effect on the social norm, $S$. Finally, analogous to models with rational expectations or consistent beliefs requirements, we assume agents correctly anticipate the actions of others in the organization so that actual action choices define the norm.}

Finally, assume the lower bound for the undesirable action, $u$, satisfies a technical condition to ensure that the agents’ action choices for $u$ are interior: $$u \leq \min \left( \frac{a}{1 - \lambda}, a \right) - \max \left( \mathcal{P}, 0 \right).$$

A post-contracting equilibrium for an organization of agents $J$ who have contracts $\{w_j, b_j\}_{j \in J}$ is defined as a set of actions, $\{a_j, u_j\}_{j \in J}$ such that each agent’s action choice maximizes objective (1) given $\{w_j, b_j\}_{j \in J}$ and $S$, and $S$ satisfies equation (3) given $\{u_j\}_{j \in J}$. The following lemma establishes the existence and uniqueness of a post-contracting equilibrium, as well as two characteristics of the equilibrium used later in the paper.\footnote{The proof is in the appendix.}
LEMMA 1: For any set of contracts, \( \{w_j, b_j\}_{j \in J} \) for organization \( J \), there exists a unique post contracting equilibrium, \( \{a_j, u_j\}_{j \in J} \). In equilibrium, \( a_i = u_i + N_i \) for all \( i \), and

\[
S = -\frac{\int_J a_j \, dj - \int_J (1 - \lambda_j) P_j \, dj}{\int_J \, dj - \int_J \lambda_j \, dj}.
\]

Because \( a_i = u_i + N_i \) the mix of actions selected by agent \( i \) shifts towards the desired action as his norm, \( N_i \), increases. The shift occurs because a higher norm makes the undesired action relatively more costly to the agent. Equation (4) implies that the social norm is lower when the desired actions induced are higher, on average. The relation between the social norm and the average desired action arises because a higher desired action requires higher-powered incentives, higher-powered incentives induce more of the undesirable action, and more undesirable action implies a lower social norm.

To gain some insight into how the social norm affects agent responses to changes in incentive intensity, we consider a setting in which all agents have identical preferences and contracts. We suppress subscripts because the agents and contracts are identical. In this setting, an increase in the power of incentives across all the agents in the organization causes the post-contracting undesired and desired actions for each agent to change as follows:

\[
\frac{du}{db} = \frac{h'(2u + N) + \frac{dS}{db} \lambda bh''(2u + N) - f''(u + N)}{2bh''(2u + N) - f''(u + n)}
\]

and

\[
\frac{da}{db} = \frac{h'(2a - N) - \frac{dS}{db} \lambda bh''(2a - N)}{2bh''(2a - N) - f''(a)}.
\]

Ignoring the effect of the social norm (i.e., the term in each numerator involving \( dS/db \)), the fact that \( h' > 0, h'' \leq 0 \) and \( f'' > 0 \) implies that each action choice increases in response to higher returns to those actions. The impact of incentives on the social norm, however, has opposing incremental effects for each action choice. Noting that \( S = -u \) in this setting, the change in the social norm satisfies

\[
\frac{dS}{db} = \frac{h'(2u + N)}{(2 - \lambda)bh''(2u + N) - (1 - \lambda)f''(u + n)} < 0.
\]
Hence, the impact of increased incentives reduces the social norm. The characterization of $du/db$ and $dS/db$, in turn, imply that the social norm magnifies the impact of increased incentives on undesirable actions. In contrast, the characterization of $da/db$ and $dS/db$ imply that the social norm mutes the impact of increased incentives on desirable actions if $h'' < 0$. Given our model specification, however, the muting effect on the desired action never dominates, so the desired action always increases in the power of incentives.

**B. Per-Capita Implementation Costs**

Having characterized the post-contracting equilibrium, we turn to the determination of the minimum-cost contracts for implementing a specified set of desirable actions, $A \equiv \{a_j\}_{j \in [0,1]}$. For a given allocation of agents to organizations, the optimal contracting scheme for implementing $A$ is a $\{W, B\}$ that minimizes the sum of the per-capita expected compensation and undesirable action costs subject to each agent attaining his reservation level of expected utility, $v$, and the action choices satisfying the equilibrium conditions for each $i$. That is, the optimal $\{W, B\}$ minimizes

$$\int_0^1 [w_j + b_j h (a_j + u_j) + k_j (u_j - u)] dj$$

subject to a reservation constraint for each $i$,

$$w_i + b_i h(a_i + u_i) - f(a_i) - f(u + N_i) \geq v, \quad (5)$$

and the equilibrium conditions for each $a_i, u_i, N_i$, and the social norm for each organization given in (1), (2) and (3), respectively.

The reservation constraint (5) is always binding because the constant in the compensation contract can be altered without affecting the agents’ action choices. As a consequence, the principal’s objective can be written solely as a function the desirable $A$ and the induced set of undesirable actions, $U$:

$$\int_0^1 [v + f(a_j) + f(u_j + N_j) + k_j (u_j - u)] dj. \quad (6)$$
The first-best optimal \( u_i \) is \( u \) for all \( j \) given any desirable \( A \). Thus, the cost of implementing \( A \) in the first-best case is simply \( \int_0^1 [v + f(a_j)] \, dj \). Comparison of the first-best minimum cost of implementing \( A \) and the second-best cost of implementing \( A \) identifies two sources of losses associated from moral hazard. First, there is the cost associated with the undesirable action imposed on the principal, \( \int_0^1 k_j(u_j - u) \, dj \). Second, there is the indirect cost of compensating the agents for the psychic cost, \( \int_0^1 f(u_j + N_j) \, dj \), that follows from putting them in a situation that leads them to take the undesirable action.

II. Organizational Boundaries

Organizations often include individuals who specialize in different desirable tasks (e.g., auditing and consulting, or sales and service), and each task may require a different incentive system. In the simplest case, which we develop below, there are two activities. The first activity requires a low level of the desirable action, denoted \( a_l \) (and hence a low incentive, denoted \( b_l \)), while the second requires a high level of the desirable action, denoted \( a_h \) (and hence a high incentive, denoted \( b_h > b_l \)). If agents’ behaviors are sensitive to the social norm for their organization, then giving one group of agents in an organization high-powered incentives to induce a high level of desirable action from them can lead to a costly erosion of norms for those in the organization who have low-powered incentives. Conversely, giving one group low-powered incentives can enhance the social norm for the organization, thereby restraining the undesirable actions of those with high-powered incentives. Because the social norm for each agent in an organization is determined solely by others in that organization, it follows that it may be optimal for an organization to split into two organizations to preserve the social norm for agents with low-powered incentives in some situations, and it may be optimal for an organization to stay intact to preserve the social norm for the agents with high-powered incentives in other situations.

To analyze this issue in isolation, consider a setting in which all agents have the same behavioral traits, \( P_i = P \) and \( \lambda_i = \lambda \) for \( i \in [0, 1] \). Some proportion, \( \delta \), of the agents are employed to exploit a high marginal production opportunity and
the remaining proportion $1 - \delta$ are employed to exploit a low marginal production opportunity. We might think of the former opportunity as a new line of business with growth prospects and the latter as a mature line of business. Throughout our analysis, we assume the desirable action in the high opportunity case, $a_h$, exceeds that in the low opportunity case, $a_l$. Furthermore, denote the per-unit cost to the organization of the undesirable action in the high opportunity case $k_h$ and that in the low opportunity case $k_l$. Call the agents assigned to the high marginal production opportunity the high-production agents and the remainder the low-production agents.

We consider two cases. In the first case, all agents work in the same organization, so the social norm is computed over all agents. In the second case, there are two organizations. One organization focuses solely in the low marginal production opportunity and employs proportion $1 - \delta$ of the agents. The other organization focuses on the high marginal production opportunity and employs the remaining agents. We compare the costs of implementing the desirable actions across the two settings to assess the implications of the social norm on organizational design.

Given the specific additional assumptions in this section, the cost of implementing $\{a_h, a_l\}$ in the single organization is

$$C_1 = v + \delta [f(a_h) + f(u_{hs} + N_s) + k_h(u_{hs} - u)]$$
$$+ (1 - \delta) [f(a_l) + f(u_{ls} + N_s) + k_l(u_{ls} - u)],$$

(7)

where $u_{hs}$ and $u_{ls}$ are the levels of undesirable actions associated with $a_h$ and $a_l$, respectively, and $N_s$ is the norm in the single organization. The cost of implementing $\{a_h, a_l\}$ in two separate organizations is

$$C_2 = v + \delta [f(a_h) + f(u_h + N_h) + k_h(u_h - u)]$$
$$+ (1 - \delta) [f(a_l) + f(u_l + N_l) + k_l(u_l - u)],$$

(8)

where $u_h$ and $u_l$ are the levels of undesirable actions associated with $a_h$ and $a_l$, respectively, and $N_h$ and $N_l$ are the norms in the organizations implementing $a_h$ and $a_l$, respectively.
The post-contracting equilibrium characteristics in Lemma 1 allow us to write the difference in costs as

\[ C_1 - C_2 = \frac{\lambda \delta (1 - \delta) (a_h - a_l) (k_l - k_h)}{1 - \lambda}. \]

Proposition 1 follows by inspection of equation (9).

**PROPOSITION 1:** The costs associated with separate organizations are less than the costs associated with a single organization if and only if \( k_l - k_h > 0 \).

Proposition 1 offers a perspective on public accounting firms that provide both consulting and auditing services. In such a professional services firm, the revenues secured by a partner are a primary performance measure for that partner. A consequence of employing revenue as a performance measure, however, is that each partner has incentives to ignore his own professional judgment and appease client management to secure revenues from that client. Appeasement, however, creates a negative externality from the firms’ perspective because it can impose subsequent reputation and legal costs on the firm as a whole. Such behavior is discouraged, in part, by norms of professional behavior. Within the context of our model, appeasement, which is a form of real manipulation, can be thought of as the undesirable action, whereas the desirable action is working hard to cultivate and service clients. Furthermore, like the undesirable action in our model, appeasement is controlled to some extent by a social norm.

Prominent commentators allege that there is a costly erosion of norms for auditors when consulting and auditing services both are provided by a public accounting firm. For example, Arthur R. Wyatt (2003, 17), formerly a senior partner and Managing Director–Accounting Principles at Arthur Andersen (a once prominent but now defunct public accounting firm), asserts that “keeping the client happy and doing what was necessary to retain the client achieved a prominence that did not exist prior to the advent of the successful consulting arms with the firms. The core values of the professional firm were undermined by primarily commercial interests.” We can frame this allegation in the context of the model by contrasting the costs of an auditor and a consultant.
appeasing a client. Appeasement in an audit engagement can result in the significant reputation and legal costs associated with an audit failure. Appeasement in a consulting engagement can result in the costs of a botched consulting assignment, which are likely to be much lower than those of an audit failure. Proposition 2 suggests that, if the incentives provided to consulting partners are more powerful than those provided to audit partners, splitting the firm into two organizations, audit and consulting, may be valuable because doing so preserves a high social norm among auditors.

To this point, we have not clearly defined how two organizations (i.e., peer groups) can be created from one. One approach that is likely to create two distinct organizations is to create two distinct firms. Another approach that may create two organizations within the same firm involves putting sufficient distance between employees engaging in different tasks by housing them in different locations. The latter approach is undermined, however, if the employees regularly interact. For example, in a public accounting firm, housing the auditors and consultants in different locations may be ineffective in fostering distinct social norms if the auditors and consultants regularly interact, say in partner/principal meetings or on client engagements.

Proposition 1 suggests that social norm considerations are relevant to organization design decisions. Of course, other factors, such as production and distribution complementarities, internal capital allocation problems, and delegation and monitoring issues, are also relevant. Therefore, it is useful to determine when the cost difference across organization forms due to social norm considerations is large in magnitude. Corollary 1, which follows from inspection of equation (9), provides some relevant insights.

**COROLLARY 1:** The magnitude of the difference in the costs across the two organization structures is greater when (i) the proportions of agents involved in each task are similar, (iii) the difference in desirable actions is greater, and (iii) the agents are more influenced by others in their organization. That is,

\[ \begin{align*}
\partial |C_s(a_h, a_l) - C(a_h, a_l)| / \partial \delta &< 0, \\
\partial |C_s(a_h, a_l) - C(a_h, a_l)| / \partial (a_h - a_l) &> 0, \text{ and} \\
\partial |C_s(a_h, a_l) - C(a_h, a_l)| / \partial \lambda &> 0.
\end{align*} \]
The intuition underlying the three comparative statics in Corollary 1 is as follows. When the proportions of agents engaged in each task are equal, the influence each task group has over the other in the single organization is magnified. Hence, the benefit from adopting the optimal organization design (with respect to the social norm) is magnified. When the desirable action choices associated with the two tasks are further apart, the difference in the intensity of incentive systems across the two tasks is larger. The larger difference in the intensity of incentive systems leads to a greater difference in undesirable actions across the two tasks. A greater difference in undesirable actions, in turn, implies a greater impact on the social norm from combining the two tasks into a single organization. Hence, the benefit from adopting the optimal organization design is larger. Finally, as agents become more sensitive to a social norm (i.e., $\lambda$ increases), the effect of the social norm on individual agent action choices increases. Therefore, the benefit from adopting the optimal organization design increases.

The comparative statics in Corollary 1 provide some insight into when norms are likely to be an important determinant of an organizational break-up. In particular, such a break-up is likely to occur when organizational units are similar in size, have significant differences in optimal incentive schemes, and when the organization is populated by agents who are more sensitive to a social norm.\textsuperscript{10}

III. Implementation Costs and Behavioral Traits

As described in the introduction, individuals differ in how much they are influenced by a social, as opposed to a personal, norm. If individuals differ in their sensitivity to a social norm, two questions naturally arise: First, what type of individuals will an organization prefer; those who are more sensitive to a social norm or those who are less sensitive? Second, will an organization naturally attract the individuals with the preferred trait?
A. Which Type of Agents are Preferred?

To address the first question, consider a variation of the model. Assume agents are identical where \( P \) and \( \lambda \) are the common behavioral parameters, and \( k \) is the common cost incurred by the principal per unit of the undesirable action. Further, assume the principal induces the same desirable action \( a \) from each agent. Finally, let \( C(a) \) denote the minimum per-capita cost of inducing \( a \) for all agents in (6).

One might initially conjecture that individuals who are more influenced by the personal norm (i.e., agents with a low \( \lambda \)) are preferable to those who are more influenced by the social norm (i.e., agents with a high \( \lambda \)), since the former’s action choices are not adversely influenced by peers who undertake high levels of the undesirable action. Such intuition fails to recognize that these agents are also not favorably influenced by peers who undertake low levels of the undesirable action. Proposition 2 shows the level of the desirable action to be induced determines which of these two influences dominates. The proof is in the appendix.

**PROPOSITION 2:** Given identical agents, the cost of implementing \( a \) for all agents in the organization is increasing (decreasing) as agents become more sensitive to the social norm if the desirable action induced is above (below) a critical threshold, \( a_t = 0 \). That is, \( dC(a)/d\lambda < 0 \) if \( a < a_t = 0 \) and \( dC(a)/d\lambda > 0 \) if \( a > a_t = 0 \).

The intuition underlying Proposition 2 stems from the observation that a higher desired action requires greater incentives, which erode the social norm. Hence, if the principal wants to implement a high level of the desired action, he prefers agents who are influenced less by the associated weak social norm. In contrast, when he wants to implement a low level of the desired action, he prefers agents who are influenced more by the associated strong social norm.
B. Will Agents Self-select?

Having identified the agent traits an organization values, we turn to our second question: When agent characteristics are not observable, will individuals self-select into organizations that most value their traits? To develop some insight, we again consider a particular setting. Assume half the agents are more sensitive to the social norm and have $\lambda = \lambda_S$ and the other half are less sensitive with $\lambda = \lambda_P$, where $\lambda_P < \lambda_S$. The agents are otherwise identical with a common personal norm, denoted $P$. The principal must allocate one half the agents to an organization, called $h$, that must induce $a_h$ and the other half to an organization, called $l$ that requires $a_l$, where $a_l < a_t < a_h$. We call this setting the “example setting.”

Consider first the optimal assignment of types to organizations assuming agent traits are observable. When traits are observable, the principal can dictate the assignment of types to organizations. In this case, Proposition 2 implies that the principal prefers to allocate all the agents that are least sensitive to the social norm to organization $h$ and all the agents that are most sensitive to the social norm to organization $l$.

Now suppose agent types are unobservable, so that each agent joins the organization that provides him the highest payoff, where payoff is determined by the contract offered by the organization, the other agents who also join the organization, and the resulting social norm. Two questions are: Do the contracts that are optimal when the principal can assign agents to organizations implement the desired assignments and actions when each agent selects his preferred organization-and-contract pair? If not, what contracts implement the desired actions across the two organizations?

Observation 1, which is proven in the appendix, establishes that the answer to the first question is no. Further, the answer to the second question is that implementing the desirable actions requires contracts that attract the less-desired type to each organization. That is, the assignment of types to organizations is opposite to the assignment preferred by the principal.

**OBSERVATION 1:** In the example setting, it is optimal to assign agents who are least sensitive to a social norm to organization $h$ and those who are most sensitive to a
social norm to organization \( l \) when types are observable. When agents must self-select, the only contracts that implement the desirable actions cause the agents who are most sensitive to the social norm to join organization \( h \) and those who are least sensitive to the social norm to join organization \( l \).

The self-selection problem identified in Observation 1 arises because the principal wants the agents most sensitive by the social norm to join the organization with the highest social norm (i.e., organization \( l \)), but the agents who are most sensitive to the social norm are first to be attracted to an organization with the lowest social norm (i.e., organization \( h \)). More specifically, the social norm in organization \( h \) is lower than the social norm in organization \( l \) irrespective of how agents are matched to organizations. Moreover, any differential in social norms exerts greater influence on the preference ranking of agents who are most sensitive to a social norm. Hence, the lower social norm in \( h \) attracts the \( \lambda_S \) agents whenever it attracts the \( \lambda_P \) agents, but not vice versa. Consequently, organization \( h \) attracts the agents that are most sensitive to the social norm.

Observation 1 implies that organizations with higher-powered incentives prefer agents who are least sensitive to a social norm, but they may attract those who are most sensitive to a social norm. Organizations with lower-powered incentives prefer agents that are most sensitive to a social norm, but they may attract those who are least sensitive to a social norm.\(^{11}\)

Two caveats to Observation 1 are in order. First, underlying Observation 1 is the assumption that the two desirable action levels lie on opposite sides of the threshold. In settings where the two desirable action lie on the same side of the threshold, Observation 1 may not hold. In particular, when both actions lie on the same side of the threshold, it is possible that the optimal assignment of agent types to organizations can be implemented when types are unobservable. In these cases, however, the optimal matching can only be achieved when one agent type earns rents.\(^{12}\)

Second, Observation 1 applies to a setting where the agents are assigned to one of two organizations, each of which induces all agents in the organization to implement
the same level of the desirable action. If instead the agents are assigned to execute one of two desirable action levels within a single organization, the optimal assignment of types to desirable action levels when type is observable depends on the difference in the per-unit costs to the principal of the undesirable action associated with each desirable action, and the difference in the levels of the desirable actions to be implemented. When type is unobservable, the optimal assignment of types to tasks is not implementable in some cases, is implementable at a cost in others, and is implementable at no cost in yet others. Thus, the problem of self-selection can arise within a single organization as well.

We also point out that issues of self-selection are likely to arise when agents have different personal norms, (i.e., the $P_i$’s differ from each other). The principal always prefers agents with a higher personal norm. Furthermore, the value of a high personal norm to the principal is greatest when the desirable action to be implemented is high. Agents with lower personal norms, however, are always more eager to accept a contract with higher monetary incentives, all else equal. Hence, a principal who wants to induce a high level of the desired action may not succeed in attracting the agents with the most desired personal norms. We have focused on differences in agent sensitivity to social norms because it is more interesting analytically (i.e., the principal’s preferences vary with the desired action), and that sensitivity is measured by social psychologists, including Ponemon and Gabhart (1994).

IV. Norms for Desirable Actions

To this point we have focused on the role played by norms in thwarting undesirable actions. With respect to this type of norm, we have shown that: the presence of a social norm undermines the effectiveness of monetary incentives; the presence of the norm can make it optimal to divide an organization when the cost to the principal of undesirable actions is high for low-productivity agents; agents who are less sensitive to a social norm are more valuable when the power of incentives is high; and a mismatch between agent types and incentive intensity can arise. Alternatively, norms may promote the desirable action. For example, consider a work ethic. Within our framework, a work ethic is a
norm that influences the cost of the desirable action. In particular, an organization with a strong work ethic is one in which the cost of the desirable action is lower for the agents. In this section, we reconsider our results when the norm reinforces the desirable action. The proofs of the results in this section parallel the proofs of earlier results, and so are omitted.

We alter the set-up as follows: the cost to the agent of the desirable and undesirable actions are \( f(a_i - N_i) \) and \( f(u_i) \), respectively, and the social norm for organization composed of agents in \( J \) satisfies \( S \equiv \frac{\int_{j \in J} a_j dj}{\int_{j \in J} dj} \). These assumptions imply that an agent is more willing to undertake the desirable action if others undertake more of that action. Finally, to ensure that the agents’ action choices are interior, we assume that \( \min(P, \bar{\lambda}a + (1 - \bar{\lambda})P) \geq 0 \) and \( u \leq a \).

As in the case where the norm affects the cost of the undesirable action, there exists a unique post-contracting equilibrium for an organization and set of contracts.

**Lemma 1':** For any set of contracts, \( \{w_j, b_j\}_{j \in J} \) for organization \( J \), there exists a unique post contracting equilibrium, \( \{a_j, u_j\}_{j \in J} \). In equilibrium, \( a_i = u_i - N_i \) and \( S = \frac{\int_{j} a_j dj}{\int_{j} dj} \).

Similar to the setting where the norm applied to the undesired action, a greater value for the norm induces the agent to shift the action mix towards the desirable action. The shift occurs, however, because a higher norm makes the desired action less costly to the agent, as opposed to making the undesirable action more costly to the agent. In addition, the social norm is, by definition, simply equal to the average level of the desired action. Finally, in contrast to the setting where the norm thwarted the undesirable action, the presence of the social norm enhances the effectiveness of monetary incentives. In particular, the direct effect of increasing the power of individual incentives on the desirable action is magnified, and the effect on undesirable action is muted, by an indirect effect associated with increases in the social norm.

First we reconsider whether it is optimal to have one organization or two when proportion \( \delta \) of the agents are employed to exploit a high marginal production opportunity and proportion \( 1 - \delta \) are employed to exploit a low marginal production opportunity.
PROPOSITION 1′: Given a norm that applies to desirable actions, there exists a $q > 0$ such that the costs associated with separate organizations are less than the costs associated with a single organization if and only if $k_l - k_h < q$.

As in the case where the norm applies to the undesirable action, the optimal organizational design depends on the difference between the cost of undesirable action for organization $h$, $k_h$, and organization $l$, $k_l$; however, separate organizations are preferred when the difference, $k_l - k_h$, is low rather than high. The reason is the indirect effect of the norm on the undesirable action. A higher value for the norm reduces the undesirable action by making the desirable action relatively less costly to the agents. With separate organizations, the high-production agents face a higher norm and the low-production agents face a lower norm than with a single organization. Hence, when the cost of undesirable actions by high-production agents is high, it is beneficial to have separate organizations. Together, Propositions 1 and 1′ illustrate how norms affect organizational design. How a norm influences the organization structure depends upon whether the norm discourages undesirable actions or encourages desirable actions.

Consider next the matching of agent types to tasks. When the norm affects the desirable action, we obtain a result that is opposite to that attained when the norm affects the undesirable action—agents that are more responsive to a social norm are preferred when the level of the desirable action (and the associated power of incentives) is high, and agents that are less responsive to a social norm are preferred when it is low. This flip arises because the benefit derived from the social norm is positively, rather than negatively, related to the power of incentives necessary to induce the desirable action (i.e., a higher level of the desirable action is associated with a more effective social norm). Hence, agents who are more (less) responsive to the social norm are preferred when the power of incentives necessary to induce the desirable action is high (low). More formally, we have:

PROPOSITION 2′: Given identical agents and a norm that applies to desirable actions, the cost of implementing $a$ for all agents in the organization is decreasing (increasing) as the agents become more sensitive to the social norm if the desirable
action is above (below) a critical threshold, $a'_t = P$. That is, $dC(a)/d\lambda < 0$ if $a < P$ and $dC(a)/d\lambda > 0$ if $a > P$.

Whether agents naturally self-select hinges on whether the norm pertains to the desirable or undesirable action. We develop our insights from an example setting that parallels the one from the case where the norm applies to the undesirable action. Assume that half the agents are more sensitive to a social norm with $\lambda = \lambda_S$ and the other half are less sensitive with $\lambda = \lambda_P$, where $\lambda_P < \lambda_S$. The agents are otherwise identical with personal norm denoted $P$ and the principal must allocate half the agents to an organization $h$ that must induce $a_h$ and the other half to organization $l$ that must induce $a_l$, where $a_l < a'_t < a_h$. Observation 1’ indicates that there is no incremental cost incurred when type is unobservable because agents naturally select the organization which prefers their type.

**OBSERVATION 1’**: In the example setting with a norm that applies to the desirable action, it is optimal to assign agents that are most sensitive to a social norm to organization $h$ and those that are least sensitive to a social norm to organization $l$. The optimal contracts when type is observable also implement the optimal agent assignments and desirable actions when type is not observable.

In contrast to the case when norms apply to the undesired actions, Observation 1’ suggests that self-selection work in the principal’s favor when norms apply to desired actions. The reason is that, when norms apply to desired actions, the principal wants the agents most sensitive to the social norm to join the organization with the highest social norm (i.e., organization $h$). Happily from the principal’s perspective, the agents who are most sensitive to the social norm are also first to be attracted to an organization with the highest social norm (i.e., organization $h$).
V. Conclusion

We derive some implications of social norms for organizational design by analyzing a model where a principal employs a continuum of agents who each choose two actions, a desirable action that is beneficial to the principal and an undesirable action that is costly to the principal. Each agent chooses a mix of these actions because both favorably affect the performance measure used for contracting with that agent. The critical aspect of social norms in our model is that an agent’s behavior is influenced by the behavior of others in his organization. In particular, the cost of an action to an agent is reduced as other agents engage in more of that action.

In our primary analysis, the cost of undesired action to an individual agent is decreasing as other agents engage in more of the undesirable action. In that analysis we highlight three points. First, the presence of a social norm magnifies the impact of increased financial incentives on undesired actions and mitigates the impact of increased financial incentives on desired actions. Second, when agents with high-powered incentives are influential in setting social norms, splitting an organization into two independent organizations may be beneficial. Splitting the organization avoids the adverse impact the agents with high-powered incentives exert on the behavior of the organization’s other agents. Third, agents that are more sensitive to a social norm are preferred when incentives are low-powered because they are heavily influenced by the relatively good behavior of their peers. In contrast, agents that are insensitive to a social norm are preferred when incentives are high-powered because they are hardly influenced by the relatively bad behavior of their peers. Unfortunately, when agents cannot be assigned to organizations because behavioral traits are not observable, agents do not necessarily self-select into the organizations that most value their traits.

We also analyze a setting in which the norm applies to the desirable action. This alternative setting offers different insights, suggesting that the nature of the norm is important to its role in organizational design. In particular, we find that the norm magnifies the impact of increased financial incentives on the desired action and mitigates the impact on undesired actions; splitting up an organization may be beneficial in
different circumstances than in the primary setting; agents that are more (less) sensitive to a social norm are preferred when incentives are high-powered (low-powered); and agents will naturally self-select into the organizations that most value their traits.

Because the norms affecting undesired actions and desired actions can have different implications, determining which type of norm is dominant is critical to assessing how norms are expected to influence organization design choices. Although such a determination awaits empirical testing, we conjecture that norms for undesired actions are likely to be more important in settings where other barriers to performance measure manipulation are weak or absent. For example, we might expect that norms relating to undesirable actions are relatively more important when: the firm’s other monitoring mechanisms, such as an active independent board, are weak; the firm engages in transactions where the accounting is complex and requires significant estimates; and when the firm has significant leverage over its suppliers and customers, which makes it easier to use them to manipulate reports.

Finally, we have restricted our attention to the implications of social norms for organizational design. Social norm considerations, however, are relevant to other issues of economic interest such as understanding how technological changes undermine legal systems and honor codes that are supported, in part, through social norms. For example, the development of the internet and the digitization of copyrighted materials has facilitated the undesirable acts of copyright violation (e.g., music file sharing) and plagiarism not only by making such acts physically easier, but also by eroding the norms that discouraged such acts.
APPENDIX

PROOF OF LEMMA 1: Provided \( S \) is finite and less than \(-u\), it follows from the assumed properties of \( P, f, h, \) and \( u \) that agent \( i \)'s optimal choices for \( a_i \) and \( u_i \) are interior for \( b_i > 0 \). Hence, the first-order conditions on (1) with respect to \( a_i \) and \( u_i \) uniquely characterize agent \( i \)'s optimal choices as a function of the contract parameters and the norm. These conditions, 

\[ b_i h'(a_i + u_i) = f'(a_i) \quad \text{and} \quad b_i h'(a_i + u_i) = f'(u_i + N_i) \]

imply that \( a_i = u_i + N_i \), and so they may be rewritten:

(A1) \[ 0 = b_i h'(2a_i - N_i) - f'(a_i), \] \quad \text{and} \quad \[ (A2) 0 = b_i h'(2u_i + N_i) - f'(u_i + N_i). \]

Equations (A1) and (A2) allow \( a_i \) and \( u_i \), respectively, to be written as implicit functions of \( S \). Let \( u_i(S) \) be the implicit function for \( u_i \) derived from (A2). To prove the existence of a unique equilibrium, we show there exists a unique finite value for \( S \leq -u \) such that (3) is satisfied when the set of undesirable actions is determined by the implicit function characterized by (A2). Equilibrium requirement (3) is satisfied by a finite \( S \) and the associated implicit functions characterized by (A2) if and only if

\[ Q(S) \equiv S + \frac{\int_j u_j(S) dj}{\int_j dj} = 0. \]

The derivative of \( Q \) with respect to \( S \) is

(A3) \[ Q'(S) = 1 - \left( \int_j \lambda_j \frac{b_j h''(2u_j + N_j) - f''(u_j + N_j)}{2b_j h''(2u_j + N_j) - f''(u_j + N_j)} dj / \int_j dj \right). \]

By inspection, \( Q'(S) > 0 \), so \( Q(S) \) increases monotonically in \( S \). Furthermore, \( Q(-u) > 0 \) because all agents choose a level of undesirable action that is in the interior (i.e., greater than \( u \)) for all \( S \leq -u \). Hence, if a finite \( S < -u \) that satisfies \( Q(S) = 0 \) exists, it is unique. It also follows from (A3), that \( 1 - \left( \int_j \lambda_j / \int_j dj \right) \leq Q'(S) \leq 1 \) for all \( S \). Hence \( Q(S) \) is bounded above by \( Q(-u) + (1 - \lambda_j)(S + u) \) for \( S \leq -u \). There must
exist an $S < -u$ such that $Q(S) = 0$ because there exists a finite $S < -u$ such that $Q(-u) + (1 - \overline{X}_j)(S + u) = 0$.

The observation that $a_i = u_i + N_i$ for all $i$ follows directly from the first order conditions for each agents' optimization problem. To characterize $S$ as a function of the desired actions induced, we first note that the relation $a_i = u_i + N_i$ implies

$$a_i - N_i = a_i - (1 - \lambda_i)P_i - \lambda_i S = u_i$$

for all $i$ in any organization. It follows that the social norm for any organization, which is minus the average level of undesirable activity across the organization, satisfies

$$\int_J a_j dj - \int_J ((1 - \lambda_j) P_j) dj - S \int_J \lambda_j dj = -S \int_J dj.$$ 

Solving this equation for $S$ yields

$$S = -\frac{\int_J a_j dj - \int_J (1 - \lambda_j) P_j dj}{\int_J dj - \int_J \lambda_j dj}.$$ 

**PROOF OF PROPOSITION 2:** From (4), the social norm for each agent can be written as $S = -a/(1 - \lambda) + P$. This observation, coupled with the observation that $a_i = u_i + N_i$ for all $i$, the cost in (6) is $v + 2f(a) + k(a/(1 - \lambda) - P)$. Differentiating the cost with respect to $\lambda$ yields $ka/(1 - \lambda)^2$, which is less than 0 if $a < 0$ and is greater than 0 if $a > 0$.

**PROOF OF OBSERVATION 1:** The result follows from four lemmas:

**LEMMA A1:** In the example setting, the optimal assignment of types to organizations when type is observable has all the $\lambda_S$ assigned to organization $l$ and the $\lambda_P$ agents assigned to organization $h$.

**PROOF OF LEMMA A1:** Let $\delta$ denote the proportion of $\lambda_S$ agents assigned to organization $l$. By default, $\delta$ also denotes the proportion of $\lambda_P$ agents assigned to organization $h$. Taking $\delta$ as given and assuming that contracts for each type are structured to implement $a_h$ by agents in organization $h$ and $a_l$ by agents in organization
l, the post-contracting equilibrium conditions imply that the per capita of implementing these actions total

\[ v + f(a_h) + f(a_l) + \frac{1}{2} k_h \left( \frac{a_h}{1 - (1 - \delta) \lambda_S - \delta \lambda_P} - P - u \right) \]

\[ + \frac{1}{2} k_l \left( \frac{a_l}{1 - \delta \lambda_S - (1 - \delta) \lambda_P} - P - u \right). \]

Differentiating this cost with respect to \( \delta \) yields

\[ - \frac{(\lambda_S - \lambda_P)}{2} \left( \frac{a_h k_h}{(1 - (1 - \delta) \lambda_S - \delta \lambda_P)^2} - \frac{a_l k_l}{(1 - (1 - \delta) \lambda_P - \delta \lambda_S)^2} \right) < 0, \]

because \( a_l < 0 < a_h \). Therefore, the optimal proportion of \( \lambda_S \) agents allocated to \( l \) is 1 and the optimal proportion of \( \lambda_P \) agents allocated to organization \( h \) is 1.

**LEMMA A2:** In the example setting, if \( a_h \) and \( a_l \) are implemented for any \( \delta \), it must be the case that the social norm in organization \( h \), \( S_h \), and the social norm in organization \( l \), \( S_l \), satisfy: \( S_l > P > S_h \).

**PROOF OF LEMMA A2:** Given \( \delta \), the social norms satisfy \( S_h = P - a_h / (1 - (1 - \delta) \lambda_S - \delta \lambda_P) < P \), and \( S_l = P - a_l / (1 - (1 - \delta) \lambda_P - \delta \lambda_S) > P \), because \( a_h > 0 > a_l \).

**LEMMA A3:** In the example setting, \( a_h \) and \( a_l \) cannot be implemented for any \( \delta > 0 \) if type is not observable.

**PROOF LEMMA A3:** The proof is by contradiction. Let \( v_{itj} \) denote the expected utility that type \( t \in \{ S, P \} \) (where the subscript \( S \) denotes \( \lambda_S \) agents and the subscript \( P \) denotes \( \lambda_P \) agents) attains by accepting the contract offered by organization \( i \) that implements the desirable action \( a_i \) by type \( j \). Let \( N_{it} \) denote the norm of type \( t \) if he joins organization \( i \). From Lemma 2, \( S_l > P > S_h \) coupled with \( \lambda_S > \lambda_P \), implies that \( N_{hc} < N_{hp} \) and \( N_{lp} < N_{lc} \). In turn, \( N_{hc} < N_{hp} \) and \( N_{lp} < N_{lc} \), imply that \( v_{hcp} > v_{hpp} \) and \( v_{lpc} > v_{lcc} \). These observations, coupled with the assumption some nonconformists are in \( h \) and some conformists are in \( l \), imply that

\[
\begin{align*}
v_{hpp} & \geq v_{lpp} > v_{lpc}, \\
v_{lcc} & \geq v_{hcc} > v_{hcp}.
\end{align*}
\]
These two conditions, coupled with the observations \( v_{hcp} > v_{hpp} \) and \( v_{lpc} > v_{lcc} \), imply

\[
v_{hpp} \geq v_{lpp} \geq v_{lpc} \geq v_{lcc} \geq v_{hcc} \geq v_{hcp} > v_{hpp}, \quad \text{and} \quad v_{lcc} \geq v_{hcc} \geq v_{hcp} \geq v_{hpp} \geq v_{lpp} \geq v_{lpc} > v_{lcc},
\]

both of which cannot be true.

**Lemma A4:** In the example setting, \( a_h \) and \( a_l \) can be implemented for \( \delta = 0 \) if type is not observable.

**Proof of Lemma A4:** Let \( h \) offer a single contract \( w_h, b_h \) and \( l \) offer a contract \( w_l, b_l \). Let \( b_h \) and \( b_l \) satisfy the following conditions:

\[
b_h h'(2a_h - (1 - \lambda_S)P - \lambda_S S_h) = f'(a_h), \quad \text{and} \quad b_l h'(2a_l - (1 - \lambda_P)P - \lambda_P S_l) = f'(a_l),
\]

where \( S_h = P - a_h/(1 - \lambda_S) \) and \( S_l = P - a_l/(1 - \lambda_P) \). Given that the contracts satisfy these two conditions, and recalling that \( a_i = u_i + N_i \) for any agent \( i \) in an organization, \( a_h \) is implemented by the \( \lambda_S \) agents if they all join organization \( h \), and \( a_l \) is implemented by the \( \lambda_P \) agents if they are all join organization \( l \). Taking the solution values for \( b_h \) and \( b_l \) from the above conditions, let \( w_h \) and \( w_l \) satisfy

\[
w_h = v - b_h h(2a_h - (1 - \lambda_S)P - \lambda_S S_h) + 2f(a_h), \quad \text{and} \quad w_l = v - b_l h(2a_l - (1 - \lambda_P)P - \lambda_P S_l) + 2f(a_l).
\]

Recalling that \( a_i = u_i + N_i \) for any agent \( i \) in an organization, these additional two conditions imply that each type attains the reservation level of expected utility when the \( \lambda_S \) agents are all aligned with organization \( h \) and the \( \lambda_P \) agents are all aligned with organization \( l \). Finally, from Lemma 3, \( v = v_{hcc} > v_{hpc} \) and \( v = v_{lpp} > v_{lcp} \), so the \( \lambda_S \) agents strictly prefer to be in organization \( h \) and the \( \lambda_P \) agents strictly prefer to be in organization \( l \).
References


Notes

1. See, for example, work by Julian B. Rotter (1966) and Lawrence Kohlberg (1984).
2. There are undoubtedly other counterproductive actions that can be employed to enhance one's compensation. For example, Canice Prendergast (1999) discusses how counterproductive influence activities may be employed to improve subjective evaluation measures.
4. The literature on managerial myopia considers the implications of agents altering real, as opposed to accounting, decisions to enhance reported performance; see, for example, Jeremy Stein (1989) or David Hirshleifer (1993). For empirical evidence that some managers sacrifice real resources to inflate earnings, see Merle Erickson, Michelle Hanlon, and Edward Maydew (2004) and Paul Oyer (1998).
5. Laboratory and field studies in the social psychology literature suggest that individuals differ in the extent to which they are influenced by social norms. See, for example, Lawrence Kohlberg (1984) and Denis Arnold and Lawrence Ponemon (1991).
6. Models of reciprocation have interdependent preferences, which is a feature analogous to our social norm. Ernst Fehr and Armin Falk (2002) distinguish between
two human social desires that interact with economic incentives: the desire to reciprocate—see, e.g., Gary Bolton and Axel Ockenfels (2000) and Gary Charness and Matthew Rabin (2000), who examine how notions of equity and fairness influence behaviors—and the desire for social approval. These models do not address social norms with respect to unobservable acts.

7. Our assumption that the social norm only influences an agent’s cost of the undesired action implies that, all else equal, an agent always prefers to be in an organization with a lower social norm. More generally, being part of an organization with a low social norm also may impose a direct cost on an agent because he may be treated less well by his colleagues. We thank a referee for pointing this out to us.

8. Like many economic models, establishing the existence and uniqueness of equilibrium relies critically on the chosen functional forms. Indeed, there exists a seemingly reasonable alternative functional form that can result in multiple equilibria in some cases and nonexistence of a finite equilibrium in others: Consider an organization comprised of agents indexed on \([0, 1]\). Assume all agents are identical and have objective \(b(au)^\beta - a - Nu\), where \(\beta < 1/2\), \(N = P + 1/S\), \(S = \int_0^1 u_jdj\), \(P > 0\), and \(\{a, u\} \in \mathbb{R}_+^2\). This alternative is similar to our model in that the cost of the undesirable action for a particular agent is increasing in a norm and that norm is decreasing in the actions of others. In this alternative setting in which the actions are complements, as opposed to substitutes, there may not exist an equilibrium in
which $S$ is finite for some parameterizations of this model and there may exist two equilibria for others.

9. See also the comments of Securities and Exchange Commission Chairman Arthur Levitt (2000).

10. We should note that Corollary 1 is derived from an analysis of cost differences when the desirable action mix is constant across the single and separate organization structures. Norm management considerations, however, may cause the mix to optimally differ across the two structures. Nonetheless, the insights offered by Corollary 1 should still hold, although the difference in the principal’s welfare attributable to deviations from the optimal structure will be muted by alterations in the mix of desirable actions in response to norm considerations.

11. Survey and anecdotal evidence suggests that individuals’ attitudes towards undesirable earnings management activities varies across organizations. In a survey of managers from many different corporations, William J. Bruns and Kenneth A. Merchant, Jr. (1990) document a wide range of attitudes towards various types of earnings management (i.e., making operational decisions, timing the recognition of non-operational transactions, or choosing accounting methods so as to manipulate reported earnings) across firms. Lawrence Ponemon and David Gabhart (1994) survey the evidence that the degree to which individuals, on average, are sensitive to norms varies across professions, and personnel selection and promotion reduce the variation in type within public accounting firms, especially among more senior members of
the firm. For an anecdotal account of the varied responses of lower-level managers to top-level managers’ adoption of accounting treatments that violate generally accepted accounting principles, see Susan Pulliam and Deborah Solomon (October 30, 2002) “Uncooking the Books: How Three Unlikely Sleuths Discovered Fraud at WorldCom” Wall Street Journal.

12. For example, assume $a_t < a_l < a_h$ and assume the optimal assignment of types to organizations when type is observable is for the agents that are most sensitive to the social norm to be in organization $l$. Such a matching is implementable when type is not observable. Note also that the optimal matching of types to organizations when type is observable may not be the assumed matching when $a_t < a_l < a_h$.

The optimal matching is a function of the difference in the two costs of undesirable action, the difference in the two desirable action levels, and the difference in the $\lambda$'s across types. Furthermore, in some cases the optimal alignment involves mixing types across organizations.

13. As an aside, note that the threshold difference is (i) 0 for a norm applying to the undesirable action but (ii) is strictly positive for a norm applying to the desirable action. In the latter case, the threshold is strictly positive because the cost of the desirable action is increasing and strictly convex in $a - N$. As a result, ignoring the cost of undesirable actions, the benefit of increasing the norm for the high-productivity agents when the two organizations are separate more than (rather than exactly) offsets the cost from decreasing the norm for the low-productivity agents when the two organizations are separate.
Supplementary Notes

CLAIM: For any organization composed of agents in J, a linear contract can implement any interior set of actions, \( \{a_j^*\}_{j \in J} \), at a per-capita expected cost that is less than or equal to any other contract form.

PROOF OF CLAIM: Consider any set of compensation contracts \( \{\omega_j(r)\}_{j \in J} \) that implement an interior set of desirable actions \( \{a_j^*\}_{j \in J} \). Let the post-equilibrium outcomes for this contract be denoted with a superscript * and let \( C_i^* = \int_R \omega_i(r)g(r; a_i^* + u_i^*)dr \) where \( R \) is the support of \( r \). Given the social norm induced by \( \{\omega_j(r)\}_{j \in J} \), the first-order conditions for each agent’s action choice problem are

\[
\frac{\partial C_i^*}{\partial a_i} = f'(a_i^*) \\
\frac{\partial C_i^*}{\partial u_i} = f'(u_i^* + N_i^*)
\]

Because the density function for \( r \) can be written as \( g(r; a_i + u_i) \), it follows that

\[
\frac{\partial C_i^*}{\partial a_i} = \frac{\partial C_i^*}{\partial u_i}.
\]

This observation, coupled with the first-order conditions imply that \( a_i^* = u_i^* + N_i^* \) for all \( i \). The fact that \( a_i^* = u_i^* + N_i^* = u_i^* + \lambda_i S^* + (1 - \lambda_i) P_i \) and the post-contracting equilibrium requirement for the social norm given by equation (3) imply that

\[
S^* = -\frac{\int_J a_j^*dj - \int_J (1 - \lambda_j) P_jdj}{\int_J dj - \int_J \lambda_j dj}.
\]

Furthermore, it must be the case that each agent’s expected utility satisfies equation (1).

It follows that the per capita expected cost of this contract must be weakly greater than

\[
\int_0^1 [v + f(a_j^*) + f(u_j^* + N_j^*) - k_j(u_j^* - w)]dj.
\]

The proof is completed by constructing a linear contracting scheme, \( \{w_j^*, b_j^*\}_{j \in J} \) that implements \( \{a_j^*\}_{j \in J} \) at a weakly lower cost than \( \{\omega_i(r)\}_{i \in [0,1]} \). Consider a linear contract with a set of slope coefficients, \( \{b_j^*\}_{j \in J} \), such that \( b_i \) satisfies

\[
b_i^* h'(2a_i^* - N_i^*) - f'(a_i^*) = 0
\]

and

\[
b_i^* h'(2u_i^* + N_i^*) - f'(u_i^* + N_i^*) = 0.
\]
for all $i$. Such a $\{b_j^*\}_{j \in J}$ exists because $a_i^* = u_i^* + N_i^*$. Furthermore choose the constant coefficients, $\{w_j^*\}_{j \in J}$, such that $w_i^* + b_i^*h(a_i^* + u_i^*) - f(a_i^*) - f(u_i^* + N_i^*) \geq v$, for all $i$. Equations (B2) and (B3) imply that such a contract must implement $\{a_j^*, u_j^*\}_{j \in J}$ given $S^*$ and must yield each agent their reservation utility. Furthermore, given that the contract implements $\{u_j^*\}_{j \in J}$, it follows that $S^*$ satisfies equation (3), the post-contracting equilibrium requirement for the social norm. The fact that each agent attains his expected utility implies that the per capita expected cost of $\{w_j^*, b_j^*\}_{j \in J}$ is $(\int J v + f(a_j^*) + f(u_j^* + N_j^*) - k_j(u_j^* - u) dj) / \int J dj$, which is weakly less than the per capita cost associated with $\{\omega_j(r)\}_{j \in J}$.

CLAIM: Consider an organization comprised of agents indexed on $[0, 1]$. Assume all agents are identical and have objective $b(au)^{\beta} - a - Nu$, where $\beta < 1/2$, $N = P + 1/S$, $S = \int_0^1 u_j dj$, $P > 0$, and $\{a, u\} \in \mathbb{R}_+^2$. For some contracts there does not exist an equilibrium in which $S$ is finite, for others there exists two post-contracting equilibria in which $S$ is finite, and for some knife-edged contracts there exists a unique post-contracting equilibrium in which $S$ is finite.

PROOF OF CLAIM: Because agents are identical, we omit subscripts. Given any contract and any strictly positive value for the social norm, an agent’s action choices satisfy the following first order conditions: $b\beta a^{\beta - 1} u^\beta = 1$ and $b\beta a^{\beta} u^{\beta - 1} = N$. These conditions imply that $a = Nu$. Exploiting this observation allows us to derive the following solution values for $a$ and $u$:

$$a = (b\beta)^{-\frac{1}{1-2\beta}} N_{-\frac{\beta}{1 - 2\beta}},$$

and

$$u = (b\beta)^{-\frac{1}{1-2\beta}} N^{\frac{\beta - 1}{1 - 2\beta}}.$$ 

Let $F = (b\beta)^{-\frac{1}{1-2\beta}} > 0$ and $G = (\beta - 1)/(1 - 2\beta) < 0$. Given these characterizations of the action choices as a function of the social norm and contract parameters, there exists
an equilibrium characterized with social norm $S$ that satisfies the equilibrium condition $S = \int_0^1 u_j dj$ where $u_j = FN^G$ for all $j$. Such an equilibrium $S$ exists if and only if

$$F(P + 1/S)^G - S = 0.$$ 

This condition is equivalent to

$$Q(S) = F^{1/G} P + F^{1/G} S^{-1} - S^{1/G} = 0.$$ 

Note that $Q(0) > 0$, $Q'(0) = 0$, $Q' = 0$ at $S^* = (-G)^{\frac{1}{1+G}} F^{\frac{1}{1+G}} > 0$. Furthermore, $Q'' > 0$ at any $S > 0$ such that $Q'(S) = 0$. Hence, it follows that $Q(S)$ decreasing and then increasing in $S$, and is minimized at $S^*$. Therefore, there exists no equilibrium in cases where $Q(S^*) > 0$, two equilibria in the case where $Q(S^*) < 0$, and a unique equilibrium in the knife-edged case where $Q(S^*) = 0$.

**PROOF OF LEMMA 1':** Provided $S$ is finite and weakly greater than $a$, it follows from the assumed properties of $P$, $f$, $h$, and the bounds on the action choice sets that agent $i$’s optimal choices for $a_i$ and $u_i$ are interior for $b_i > 0$. Hence, the first-order conditions for $a_i$ and $u_i$ uniquely characterize agent $i$’s optimal choices as a function of the contract parameters and the norm. These conditions, $b_i h'(a_i + u_i) = f'(a_i - N_i)$ and $b_i h'(a_i + u_i) = f'(u_i)$ imply that $a_i - N_i = u_i$, and so they may be rewritten:

(B4) $$0 = b_i h'(2a_i - N_i) - f'(a_i - N_i),$$
(B5) $$0 = b_i h'(2u_i + N_i) - f'(u_i).$$

Equations (B4) and (B5) allow $a_i$ and $u_i$, respectively, to be written as implicit functions of $S$. Let $a_i(S)$ be the implicit function for $a_i$ derived from (B4). To prove the existence of a unique equilibrium, we show there exists a unique finite value for $S \geq a$ such that the equilibrium requirement for $S$ is satisfied when the set of desirable actions is determined by the implicit functions characterized by (B4). The equilibrium requirement
is satisfied by a finite \( S \geq a \) and the associated implicit functions characterized by (B2) if and only if

\[
Q(S) \equiv S - \frac{\int_j a_j(S) dj}{\int_j dj} = 0.
\]

The derivative of \( Q \) with respect to \( S \) is

\[
Q'(S) = 1 - \left( \int_J \lambda_j \frac{b_j h''(2a_j - N_j) - f''(a_j - N_j)}{2b_j h''(2a_j - N_j) - f''(a_j - N_j)} dj \right) / \int_J dj
\]

By inspection, \( Q'(S) > 0 \), so \( Q(S) \) increases monotonically in \( S \). Furthermore, \( Q(a) < 0 \) because all agents choose a level of undesirable action that is in the interior (i.e., greater than \( a \)) for all \( S \geq a \). Hence, if a finite \( S > a \) that satisfies \( Q(S) = 0 \) exists, it is unique. It also follows from (B6), that

\[
1 - \left( \int_J \lambda_j dj / \int_J dj \right) \leq Q'(S) \leq 1
\]

for all \( S \). Hence \( Q(S) \) is bounded below by \( Q(a) + (1 - \bar{\lambda} dj)(S - a) \) for all \( S \geq a \). There must exist an \( S > a \) such that \( Q(S) = 0 \) because there exists a finite \( S > a \) such that \( Q(a) + (1 - \bar{\lambda} dj)(S - a) \) for all \( S \geq a = 0 \). Finally, the observation that \( a_i - N_i = u_i \) for all \( i \) follows directly from the first order conditions for each agents’ optimization problem and the observation that \( S = \int_{j \in J} a_j dj / \int_J dj \) follows directly from the definition of \( S \) in this setting.

**PROOF OF PROPOSITION 1’:** As in Proposition 1, suppose some proportion, \( \delta \), of the agents are employed to exploit a high marginal production opportunity and the remaining proportion \( 1 - \delta \) are employed to exploit a low marginal production opportunity. The cost of implementing \( a_h \) and \( a_l \) in a single organization is

\[
C_1 = v + 2\delta f(a_h - (1 - \lambda)P) - \lambda(\delta a_h + (1 - \delta) a_l)) + \delta k_h (a_h - (1 - \lambda)P - \lambda(\delta a_h + (1 - \delta) a_l) - u) + 2(1 - \delta) f(a_l - (1 - \lambda)P - \lambda(\delta a_h + (1 - \delta) a_l)) + (1 - \delta) k_l (a_l - (1 - \lambda)P - \lambda(\delta a_h + (1 - \delta) a_l) - u).
\]
The cost of implementing \( a_h \) and \( a_l \) in separate organizations is
\[
C_2 = V + 2\delta f (a_h - (1 - \lambda)P - \lambda a_h) \\
+ \delta k_h (a_h - (1 - \lambda)P - \lambda a_h - u) \\
+ 2(1 - \delta) f (a_l - (1 - \lambda)P - \lambda a_l) \\
+ (1 - \delta) k_l (a_l - (1 - \lambda)P - \lambda a_l - u).
\]

It is straightforward to show that
\[
(C7) \quad C_1 - C_2 = Z - \alpha \delta \lambda (1 - \delta)(k_l - k_h),
\]
where
\[
Z = 2\delta \left[ f (H + \alpha (1 - \delta \lambda)) - f (H + \alpha (1 - \lambda)) \right] + 2(1 - \delta) \left[ f (H - \alpha \delta \lambda) - f (H) \right],
\]
\[
H = (a_l - P)(1 - \lambda), \quad \text{and}
\]
\[
\alpha = a_h - a_l.
\]
Because (B7) is linear and decreasing in \( k_l - k_h \), there is a threshold value \( q \) such that the costs associated with separate organizations are less than the costs associated with a single organization if and only if \( k_l - k_h < q \). We show that the threshold \( q \) is strictly positive by showing that \( Z > 0 \). Note that \( Z = 0 \) when \( \alpha = 0 \). Hence, \( \partial Z / \partial \alpha > 0 \) implies \( Z > 0 \).

\[
\frac{\partial Z}{\partial \alpha} = 2\delta \left[ (1 - \delta \lambda)f' (H + \alpha (1 - \delta \lambda)) - (1 - \lambda)f' (H + \alpha (1 - \lambda)) \right] \\
- 2(1 - \delta) \delta \lambda f' (H - \alpha \delta \lambda) \\
= 2\delta (1 - \lambda) \left[ f' (H + \alpha (1 - \delta \lambda)) - f' (H + \alpha (1 - \lambda)) \right] \\
+ 2\delta (1 - \delta) \lambda \left[ f' (H + \alpha (1 - \delta \lambda)) - f' (H - \alpha \delta \lambda) \right].
\]

Because \( \delta \in (0, 1) \) and \( \lambda \in (0, 1) \), \( (1 - \delta \lambda) > (1 - \lambda) \) and \( 1 - \delta \lambda > -\delta \lambda \). Since \( \alpha > 0 \) and \( f' > 0 \) by assumption, it follows that the \( \partial Z / \partial \alpha > 0 \). Thus, for any \( a_h, a_l < a_h, \delta \), and \( \lambda \) there exists a threshold value \( q, q > 0 \), such that \( k_l - k_h < q \) implies \( C_1 > C_2 \).
PROOF OF PROPOSITION 2′: The per-capita cost of implementing action $a$ is $v + 2f((1 - \lambda)(a - P)) + k((1 - \lambda)(a - P))$. The result follows immediately from differentiation of this expression because this cost is decreasing in $\lambda$ if $a > P$ and increasing in $\lambda$ if $a < P$.

PROOF OF OBSERVATION 1′:

We first show that it is optimal to assign all $\lambda_S$ agents to organization $h$ and all $\lambda_P$ agents to organization $l$. Let $\delta$ denote the proportion of $\lambda_S$ agents assigned to $h$ and the proportion of $\lambda_P$ agents assigned to $l$. Taking $\delta$ as given and assuming that contracts for each type are structured to implement $a_h$ by agents in organization $h$ and $a_l$ by agents in organization $l$, the post-contracting equilibrium conditions imply that the per capita of implementing these actions total

$$v + f(U_h) + f(U_l) + \frac{1}{2}k_h(U_h - u) + \frac{1}{2}k_l(U_l - u),$$

where $U_h = (a_h - P)(1 - \delta\lambda_S - (1 - \delta)\lambda_P)$ and $U_l = (a_l - P)(1 - (1 - \delta)\lambda_S - \delta\lambda_P)$. Differentiating this cost with respect to $\delta$ yields

$$-(\lambda_S - \lambda_P)((a_h - P)f'(U_h) - (a_l - P)f'(U_l) + (a_h - a_l)) < 0,$$

because $a_l < P < a_h$. Therefore, the optimal proportion of $\lambda_S$ agents allocated to $h$ is 1 and the optimal proportion of $\lambda_P$ agents allocated to organization $l$ is 1.

Let $\{w_l, b_l\}$ denote the optimal contract for organization $h$ when type is observable and $\lambda_S$ agents are assigned to organization $h$. Let $\{w_l, b_l\}$ denote the optimal contract for organization $l$ when type is observable and $\lambda_P$ are assigned to organization $l$. These contracts and type assignments are optimal when type is observable and result in each agent attaining $v$. The proof is completed by showing that, when type is not observable, $\{w_h, b_h\}$ and $\{w_l, b_l\}$ support an equilibrium in which $\lambda_S$ agents choose organization $h$ and action $a_h$, while $\lambda_P$ agents choose organization $l$ and action $a_l$.

Assume organization $h$ offers $\{w_h, b_h\}$ and organization $l$ offers $\{w_l, b_l\}$. Conjecture an equilibrium in which $\lambda_S$ agents choose organization $h$ and $\lambda_P$ agents choose organization $l$. If no agents defects from the conjectured equilibrium assignment, it must be that
\( \lambda_S \) agents choose \( a_h \), \( \lambda_P \) agents choose \( a_l \), and both agent types attain expected utility \( v \). The proof is completed by showing that no agent would strictly prefer to switch organizations.

Define \( v_{it} \) to be the expected utility attained by an agent of type \( t \in \{ S, P \} \) (where the subscript \( S \) denotes \( \lambda_S \) agents and the subscript \( P \) denotes \( \lambda_P \) agents) in an organization of type \( i \in \{ h, l \} \). No agent prefers to switch organizations (i.e., defect from the conjectured equilibrium assignment) if and only if

\[
(B8) \quad v_{hp} \leq v = v_{lp} \quad \text{and} \\
(B9) \quad v_{lc} \leq v = v_{hc}.
\]

Given that \( a_l < P < a_h \), it follows that \( N_{lc} < N_{lp} \) and \( N_{hp} < N_{hc} \), where \( N_{it} \) is the norm faced by an agent of type \( t \) in organization \( i \). Thus, it follows that

\[
(B10) \quad v_{lc} < v = v_{lp} \quad \text{and} \\
(B11) \quad v_{hp} < v = v_{hc}.
\]

Equations (B10) and (B11) imply that (B8) and (B9) are satisfied.