Why Does The Introduction of Monetary Compensation Produce A Reduction In Performance?

Harvey S. James, Jr.*
Contracting and Organizations Research Institute
University of Missouri
146 Mumford Hall
Columbia, MO 65211 USA
Phone: 573-884-9682
Fax: 573-882-3958
E-mail: hjames@missouri.edu

March 7, 2003

Abstract: According to empirical evidence, extrinsic incentives often crowd out intrinsic motivation, thus reducing the effort choices of workers. This article presents a simple model illustrating how the introduction of monetary incentives causes a discontinuous reduction in worker effort as well as a reduction in worker motivation to act in the interest of a principal. The primary finding is that motivation crowding out occurs when then the object of an agent's intrinsic motivation is a principal who is also the source of the extrinsic compensation the agent receives. When intrinsic satisfaction is directed at more generalized social norms of behavior, however, extrinsic rewards will not crowd out intrinsic motivation.

Keywords: Principal-agent problem, incentive compensation, intrinsic motivation, motivation crowding out

JEL Codes: D64, J33, L22

* I am grateful for helpful comments from Mike Panik, Farhad Rassekh, and Michael Sykuta. An earlier version of this paper was presented at a workshop at the University of Missouri. This research was supported in part by the Missouri Agricultural Experiment Station.

Why Does The Introduction of Monetary Compensation Produce A Reduction In Performance?

I. Introduction

Kreps (1997) affirmed that it is a stylized fact that "providing extrinsic incentives for workers can be counterproductive, because it may destroy the worker's intrinsic motivation, leading to lessened levels of quality-weighted effort and lower net profits for the employer" (p. 360). However, there is also considerable empirical validity that extrinsic incentives crowd out intrinsic motivation (Frey and Jegen, 2001). For example, Gneezy and Rustichini (2000) conducted experiments in which subjects were offered "employment" contracts that required them to answer simple questions or to perform simple tasks (e.g., collect donations). Some subjects were offered fixed wages for participating and then told to complete as many tasks as possible. Other subjects were offered a fee for participating but then an additional "incentive" payment based on their productivity. Gneezy and Rustichini found that higher incentive rates induced greater effort, but the effort level of workers given only a fixed fee often exceeded the effort level of workers paid an incentive wage. Fehr and Gächter (2000a) reported similar results from an experiment in which a principal offered a contract to an agent to provide effort in exchange for a payment. If the agent accepted the contract, then the agent chose the desired level of effort. In some contracts, the principals offered only fixed wages. In other contracts, the principals offered a fixed wage but also had the opportunity to punish (e.g., fine) shirking agents. Fehr and Gächter found that the (negative) incentive contracts, on average, elicited lower levels of effort from agents relative to the fixed wage contracts.

In spite of the growing empirical evidence that monetary incentives often crowd out intrinsic motivation and thus induce lower levels of effort, economists have been unable to provide convincing economic explanations for motivation crowding out (Kreps, 1997; Frey and Jegen, 2001). For example, Gneezy and Rustichini (2000) considered two possible explanations for their experimental findings. The first explanation was that agents are motivated by extrinsic and intrinsic factors and that the introduction of monetary compensation "crowds out" intrinsic motivation. Although they accepted that some form of "crowding out" seemed to occur, Gneezy and Rustichini rejected this explanation because it does not explain the discontinuous reduction in performance they observed when incentives are introduced. The second explanation was that performance contracts are incomplete and the introduction of incentives causes agents to change their perception of the contracts. Specifically, without explicit incentives, agents perceive the contracts as embodying an implicit obligation on their part to provide some "basic" level of effort, which obligation is based on social norms of cooperation or reciprocity (see Fehr and Gächter, 2000b). Although such social norms may exist (see Ostrom, 2000) in support of intrinsic incentives that induce agents to supply effort when contracts are incomplete and explicit incentives are absent, this does not explain *why* the introduction of incentives changes the perceived obligation of workers to supply effort.

This paper presents a model illustrating *how* the combination of extrinsic and intrinsic motivations explains the discontinuous shift in worker effort when incentives are introduced. This paper also answers the question of *why* the introduction of incentives produces a reduction in performance by crowding out intrinsic motivation. To say that extrinsic incentives crowd out intrinsic motivation suggests that when extrinsic incentives are introduced into a utility function comprising both extrinsic as well as intrinsic components, the agent behaves *as if* the intrinsic

features of his utility are no longer salient. What conditions must exist to make an agent behave as if the intrinsic features of his utility are no longer important when extrinsic incentives are introduced? This paper answers this important question. Unlike previous models that illustrate motivational crowding out by making assumptions about the signs of partial derivatives (as in Frey, 1993 and 1994; Frey and Oberholzer-Gee, 1997), this paper shows that the crowding out of intrinsic motivation is an optimal response of the agent to the introduction of incentives *when intrinsic motivation is directed at the interests of the principal* (e.g., when the agent is motivated by loyalty to the principal) because of the offsetting effects agent and principal interests have in an agent's utility comprising both extrinsic and intrinsic features. Thus, when an agent derives intrinsic satisfaction from working in the interest of a principal, and this principal also controls the agent's compensation, then extrinsic incentives can be shown to crowd out intrinsic motivation under certain conditions. The reason is that the combination of fixed and incentive compensation creates a latent "conflict of interest" between the agent's extrinsic interests and his intrinsic interests if the latter are tied to the well-being of the principal. However, when intrinsic motivation is tied to more generalized norms of behavior (e.g., to provide an honest day's work for an honest day's pay) rather than to principal interests, then compensation does not crowd out intrinsic motivation. In this sense this paper supplements Murdock's (2002) article that shows how incentive contracts in a principal-agent framework optimally respond to the presence of agent intrinsic motivation. In his paper, intrinsic motivation complements an implicit contract between a firm and worker, in that the firm agrees to implement projects generating positive intrinsic utility for the agent but negative expected returns for the firm. The firm gains in the long run, however, because the intrinsic motivation of the worker increases the effect of effort incentives attached to other projects that are profitable, thus increasing the overall returns of the

firm. In contrast, this paper shows that under certain circumstances the use of extrinsic effort incentives could backfire by reducing the effort choices of agents engaged in activities generating positive profits for the firm.

II. Model

Suppose a principal hires an agent to supply effort, *e*, in exchange for compensation, *w*. If *p* is the revenue generated per unit of effort, then the principal's profits are $pe - w$. If the disutility of effort for the worker is e^2, then the agent's extrinsic utility is $w - e^2$. Optimally, the worker should provide effort $e^* = p/2$, which maximizes social welfare (profits plus utility).

A. Explicit Incentives and the Discontinuous Reduction of Effort

How do we account for evidence that the introduction of incentives reduces the effort choice of the agent? To answer this question, suppose the agent chooses effort, *e*, to maximize the following utility function comprising both extrinsic and intrinsic elements:

$$U = \overline{w} + re - e^2 + \delta(pe - \overline{w} - re). \quad (1)$$

The extrinsic motivation of the agent is represented as $\overline{w} + re - e^2$, which consists of compensation from a fixed fee, $\overline{w}$, and an incentive payment at rate *r*, net of effort cost e^2. To this is added a term depicting intrinsic motivation, represented as $\delta(pe - \overline{w} - re)$. The parameter δ represents the degree to which the agent is intrinsically motivated to provide effort generating profits of $pe - \overline{w} - re$ for the principal (see Murdock, 2002). The parameter δ could also be interpreted as the degree to which the agent feels an innate sense of duty (such as loyalty) toward the principal. Consistent with Kant (1949), who likened the honoring of one's duty to an "inner

satisfaction" and a "genuine incentive," the ideas of honoring a perceived duty and being intrinsically motivated may be used synonymously. Thus, the stronger is the agent's innate sense of duty toward the principal, the greater is the intrinsic motivation of the agent to provide the required effort. Assume $p > r$ and $\delta \geq 0$ (so that the agent is motivated to be productive rather than destructive). In this model, to say that intrinsic motivation is crowded out means that the agent behaves *as if* $\delta = 0$. An agent who is intrinsically motivated, on the other hand, behaves as if $\delta > 0$.

For simplicity, assume a two-period framework. In period one a principal offers an employment contract to an agent randomly drawn from a population consisting of workers whose reservation wage is zero and who are identical in their abilities, but are different in the size of δ defining the strength of the intrinsic motivation they feel, so that agents with higher δs are more intrinsically motivated than agents with lower δs, other things being equal. Assume further that the level of δ is private information and not knowable to the principal at any cost. Therefore, we are abstracting away from the principal's problem of identifying agents with high δs. The contract offered to the agent consists of a fixed wage, $\overline{w}$, and an incentive rate, $r \geq 0$, in exchange for effort, *e*, which the agent selects. If the agent agrees to the contract, then in period two the agent makes the effort choice and the principal compensates the worker, which occur simultaneously. In choosing effort, the agent can choose whether or not to be intrinsically motivated but does not choose the actual value of δ. Again, an intrinsically motivated agent is one who chooses effort as if $\delta > 0$ in his utility, while an agent whose intrinsic motivation has been crowded out chooses effort as if $\delta = 0$.

Given equation (1) and the assumptions of the model, the agent's optimal effort choice, $\hat{e}$, obtained by maximizing (1) with respect to effort, is

$$\hat{e} = \frac{\delta(p - r) + r}{2}. \tag{2a}$$

Now consider two possible contractual scenarios. In the first, the agent is offered only a fixed fee "for showing up," but he also feels a sense of duty to supply effort. In this case $r = 0$ and $\delta > 0$ so that the agent supplies effort $\hat{e} = \delta p / 2 \equiv \hat{e}_1$. In the second, the agent is offered a fixed fee and an incentive, *r*, for effort, but he feels no obligation to provide effort above what the incentive offers. In this case $r > 0$ and $\delta = 0$ so that the agent supplies effort $\hat{e} = r / 2 \equiv \hat{e}_2$. Note that in this second case, effort increases as the piece rate increases (the optimal effort-incentive profile is positive). These two contractual scenarios conform to the stylized fact that workers are often intrinsically motivated when no incentive compensation is offered (i.e., $r = 0$ and $\delta > 0$) but are not intrinsically motivated when incentives are offered (i.e., $r > 0$ and $\delta = 0$). Equation (2b) summarizes the effort choice of the agent under each of these two scenarios.

$$\hat{e} = \begin{cases} \delta p / 2 \equiv \hat{e}_1 & \text{if} \quad r = 0, \delta > 0 \\ r / 2 \equiv \hat{e}_2 & \text{if} \quad \text{r} > 0, \delta = 0 \end{cases}. \tag{2b}$$

According to the experimental evidence, effort supplied under a fixed-fee and positive "duty" contract will often equal or exceed effort supplied under the incentive contract (i.e., $\hat{e}_1 \geq \hat{e}_2$). This would be true when $\delta \geq r / p$, or when the intrinsic motivation or sense of duty of the agent is sufficiently "strong." In other words, when $\delta \geq r / p$, then effort provided by the agent when no extrinsic incentives are offered ($\hat{e}_1$) will not be less than effort provided by the agent when incentives are offered and the agent's sense of intrinsic motivation is zero ($\hat{e}_2$). If incentives are subsequently introduced, however, then the effort choice of the agent discontinuously declines from $\hat{e}_1$ to $\hat{e}_2$, if the introduction of incentives also results in the agent

behaving as if $\delta = 0$. Thereafter, any increase in piece rates results in increased effort (i.e., as *r* increases, $\hat{e}_2$ increases continuously). This is illustrated in Figure 1.

B. Motivation Crowding Out When Duty is Directed Toward a Principal

Why is the introduction of a piece rate associated with an apparent disappearance of a sense of duty by the worker? That is, why does $r > 0$ cause the agent to behave as if $\delta = 0$? This question is especially germane when one recognizes that the agent's utility is increasing in δ. Thus, wouldn't the agent prefer that $\delta > 0$ when $r > 0$? Gneezy and Rustichini speculated that the change from a fixed-fee to an incentive contract is associated with a change in the agent's perception of the contract because the contract is incomplete. The agent's "perception," in effect, fills in the missing parts (e.g., how much effort to supply). When incentives are not offered, the agent's belief that "effort should be given when one is paid" – which might be derived from norms of reciprocity or other sources – may help the agent determine how much effort to supply; hence, the agent supplies effort $\hat{e}_1$. In the case of an incentive contract, however, the perception might be that effort should be rewarded according to the piece rate only. In this case, social norms for effort are not required because the level of the piece rate helps the agent determine how much effort to supply; hence, the agent supplies effort $\hat{e}_2$. Although perceptions may help explain how the agent chooses effort in the absence of incentives, it is not obvious why social norms of reciprocity and the associated feelings of obligation should suddenly disappear when perceptions shift following the introduction of incentives. Why doesn't the worker select effort according to the incentive rate as well as the social norm for work, since fulfilling a duty is intrinsically rewarding to the agent (i.e., it adds to the agent's utility)? Why does the incentive "crowd out" the agent's innate sense of duty rather than reinforce it?

In order to understand why compensation might diminish the agent's intrinsic motivation to adhere to a perceived duty – that is, why an agent with utility comprising extrinsic and intrinsic elements would choose effort as if intrinsic utility no longer mattered just because incentives are offered – we must examine the utility of the agent when $\delta = 0$ (no intrinsic motivation) and $\delta > 0$ (positive intrinsic motivation) under the two cases of $r = 0$ (no incentive compensation) and $r > 0$ (positive incentive compensation). This is done by evaluating the utility of the agent from equation (1) when effort is selected according to the agent's optimal effort choice from equation (2a) under each of these four scenarios: $r = 0$ and $\delta = 0$; $r = 0$ and $\delta > 0$; $r > 0$ and $\delta = 0$; and $r > 0$ and $\delta > 0$.

First, consider the case in which no incentive pay is offered so that $r = 0$. If the agent ignores his duty, so that $\delta = 0$, then the utility of the agent is

$$U(\hat{e} \mid r = 0, \delta = 0) = \overline{w}, \tag{3a}$$

which corresponds to an effort choice of $\hat{e} = 0$. If, however, the agent derives positive satisfaction by fulfilling a perceived duty to act in the principal's interest, so that $\delta > 0$, then the utility of the agent is

$$U(\hat{e} \mid r = 0, \delta > 0) = \frac{\delta^2 p^2 - 4\delta\overline{w} + 4\overline{w}}{4}, \tag{3b}$$

which corresponds to an effort choice by the agent of $\hat{e}_1$ $(= \delta p / 2)$. The agent will behave in a manner reflecting intrinsic motivation when equation (3b) is not less than (3a), or when

$$\delta \geq \frac{4\overline{w}}{p^2} \equiv \delta_{FW}. \tag{4}$$

Equation (4) defines the threshold determining whether an agent with a given innate sense of duty, δ, will be intrinsically motivated to provide effort for a principal when offered only a fixed wage (FW) contract. When $\delta \geq \delta_{FW}$, or when the degree to which the agent feels an innate sense

of duty toward the principal is relatively high, then the agent will behave as if intrinsically motivated when choosing the effort he will supply for the principal. However, when $\delta < \delta_{FW}$, or when the agent's innate sense of duty toward the principal is relatively low, then he will have higher utility by behaving as if he were ignoring his duty in that $\delta = 0$; that is, intrinsic motivation would be crowded out. Observe that the threshold δ_{FW} increases in the amount of the fixed wage but decreases in the revenue generated per unit of effort. That is, the larger the fixed wage offered to the agent, the higher the agent's innate sense of duty would have to be in order to induce him to consider his obligation to provide effort, other things being equal. Simply, if the fixed wage offered to the agent is too large, the agent may decide to take the money and run (i.e., provide no effort). According to the experimental evidence, this does not seem to be the case, however. Most individuals, when offered a fixed wage but no incentive compensation, seem to be willing to work as if intrinsically motivated (i.e., $\delta > 0$), suggesting that $\delta \geq \delta_{FW}$ and thus effort corresponds to $\hat{e}_1$.

Now, consider the case in which incentive pay is offered so that $r > 0$. If the agent is not intrinsically motivated, or if the agent behaves as if ignoring his duty to the principal, so that $\delta = 0$, then the utility of the agent is

$$U(\hat{e} \mid r > 0, \delta = 0) = \frac{r^2 + 4\overline{w}}{4}, \tag{5a}$$

which corresponds to an effort choice of $\hat{e}_2$ $(= r/2)$. If, however, the agent is intrinsically motivated, so that $\delta > 0$, then the utility of the agent is

$$U(\hat{e} \mid r > 0, \delta > 0) = \frac{\delta^2 (p - r)^2 + 2\delta(pr - r^2 - 2\overline{w}) + r^2 + 4\overline{w}}{4}, \tag{5b}$$

which corresponds to effort choice $\hat{e} = \frac{\delta(p-r)+r}{2}$. The agent will behave as if intrinsically motivated when equation (5b) is not less than (5a), or when

$$\delta \geq \frac{4\overline{w} - 2r(p-r)}{(p-r)^2} \equiv \delta_{IC}. \tag{6}$$

Equation (6) defines the threshold determining whether an agent will be intrinsically motivated to provide effort for a principal when offered an incentive compensation (IC) contract. If $\delta \geq \delta_{IC}$, or if the agent's innate sense of duty is sufficiently strong, then even when incentive compensation is offered the agent will have higher utility if he were not to forgo his duty when choosing the level of effort to provide for the principal. If, however, the agent's innate sense of duty is too low such that $\delta < \delta_{IC}$, then the agent will have higher utility by acting as if he were ignoring that obligation and thus by choosing effort based solely on the size of the incentive rate, *r*; that is, intrinsic motivation would be crowded out. Like the case in which no incentive compensation is offered, the threshold δ_{IC} increases in the amount of the fixed wage. How the threshold changes as *p* and *r* increase, however, is indeterminate.

Figure 2 summarizes the effort choices of the agent as described in the preceding discussion. The experimental evidence suggests that agents seem to behave as if intrinsically motivated when a fixed wage but no incentive rate is offered, suggesting that $\delta \geq \delta_{FW}$. This corresponds to effort choice $\hat{e}_1$ in cell III of Table 2. Therefore, assume that the $\delta \geq \delta_{FW}$ condition is satisfied so that the agent is providing effort $\hat{e}_1$. If an incentive is now offered, would $\delta \geq \delta_{IC}$ so that effort in cell IV of Figure 2 is selected, or would $\delta < \delta_{IC}$ so that effort $\hat{e}_2$ in cell II is selected? The experiment evidence suggests that when workers are offered incentive compensation, many who would have been or who were intrinsically motivated when no

incentives were paid now behave as if they have dropped consideration of duty to the principal when choosing their effort level, suggesting that $\delta < \delta_{IC}$ rather than $\delta \geq \delta_{IC}$. Why? That is, why would the introduction of incentive compensation imply that for some worker cell II is chosen over cell IV? Why would the introduction of an incentive imply that $\delta < \delta_{IC}$ rather than $\delta \geq \delta_{IC}$? To answer these questions we must determine whether and, if so, when, it is possible for $\delta_{FW} \leq \delta < \delta_{IC}$. This expression would be true when $\frac{4\overline{w}}{p^2} < \frac{4\overline{w} - 2r(p-r)}{(p-r)^2}$, or when

$$\overline{w} > \frac{p^2(p-r)}{2(2p-r)}. \tag{7}$$

Equation (7) says that if the fixed compensation offered by the principal is "too high," then some agents intrinsically motivated to provide effort in the absence of incentive compensation (i.e., with $\delta \geq \delta_{FW}$) will no longer be intrinsically motivated when incentive compensation is offered (i.e., $\delta < \delta_{IC}$). That is, their intrinsic motivation would appear to be crowded out by the extrinsic incentives. If, however, $\overline{w} \leq \frac{p^2(p-r)}{2(2p-r)}$, or if the fixed compensation is not too large, then $\delta_{FW} \geq \delta_{IC}$. Therefore, agents who are already intrinsically motivated when no incentive compensation is offered (i.e., $\delta \geq \delta_{FW}$) will also behave as if intrinsically motivated when incentives are offered, since $\delta \geq \delta_{FW}$ implies $\delta > \delta_{IC}$ in this case. In other words, for some agents the question of why the introduction of extrinsic compensation "crowds out" intrinsic motivation or duty, thus resulting in lower effort levels, appears to depend on the size of the fixed portion of the compensation offered to the agent.

The reason the fixed fee plays such an important role in determining whether incentive compensation drives out intrinsic motivation is because both forms of compensation (fixed and

incentive pay) have opposing effects on the extrinsic and intrinsic portions of the agent's utility. Unlike effort, which increases the agent's extrinsic utility $\overline{w} + re - e^2$ (but at a decreasing rate) as well as the agent's intrinsic utility $\delta(pe - \overline{w} - re)$ (since $p > r$ by assumption), increases in both the fixed wage, $\overline{w}$, and the incentive rate, *r*, increase extrinsic utility but decrease intrinsic utility. This is because the agent recognizes that higher compensation reduces the overall level of profits that accrue to the principal, *other things being equal*, and it is the overall level of profits that the agent – by assumption – feels a duty to support. When the fixed compensation is too high and incentive rates are offered, the agent must balance the diminished importance of satisfaction derived from fulfilling a duty to the principal with the increased rewards extrinsic compensation provides (e.g., "if the principal's profit is not going to be very large, what's the point of my worrying about it?"). In other words, a fixed compensation that is too large combined with the presence of an incentive rate literally forces the agent to choose between his extrinsic interests and his intrinsic interests he feels for the principal. Only when the innate sense of satisfaction that the agent derives from acting in the principal's interest is large enough (i.e., $\delta \geq \delta_{IC}$ and $\delta \geq \delta_{FW}$), which we would expect for some workers, will the intrinsic satisfaction from acting in the principal's interest be strong enough to outweigh the temptation to ignore his duty to the principal.

C. Intrinsic Motivation When Duty is Directed Toward a Social Norm

In the preceding model it was assumed that the agent derived satisfaction from acting in the interest of the principal, where the principal was assumed to have an interest in maximizing profits of $pe - \overline{w} - re$. Now suppose that rather than feeling a duty toward the principal, the agent felt an obligation to adhere to a generalized social norm, such as "provide an honest day's

work for an honest day's pay" or "honor your contractual obligations." This could be interpreted to mean that the agent believes he should provide at least some minimal level of effort, $\bar{e}$, when employed by the principal. Thus, instead of an objective function defined by equation (1), the agent's objective function would be represented as

$$U = \overline{w} + re - e^2 + \delta(e - \bar{e}). \tag{8}$$

In this case the object of the agent's intrinsic motivation is not the profits of the principal *per se*, although providing higher levels of effort could increase the principal's profits. Rather, the object of the agent's duty is providing effort at least equal to some minimum, $\bar{e}$, so that the intrinsic utility of the agent is $\delta(e - \bar{e})$. In this formulation the utility of the agent is increased when $e > \bar{e}$ and decreased when $e < \bar{e}$, thus giving the agent an incentive to adhere to the general social norm, other things being equal. How does the agent know what the minimum level of effort should be? This minimum effort could be specified in an employment contract, in which case the social norm to which the agent would have an interest in following might be interpreted as "honor your contractual obligations." But this need not be the case. The minimum effort could be defined by a social norm to "provide an honest day's work for an honest day's pay." In either case, observe that the agent's intrinsic motivation is not focused directly on the principal's interests but rather on a more generalized norm of behavior.

Given equation (8), the agent's optimal effort choice is

$$\tilde{e} = \frac{\delta + r}{2}. \tag{9}$$

Consistent with the findings above, effort under a fixed-wage contract with positive intrinsic motivation will always be greater than effort under an incentive contract when intrinsic motivation is absent when the innate sense of duty felt by the agent is large enough. Specifically,

if $\tilde{e}_1 = \delta/2$ is the optimal level of effort when $r = 0$ and $\delta > 0$, and $\tilde{e}_2 = r/2$ is the optimal effort level when $r > 0$ and $\delta = 0$, then $\tilde{e}_1 \geq \tilde{e}_2$ when $\delta \geq r$.

Consider now the utility of the agent under the cases of $\delta = 0$ or $\delta > 0$ when $r = 0$, and $\delta = 0$ or $\delta > 0$ when $r > 0$. First, suppose no incentive compensation is offered, so that $r = 0$. If the agent behaves as if feeling no sense of duty to follow the generalized social norm, then his utility is

$$U(\tilde{e} \mid r = 0, \delta = 0) = \overline{w}, \tag{10a}$$

which corresponds to an effort choice of $\tilde{e} = 0$. If, however, the agent is intrinsically motivated, then the agent's utility is

$$U(\tilde{e} \mid r = 0, \delta > 0) = \frac{\delta^2 - 4\delta\overline{e} + 4\overline{w}}{4}, \tag{10b}$$

which corresponds to effort choice $\tilde{e}_1$. The agent will be intrinsically motivated when equation (10b) is not less than equation (10a), or when

$$\delta \geq 4\overline{e} \equiv \tilde{\delta}_{FW}. \tag{11}$$

Suppose now the principal offers an incentive wage to the agent, so that $r > 0$. In this case if the agent behaves as if not intrinsically motivated, then his utility is

$$U(\tilde{e} \mid r > 0, \delta = 0) = \frac{r^2 + 4\overline{w}}{4}, \tag{12a}$$

which corresponds to effort choice $\tilde{e}_2$. If, on the other hand, the agent is intrinsically motivated when incentive compensation is provided, then his utility is

$$U(\tilde{e} \mid r > 0, \delta > 0) = \frac{\delta^2 - 4\delta\overline{e} + 2\delta r + r^2 + 4\overline{w}}{4}, \tag{12b}$$

which corresponds to effort $\tilde{e} = \frac{\delta + r}{2}$. Equation (12b) would not be less than equation (12a), or the agent will be intrinsically motivated to adhere to the social norm of providing a minimal level of effort, when

$$\delta \geq 4\bar{e} - 2r \equiv \tilde{\tilde{\delta}}_{IC}. \tag{13}$$

Because the incentive rate is assumed to be positive (when it is offered), the threshold degree of duty necessary for the agent to behave as if intrinsically motivated when incentive are offered will always be lower than the threshold level of duty when only fixed compensation is offered; that is, $\tilde{\tilde{\delta}}_{FW} \geq \tilde{\tilde{\delta}}_{IC}$ when $r > 0$. Therefore, if the agent is intrinsically motivated to work when there are no extrinsic incentives to providing higher levels of effort (i.e., $\delta \geq \tilde{\tilde{\delta}}_{FW}$), then the agent will always behave as if intrinsically motivated when incentive compensation is offered (i.e., $\delta \geq \tilde{\tilde{\delta}}_{IC}$). Hence, incentive compensation will not crowd out intrinsic motivation when the object of the agent's intrinsic motivation is a general social norm or contractual provision, rather than the interests of the principal. The reason is that because the agent's intrinsic motivation is not grounded in the direct interests of the principal, there is no potential conflict of interest between the agent's extrinsic interests and his intrinsic motivation to act in accordance with a generalized social norm. Consequently, there is no reason for extrinsic incentives to crowd out intrinsic motivation.

III. Conclusion

For some agents with a willingness to work in the absence of explicit incentives, the introduction of monetary incentives causes a discontinuous decline in their effort choices. This

shift in effort occurs if the agent's innate sense of duty is directed toward a principal who also controls the extrinsic compensation of the agent. The erosion of duty in this case reflects a tradeoff the agent faces between intrinsic satisfaction derived from doing one's duty, even when the effort required is personally costly, and extrinsic satisfaction derived from being compensated. When the object of the agent's intrinsic motivation is on more generalized norms of behavior, explicit incentives would not crowd out the intrinsic motivation of agents.

This analysis suggests that a consideration of *what* an agent is intrinsically motivated to do is just as important as the choice of optimal incentive contracts within a principal-agent framework. For example, indoctrinating workers on the importance of acting in the firm owners' interests when incentives are also used to help align the interests of owner and worker may be counterproductive. The reason is that workers may "see through" attempts by firm owners to them to maximize profits as enriching the owners at the expense of the firm. According to Miller (2001), "Holmstrom's impossibility result has supplied an alternative vision of opportunism – opportunism on the part of owners themselves. It is an opportunism that follows from their ownership of residual profits, and it paradoxically puts them at odds with overall efficiency of the firm" (p. 329). The idea is that under certain circumstances firm owners could increase their residual returns by fostering inefficiencies within the organization. If workers expect this, they may be less inclined to "trust" firm owners who claim that certain activities or programs are for the good of the firm, resulting in lower returns for the firm owners. This might explain why Fehr and Gächter (2000a) found in their experiments that principals often preferred the less effective incentive contracts because they allowed the principals to appropriate a larger share of the surpluses produced from the principal-agent relationship, even though the surpluses were, in total, smaller than those produced from fixed-wage contracts. Principals who stress the

importance of a healthy work ethics rather than inherent owner interests, on the other hand, may see such efforts as complementing their compensation programs for the good of the firm, consistent with Murdock's (2002) analysis that intrinsic motivation and incentive contracts are complementary. The reason is that workers may perceive these actions as credible efforts by firm owners to improve organization efficiency, thus encouraging agents to be intrinsically motivated for the good of the firm.

References

Fehr, Ernst, and Simon Gächter, (2000a), "Do Incentive Contracts Crowd out Voluntary Cooperation?" working paper no. 34, Institute for Empirical Research in Economics, University of Zurich.

Fehr, Ernst, and Simon Gächter, (2000b), "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives*, 14(3), pp. 159-181.

Frey, Bruno S., (1993), "Does Monitoring Increase Work Effort? The Rivalry with Trust and Loyalty," *Economic Inquiry*, 31(October), pp. 663-670.

Frey, Bruno S., (1994), "How Intrinsic Motivation is Crowded Out and In," *Rationality and Society*, 6, pp. 334-352.

Frey, Bruno S., and Reto Jegen, (2001), "Motivation Crowding Theory," *Journal of Economic Surveys*, 15(5), pp. 589-611.

Frey, Bruno S., and Felix Oberholzer-Gee, (1997), "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out," *American Economic Review*, 87, pp. 746-755.

Kant, Immanuel, (1949), "The Noble Descent of Duty," *The Critique of Practical Reason and Other Writings in Moral Philosophy*, trans. L.W.Beck, Chicago, IL: University of Chicago Press, pp. 193-195. First published in 1788. Reproduced in *Ethics*, P. Singer (ed.), New York, NY: Oxford University Press, 1994, pp. 39-41.

Kreps, David M., (1997), "Intrinsic Motivation and Extrinsic Incentives," *American Economic Review*, Papers and Proceedings, 87(2), pp. 359-364.

Gneezy, Uri, and Aldo Rustichini, (2000), "Pay Enough or Don't Pay At All," *Quarterly Journal of Economics*, 115(3), pp. 791-810.

Miller, Gary, (2001) "Why Is Trust Necessary in Organizations? The Moral Hazard of Profit Maximization," in K.S. Cook (ed), *Trust in Society*, New York, NY: Russell Sage Foundation, pp. 307-331

Murdock, Kevin, (2002) "Intrinsic Motivation and Optimal Incentive Contracts," *Rand Journal of Economics*, 33(4), pp. 650-671.

Ostrom, Elinor, (2000) "Collective Action and the Evolution of Social Norms," *Journal of Economic Perspectives*, 14(3), pp. 137-158.

Figures

Figure 1. Agent effort choices under two contracting scenarios. When incentives are zero ($r = 0$) and intrinsic motivation is positive ($\delta > 0$), effort is $\hat{e}_1$. When incentives are offered ($r > 0$) and intrinsic motivation disappears ($\delta = 0$), effort is $\hat{e}_2$, which increases in the incentive rate, *r*. Observe that $e_1 \geq e_2$ when $\delta \geq r / p$.

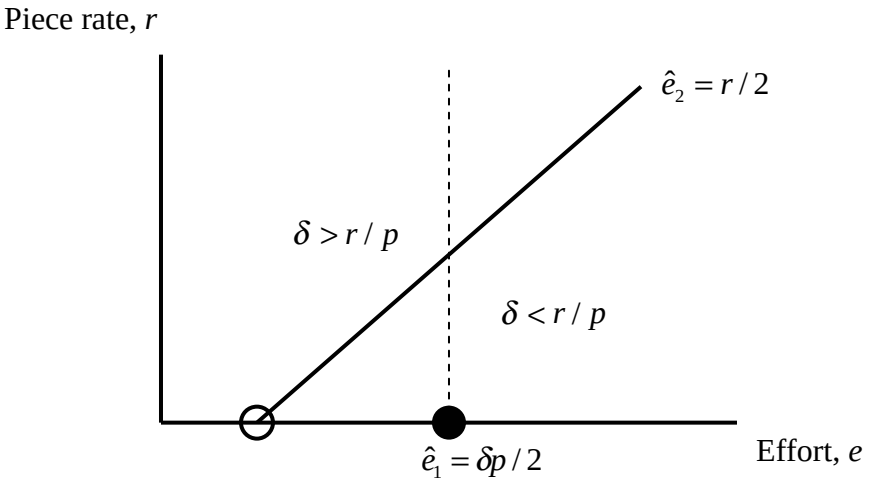

Figure 2. Optimal effort choices for an agent under each of these four scenarios: $r = 0$ and $\delta = 0$; $r = 0$ and $\delta > 0$; $r > 0$ and $\delta = 0$; and $r > 0$ and $\delta > 0$. When incentives are zero, the agent will choose cell II over cell I when $\delta \geq \delta_{FW}$ (see equation (4)). When incentives are positive, the agent will choose cell IV over cell III when $\delta \geq \delta_{IC}$ (see equation (6)).

Duty, δ	Incentive Rate, r: $r = 0$	Incentive Rate, r: $r > 0$
$\delta = 0$	I: 0	II: $\hat{e}_2 \equiv r/2$
$\delta > 0$	III: $\hat{e}_1 \equiv \delta p/2$	IV: $[\delta(p-r)+r]/2$