




ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE  DIRECT®

Journal of Economic Psychology 26 (2005) 549–566

JOURNAL OF  
**Economic  
Psychology**

[www.elsevier.com/locate/joep](http://www.elsevier.com/locate/joep)

# Why did you do that? An economic examination of the effect of extrinsic compensation on intrinsic motivation and performance

Harvey S. James Jr. \*

*Contracting and Organizations Research Institute, University of Missouri,  
146 Mumford Hall, Columbia, MO 65211, USA*

Received 27 September 2003; received in revised form 7 September 2004; accepted 22 November 2004

Available online 7 January 2005

---

## Abstract

According to empirical evidence, extrinsic incentives often crowd out intrinsic motivation, thus reducing the effort of workers. This article presents a principal–agent model utilizing a Benthamite interpretation of utility as overall satisfaction and incorporating insights from cognitive evaluation theory to explain motivation crowding out. In this context, motivation crowding out occurs when explicit rewards are perceived as controlling, which results in individuals having greater satisfaction by not being intrinsically motivated. The model shows that rewards could be perceived as controlling under two conditions. The first is when the object of an agent's intrinsic motivation is also the source of the agent's extrinsic compensation. The second is when the incentives offered to the agent are too large.

© 2004 Elsevier B.V. All rights reserved.

*JEL classification:* J30; L20

*PsycINFO classification:* 2360; 3000

*Keywords:* Principal–agent problem; Incentive compensation; Intrinsic motivation; Motivation crowding out

---

---

\* Tel.: +1 573 884 9682; fax: +1 573 882 3958.

E-mail address: [hjames@missouri.edu](mailto:hjames@missouri.edu)

## 1. Introduction

Although recognized by social psychologists (see deCharmes, 1968; Deci, 1975; Deci & Ryan, 1985), the idea 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” is just a stylized fact within economics (Kreps, 1997, p. 360). Nevertheless, economists are accumulating increasing evidence for this effect, known as motivation crowding out (see Frey & 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 of workers given only a fixed fee often exceeded the effort of workers paid an incentive wage. Fehr and Gächter (2000a) reported similar results from an experiment in which a principal offered a contract to agents to provide effort in exchange for payment. In some contracts, the principal offered only fixed wages. In other contracts the principal offered a fixed wage but retained the right 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 evidence that explicit incentives often crowd out intrinsic motivation and thus induce lower levels of effort, economists have been unable to provide convincing economic explanations for the crowding out effect (Frey & Jegen, 2001; Kreps, 1997). 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 the change from a fixed-fee to an incentive contract is associated with a change in agent perceptions of the contract because the contract is incomplete. An agent’s “perception”, in effect, fills in the missing parts (e.g., how much effort to supply). Without explicit incentives, agents perceive the contracts as embodying an implicit obligation on their part to provide effort, which obligation might be based on social norms of cooperation or reciprocity (see Fehr & Gächter, 2000b). However, when incentives are introduced the perception might be that effort should be rewarded according to the piece rate only, thus negating the need for a perception based on social norms. Although 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* a utility function comprising both extrinsic and intrinsic motivations can be used to explain the discontinuous shift in worker effort when incentives are introduced. The model is also used to explain *why* extrinsic incentives might crowd out intrinsic motivation. In contrast to extrinsic motivation, which is oriented toward obtaining a valued external reward, intrinsic drives “are the prototypes of autonomous or self-determined actions because they are performed out of interest and for the inherent satisfaction they yield. When intrinsically motivated, people *want* to engage in the activity; thus no external or intrapsychic prods, promises, or threats are required” (Ryan, Sheldon, Kasser, & Deci, 1996, p. 10; emphasis in original). In this context, motivation crowding out (MCO) means that an agent behaves *as if* the intrinsic properties of his utility no longer factor in his decision-making processes – that is, as if his intrinsic motivation has been “crowded out”, “neutralized”, or “deactivated”. The reason explored in this paper is that the agent has greater overall satisfaction by not being intrinsically motivated than by being intrinsically motivated when explicit incentives are utilized.

In order to explain MCO, two digressions are necessary. The first draws on a Benthamite interpretation of utility, illustrated for instance by the distinction between *decision utility* and *experienced utility* described by Kahneman, Wakker, and Sarin (1997), in order to show that individuals might have preferences over different types of motivations. This is necessary in order to justify the argument that some individuals have greater satisfaction by not being intrinsically motivated. The second draws on the *cognitive evaluation theory* of Deci and Ryan (1980, 1985), which is necessary to provide an empirical basis for the theoretical conditions hypothesized to cause the MCO effect.

### 1.1. Motivation crowding out and individual satisfaction

There is considerable empirical evidence showing that people often behave as if intrinsically motivated. However, when offered explicit incentives, many people suddenly behave as if not intrinsically motivated, an effect referred to as MCO. This observation raises the following question: Is MCO the result of a conscious choice of the agent not to be intrinsically motivated, or does it reflect subconscious automatic processes beyond the relevant control of the agent?<sup>1</sup>

If motivation crowding out is the result of unintended, automatic processes, then economic theory grounded in assumptions of rational, self-interested behavior will be limited in its ability to explain the MCO effect. This limitation is manifested in current theoretical explanations for motivation crowding out, which provide only ad hoc explanations for MCO. For example, in the models of Frey (1993, 1994) and Frey and Oberholzer-Gee (1997), MCO is explained by the fact that the signs of certain cross-partial derivatives are negative rather than positive (or zero). These models are ad hoc because they do not explain why cross-partials are negative for some

---

<sup>1</sup> For instance, Sorrentino (1996, p. 619) says that either “motivation . . . stems from conscious decision making, and non-conscious motives or cognitions have little relevance”, or “non-conscious memories and unintended thought play major roles in determining much of human behavior.”

people but not for others. In effect, the models presume that something happens to some agents when extrinsic incentives are introduced, thus causing MCO, but what that “something” is and why it occurs have not been made *theoretically* explicit.

An alternative approach is to assume a rational basis for the MCO effect, such that people behave as if not intrinsically motivated when doing so results in greater utility or satisfaction for them. This approach is consistent with a Benthamite view of utility, which conceives of utility as a complex web of satisfactions affected by actual experiences of pleasure, pain, feelings, emotions, and thoughts, in contrast to the standard conception of utility as a function of individual choices. In this sense, experiences, the procedures by which decisions are made, and perhaps even the various motivations driving individual choice, are seen to affect the overall satisfaction of agents, thus implying that agents will have preferences for the experiences, procedures, or motivations they have. For example, Kahneman et al. (1997) distinguish between *decision utility* and what they refer to as the Benthamite notion of *experienced utility*. Whereas decision utility “is the weight of an outcome in a decision”, experienced utility consists of hedonic measures of current as well as past, or remembered utility (pp. 375, 376). Kahneman et al. show not only how experiences matter for agents but also why they are important in explaining observed agent behavior. In related work, Frey and Stutzer (forthcoming) distinguish between standard utility and what they call *procedural utility*. Their idea is that people have preferences not only over outcomes but also “about *how* allocative and redistributive decisions are taken” (p. 3; emphasis in original). For instance, the authors find that “participation rights provide procedural utility in terms of a feeling of self-determination and influence” (p. 1).

This paper presents the idea that MCO can be understood from a model in which agent utility comprises both extrinsic and intrinsic utilities and reflects the Benthamite notion of satisfaction, so that how extrinsic and intrinsic utilities affect overall satisfaction is an important analytical consideration. Moreover, agents are assumed to have preferences over what motivates them. This allows us to examine the conditions in which some agents are worse off by being intrinsically motivated when explicit incentives are introduced, so that they will prefer not to be intrinsically motivated, thus inducing the MCO effect.

What conditions might cause the MCO effect?

### 1.2. Social psychological foundations of motivation crowding out

The social psychology literature, particularly cognitive evaluation theory (CET) developed by Deci and Ryan (1980, 1985), provide a basis for describing what conditions might cause some people to prefer not to be intrinsically motivated, thus causing motivation crowding out. According to CET, the basic factors

underlying intrinsic motivation are the psychological needs for autonomy and competence, so the effect of an event such as a reward depend on how it affects perceived self-determination and perceived competence. Events that allow need

satisfaction tend to increase intrinsic motivation whereas those that thwart need satisfaction tend to decrease intrinsic motivation. CET proposes that rewards can be interpreted by recipients primarily as controllers of their behavior or, alternatively, as indicators of their competence. (Deci, Koestner, & Ryan, 1999a, p. 628)

In other words, if an agent perceives that an event supports his innate need for self-determination and competence, then even if the event provides extrinsic value (i.e., increases his extrinsic utility), his behavior vis-à-vis the event will also provide intrinsic satisfaction (i.e., increase his intrinsic utility). Hence, we could say the agent will have higher overall satisfaction by behaving *as if* extrinsically and intrinsically motivated – that is, it would be rational for him to be both extrinsically and intrinsically motivated. On the other hand, if an agent perceives that an event is controlling rather than the result of self-determined behavior, then it will not provide intrinsic satisfaction. Accordingly, we suppose that the agent will have higher utility by behaving *as if* extrinsically motivated only (i.e., intrinsic utility is not factored in). This characterizes motivation crowding out. Therefore, the key to understanding MCO is through a model in which an agent's utility reflects either (a) extrinsic and intrinsic factors if extrinsic rewards are not perceived as controlling, or (b) extrinsic factors only if extrinsic rewards are perceived as controlling.

Extrinsic rewards might be perceived as controlling under two conditions: The first condition is when the size of the reward is large. Empirical evidence suggests that reward size is negatively related to intrinsic motivation (e.g., Newman & Layton, 1984). The reason is that the processes by which an agent rationalizes his behavior may become so overwhelmed by the salience (e.g., size) of extrinsic rewards that he is rationally compelled to attribute his behavior to the compensation rather than to his intrinsic preferences. This is known as the “overjustification effect” (see Lepper, 1981; Lepper, Greene, & Nisbett, 1973). This idea was recognized by Kreps (1997, p. 362) who said

Turning revealed preference on its head, the idea is that when a person performs some act, he looks for rationales that justify his actions. Specifically, if an employee undertakes some effort without the spur of some extrinsic incentive, he will rationalize his efforts as reflecting his enjoyment of the task. And since he enjoys it, he works harder at it. But if extrinsic incentives are put in place, he will attribute his efforts to those incentives, developing a distaste for the required effort.

Similarly, deCharmes (1968, p. 328) stated

As a first approximation, we propose that whenever a person experiences himself to be the locus of causality for his own behavior . . . , he will consider himself to be intrinsically motivated. Conversely, when a person perceives the locus of causality for his behavior to be external to himself . . . , he will consider himself to be extrinsically motivated.

In other words, large rewards may compel an agent to acknowledge that the “locus of causality” is external rather than internal and is intended as a means of control. This would undermine the agent’s inherent need for autonomy and hence provide no intrinsic satisfaction for engaging in the “controlled” behavior (see [Deci, Koestner, & Ryan, 1999b](#)). The result is that the agent would behave based only on the extrinsic utility created, resulting in MCO.

The second condition is that the object of an agent’s intrinsic motivation is also the source of the rewards. For example, if an agent is initially intrinsically motivated to act in the interest of a principal who also pays the worker’s wages, then the introduction of incentives by the principal could be perceived by the agent as an attempt at manipulation. This perception would undermine the agent’s need for autonomy, thus causing MCO. However, if the agent is motivated by generalized norms (such as “provide an honest day’s work for an honest day’s pay”) which by nature are external to the more narrow interests of the principal, then extrinsic incentives introduced by the principal might not be perceived as controlling but rather as an affirmation of competence, thus supporting rather than crowding out intrinsic motivation. This idea is consistent with empirical evidence. Although studies suggest that expected tangible rewards generally reduce intrinsic motivation (see [Deci et al., 1999a](#)), rewards offered for exceeding a norm or a minimum level of effort not only do not cause MCO but also often have positive effects (see [Cameron, Banko, & Pierce, 2001](#)).

### *1.3. Outline of paper*

This paper utilizes a Benthamite view of utility to show how the size of compensation and the object on which intrinsic motivation is placed interact result in the MCO effect. In doing so, this paper supplements [Murdock’s \(2002\)](#) article showing 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 incentives could backfire by reducing the effort choices of agents interested in generating profits for the firm.

The paper begins with a simple principal–agent model in which an agent’s utility incorporates both extrinsic and intrinsic elements. Initially it is assumed that the object of the agent’s intrinsic motivation is also the source of his compensation in that the agent is intrinsically motivated to generate profits for the principal. The model is used to demonstrate how the introduction of incentives could result in a discontinuous reduction in worker effort if the explicit incentives crowd out intrinsic motivation. The model is also used to show why the size of the incentive compensation is important in understanding why an agent would behave as if motivated by extrinsic

factors only. The model is then revised to reflect an agent's intrinsic motivation that is tied to a generalized norm of behavior rather than to the interests of a principal. This revision illustrates why the object on which intrinsic motivation is tied is important in understanding why MCO occurs. The paper concludes with an appropriate discussion.

## 2. 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 worker's disutility of effort is  $e^2$ , then his extrinsic utility is  $w - e^2$ . Optimally, the worker should provide effort  $e^* = p/2$ , which maximizes social welfare (profits plus utility).

### 2.1. Explicit incentives and the discontinuous reduction of effort

How do we account for evidence that the introduction of incentives results in a discontinuous reduction of worker effort? In order to answer this question, a model of utility is required that contains both extrinsic and intrinsic sources of utility. Thus, suppose the agent chooses effort,  $e$ , to maximize the following utility function comprising both extrinsic and intrinsic elements:

$$U = \bar{w} + re - e^2 + I\delta s. \quad (1)$$

The extrinsic motivation of the agent is  $\bar{w} + re - e^2$ , which consists of compensation from a fixed fee,  $\bar{w}$ , and an incentive payment,  $r$ , net of effort cost  $e^2$ . The agent's intrinsic motivation is represented as  $I\delta s$ . The variable  $I$  is an indicator of the presence or absence of the agent's intrinsic motivation, such that  $I = 1$  if the agent is intrinsically motivated and  $I = 0$  if his intrinsic motivation is absent (e.g., crowded out). Following [Murdock \(2002\)](#), the parameter  $\delta$  represents the intensity of the agent's intrinsic motivation for some object,  $s$ , where an agent with a high  $\delta$  will have a stronger motivation toward  $s$  than an agent with a low  $\delta$ .

For simplicity, assume  $p > r$  (so that incentives are marginally profitable to the principal) and  $\delta > 0$  (so that the agent is motivated to be productive rather than destructive). Moreover, 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  $\delta$  defining the strength of the intrinsic motivation they feel, so that agents with higher  $\delta$ s are more intrinsically motivated than agents with lower  $\delta$ s, other things being equal. Assume further that the level of  $\delta$  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  $\delta$ s. Finally, assume that the intensity of an intrinsically motivated agent's innate motivation,  $\delta$ , is fixed; that is, we are abstracting away from the problem of how to



increase or decrease at the margin the intrinsic motivation of workers.<sup>2</sup> The contract offered to the agent consists of a fixed wage,  $\bar{w}$ , and an incentive rate,  $r \geq 0$ , in exchange for effort,  $e$ , chosen by the agent. In period two, the agent makes the effort choice and the principal compensates the worker, which occur simultaneously.

Assume initially the object of the agent's intrinsic motivation is the interests of the principal (e.g., the agent is motivated to generate profits for the principal), so that  $s \equiv pe - \bar{w} - re$  (the principal is also assumed to be interested in securing maximized profits). Accordingly, Eq. (1) is rewritten as

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

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

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

Now consider two different stylized scenarios. In the first, the agent is offered only a fixed fee "for showing up", but he is also intrinsically motivated to supply effort. In this case  $r = 0$  and  $I = 1$ , 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 has no intrinsic motivation to provide effort above what the incentive offers. In this case  $r > 0$  and  $I = 0$ , 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 scenarios conform to the stylization of the MCO problem in the literature, in that workers are often intrinsically motivated when no incentive compensation is offered (i.e.,  $r = 0$  and  $I = 1$ ) but are not intrinsically motivated when incentives are offered (i.e.,  $r > 0$  and  $I = 0$ ). Eq. (2b) summarizes the effort choice of the agent under each of these two scenarios:

$$\hat{e} = \begin{cases} \delta p/2 \equiv \hat{e}_1 & \text{if } r = 0, \ I = 1, \\ r/2 \equiv \hat{e}_2 & \text{if } r > 0, \ I = 0. \end{cases} \quad (2b)$$

According to the experimental evidence, effort supplied under a fixed-fee contract with an intrinsically motivated agent will often equal or exceed effort supplied by an agent not intrinsically motivated when governed by an incentive contract (i.e.,  $\hat{e}_1 \geq \hat{e}_2$ ). This would be true when  $\delta p/2 \geq r/2$ , or when  $\delta \geq r/p$  (i.e., when the intrinsic motivation 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 (equal to  $\hat{e}_1$ ) will not be less than effort provided by the agent when incentives are offered but the agent's intrinsic motivation is absent (equal to  $\hat{e}_2$ ). If incentives are subse-

<sup>2</sup> To be clear here, the question of interest is only whether the agent will behave as if intrinsically motivated (i.e., whether  $I = 1$  rather than  $I = 0$ ), not how intrinsically motivated he would be (i.e., how large or small  $\delta$  is). This is consistent with the assessment of the empirical evidence by Ryan et al. (1996, p. 9) who conclude that the "amount or level of motivation does not necessarily differ when people are autonomous versus controlled, but the type or orientation of motivation does."



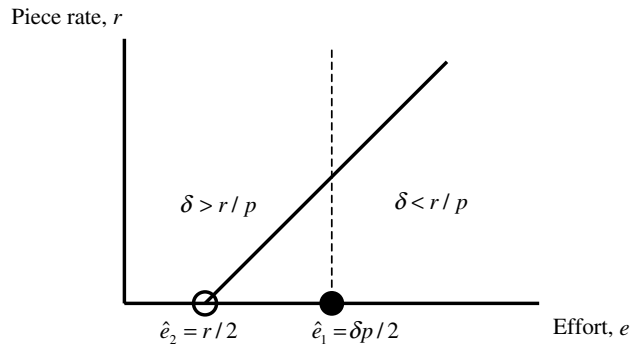


Fig. 1. Agent effort choices under two stylized scenarios. When incentives are zero ( $r = 0$ ) and intrinsic motivation is present ( $I = 1$ ), effort is  $\hat{e}_1$ . When incentives are offered ( $r > 0$ ) and intrinsic motivation is absent ( $I = 0$ ), effort is  $\hat{e}_2$ , which increases in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

quently introduced and cause the agent's intrinsic motivation to be crowded out, so that  $I = 0$ , then the effort choice of the agent discontinuously declines from  $\hat{e}_1$  to  $\hat{e}_2$ . Thereafter, any increase in piece rates results in increased effort (i.e., as  $r$  increases,  $\hat{e}_2$  increases continuously). This is illustrated in Fig. 1.

## 2.2. Motivation crowding out when duty is directed toward a principal

Why might the introduction of an explicit incentive be associated with an apparent disappearance of an agent's intrinsic motivation? That is, why might positive incentives ( $r > 0$ ) result in some agents behaving as if  $I = 0$ ? We examine the utility of an agent when  $I = 0$  (intrinsic motivation crowded out) and  $I = 1$  (positive intrinsic motivation) under the two cases of  $r = 0$  (no incentive compensation) and  $r > 0$  (positive incentive compensation) to answer the following simple question: Will the agent be better off (i.e., prefer or have higher utility) being intrinsically motivated or not being intrinsically motivated?

First, consider the case in which no incentive pay is offered so that  $r = 0$ . If the agent's intrinsic motivation is absent so that  $I = 0$ , then the utility of the agent is

$$U(\hat{e}|r = 0, I = 0) = \bar{w}, \quad (3a)$$

which corresponds to an effort choice of  $\hat{e} = 0$ . If, however, the agent is intrinsically motivated so that  $I = 1$ , then the utility of the agent is

$$U(\hat{e}|r = 0, I = 1) = \frac{\delta^2 p^2 - 4\delta\bar{w} + 4\bar{w}}{4}, \quad (3b)$$

which corresponds to an effort choice by the agent of  $\hat{e}_1 (= \delta p/2)$ . The agent will have greater utility when intrinsically motivated when Eq. (3b) is not less than (3a), or when

$$\delta \geq \frac{4\bar{w}}{p^2} \equiv \delta_{FW}. \quad (4)$$

Eq. (4) defines the threshold determining whether an agent 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 an agent's intrinsic motivation to act in the interest of the principal is relatively high, then he will have higher utility when intrinsically motivated. However, when  $\delta < \delta_{FW}$ , or when the agent's intrinsic motivation is relatively low, then he will have higher utility by choosing effort as if not intrinsically motivated such that  $I = 0$ ; that is, intrinsic motivation would be crowded out. Observe that the threshold  $\delta_{FW}$  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 intrinsic motivation would have to be in order to induce him to provide effort, other things being equal. Simply, if the fixed wage offered to the agent is too large, the agent may decide it is in his interest to take the money and run (i.e., provide no effort) rather than work. According to experimental (and anecdotal) evidence, however, this does not seem to be the case. Most individuals, when offered a fixed wage but no incentive compensation, seem to be willing to work as if intrinsically motivated (i.e.,  $I = 1$ ), suggesting that  $\delta \geq \delta_{FW}$  with effort corresponding 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 so that  $I = 0$ , then the utility of the agent is

$$U(\hat{e}|r > 0, I = 0) = \frac{r^2 + 4\bar{w}}{4}, \quad (5a)$$

which corresponds to an effort choice of  $\hat{e}_2 (=r/2)$ . If, however, the agent is intrinsically motivated so that  $I = 1$ , then the utility of the agent is

$$U(\hat{e}|r > 0, I = 1) = \frac{\delta^2(p-r)^2 + 2\delta(pr - r^2 - 2\bar{w}) + r^2 + 4\bar{w}}{4}, \quad (5b)$$

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

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

Eq. (6) defines the threshold determining whether an agent will be intrinsically motivated when offered an incentive compensation (IC) contract. If  $\delta \geq \delta_{IC}$ , or if the agent's innate intrinsic motivation toward the principal is sufficiently strong, then even when incentive compensation is offered the agent will have higher utility being intrinsically motivated when choosing the level of effort to provide for the principal. If, however, the agent's innate intrinsic motivation is too low in the sense that  $\delta < \delta_{IC}$ , then the agent will derive greater satisfaction if his intrinsic motivation is crowded out in that he chooses effort based solely on the size of the extrinsic incentive,  $r$ . In effect, the incentive compensation results in what can be conceived of as a conscious "deactivation" of the agent's intrinsic motivation by the agent, because doing so results in greater overall satisfaction. Like the case in which no incentive compensation is offered, the threshold  $\delta_{IC}$  increases in the amount of the fixed wage. How the threshold changes as  $p$  and  $r$  increase, however, is indeterminate.

Fig. 2 summarizes the analysis thus far. The experimental evidence suggests that agents generally 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 Fig. 2. Therefore, assume for the following discussion that the  $\delta \geq \delta_{FW}$  condition is satisfied so that the agent is providing effort  $\hat{e}_1$ . Suppose now the agent is offered an incentive. Will the agent choose effort from cell IV or effort  $\hat{e}_2$  in cell II? The experimental evidence suggests that when workers are offered incentive compensation, many who would have been or who were already intrinsically motivated when no incentives were paid subsequently act as if they are not intrinsically motivated, suggesting that  $\delta < \delta_{IC}$  rather than  $\delta \geq \delta_{IC}$ . Why would the introduction of incentive compensation imply that effort in cell II is observed rather than effort in 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  $\frac{4\bar{w}}{p^2} < \frac{4\bar{w}-2r(p-r)}{(p-r)^2}$ , or when

$$r > \frac{p(p^2 - 4\bar{w})}{p^2 - 2\bar{w}} \equiv \bar{r}. \quad (7)$$

	Incentive Absent ( $r = 0$ )	Incentive Positive ( $r > 0$ )
Intrinsic Motivation Absent ( $I = 0$ )	I  $U = \bar{w}$  $\hat{e} = 0$	II  $U = \frac{r^2 + 4\bar{w}}{4}$  $\hat{e} = r/2 \equiv \hat{e}_2$
Intrinsic Motivation Present ( $I = 1$ )	III  $U = \frac{\delta^2 p^2 - 4\delta\bar{w} + 4\bar{w}}{4}$  $\hat{e} = \delta p/2 \equiv \hat{e}_1$	IV  $U = \frac{\delta^2 (p-r)^2 + 2\delta(pr - r^2 - 2\bar{w}) + r^2 + 4\bar{w}}{4}$  $\hat{e} = [\delta(p-r) + r]/2$
Agent intrinsically motivated (i.e., $I = 1$ ) when:	$\delta \geq \frac{4\bar{w}}{p^2} \equiv \delta_{FW}$	$\delta \geq \frac{4\bar{w} - 2r(p-r)}{(p-r)^2} \equiv \delta_{IC}$

Fig. 2. Derived utility and optimal effort choices for an agent who is intrinsically motivated ( $I = 1$ ) and is not intrinsically motivated ( $I = 0$ ) under the two cases of no incentive compensation ( $r = 0$ ) and positive incentive compensation ( $r > 0$ ). When incentives are zero, the agent will choose effort  $\hat{e}_1$  in cell III over effort in cell I when  $\delta \geq \delta_{FW}$ , or when intrinsic motivation sufficiently strong. This is consistent with empirical evidence. Motivation crowding out means that when incentives are positive, the agent will choose effort  $\hat{e}_2$  in cell II instead of effort in cell IV, unless  $\delta \geq \delta_{IC}$ , or the strength of intrinsic motivation is sufficiently large.

Eq. (7) says that if the incentive rate 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 be crowded out by the extrinsic incentives. If, however,  $r \leq \bar{r}$ , then  $\delta_{FW} \geq \delta_{IC}$ . Therefore, agents who would be intrinsically motivated when no incentive compensation is offered (i.e.,  $\delta \geq \delta_{FW}$ ) will also be 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 and thus results in lower effort levels depends in part on how large the incentive rate is. For example, although nominal or “symbolic” (e.g., non-monetary or verbal) rewards would not likely have an adverse effect on intrinsic motivation, large tangible incentives designed to elicit higher levels of effort may be sufficient to erode intrinsic motivation, a prediction consistent with empirical evidence (see Cameron et al., 2001; Deci et al., 1999a; Frey & Jegen, 2001). As explained above, a reason that “large” rewards crowd out intrinsic motivation is that they might be perceived by agents as controlling and, if so, would provide no intrinsic satisfaction to the agent. Therefore, his behavior would likely reflect extrinsic satisfaction only.

Explicit incentives need not be large in an absolute sense before motivation crowding out occurs. Even small monetary incentives could induce agents to behave as if not intrinsically motivated depending on how valuable the agent’s efforts are for a principal, the size of the fixed wage, the agent’s innate intensity of intrinsic motivation, and other factors. For instance, observe that the incentive rate threshold,  $\bar{r}$ , increases in the revenue generated per unit of effort,  $p$ ; that is,

$$\frac{\partial \bar{r}}{\partial p} = \frac{(p^2 - \bar{w})^2 + 7\bar{w}^2}{(p^2 - 2\bar{w})^2} > 0. \quad (8a)$$

In other words, the more “valuable” or important the agent’s effort is to the principal, as measured by the revenue generated by the worker’s actions, the higher is the incentive rate the principal can offer the agent before MCO occurs, other things being equal. This is intuitive. Workers who are intrinsically motivated to take actions that benefit a principal will “naturally” be pleased when their efforts actually benefit the principal. Accordingly, it would take a larger incentive before these workers will experience the MCO effect compared to workers who provide effort resulting in only a small monetary gain to the principal.

Observe further that  $\bar{r}$  decreases in the amount of the fixed wage,  $\bar{w}$ ; that is,

$$\frac{\partial \bar{r}}{\partial \bar{w}} = -\frac{2p^3}{(p^2 - 2\bar{w})^2} < 0. \quad (8b)$$

This says that as the fixed wage increases, the minimum incentive rate at which MCO occurs will decline, suggesting that it is really total compensation, not just the incentive rate, which is important for MCO. Agents paid a relatively low fixed wage will not necessarily be rationally compelled to ignore their intrinsic interests when low or

even moderate levels of incentives are offered. However, when agents are paid large fixed wages, then even small incentives could produce the MCO effect.

### 2.3. Intrinsic motivation when duty is directed toward a social norm

In the preceding model, it was assumed that the object of an agent's intrinsic motivation was also the source of the agent's extrinsic rewards, in that the agent was assumed to derive satisfaction from acting in the interest of the principal and the principal was assumed to have an interest in maximizing profits. Suppose now that rather than being directed toward a principal, the object of an agent's intrinsic motivation is 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, so that  $s \equiv e - \bar{e}$ . Thus, instead of an objective function defined by Eq. (1'), the agent's objective function would be represented as

$$U = \bar{w} + re - e^2 + I\delta(e - \bar{e}). \quad (9)$$

In this formulation, the utility of the agent is increased when  $e > \bar{e}$  and is decreased when  $e < \bar{e}$ , thus giving the agent an incentive to adhere to the 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 such as "provide an honest day's work for an honest day's pay".

Given Eq. (9), the agent's optimal effort choice is

$$\tilde{e} = \frac{I\delta + r}{2}. \quad (10)$$

Consider now the utility of the agent under the cases of  $I = 0$  or  $I = 1$  when  $r = 0$ , and  $I = 0$  or  $I = 1$  when  $r > 0$ . First, suppose no incentive compensation is offered so that  $r = 0$ . If the agent is not intrinsically motivated to follow the social norm, then his utility is

$$U(\tilde{e}|r = 0, I = 0) = \bar{w}, \quad (11a)$$

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

$$U(\tilde{e}|r = 0, I = 1) = \frac{\delta^2 - 4\delta\bar{e} + 4\bar{w}}{4}, \quad (11b)$$

which corresponds to effort choice  $\tilde{e} = \delta/2$ . The agent will have higher utility by behaving as if intrinsically motivated when Eq. (11b) is not less than Eq. (11a), or when

$$\delta \geq 4\bar{e} \equiv \tilde{\delta}_{FW}. \quad (12)$$

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

$$U(\tilde{e}|r > 0, I = 0) = \frac{r^2 + 4\bar{w}}{4}, \quad (13a)$$

which corresponds to effort choice  $\tilde{e} = r/2$ . If, on the other hand, the agent is intrinsically motivated when incentive compensation is provided, then his utility is

$$U(\tilde{e}|r > 0, I = 1) = \frac{\delta^2 - 4\delta\bar{e} + 2\delta r + r^2 + 4\bar{w}}{4}, \quad (13b)$$

which corresponds to effort  $\tilde{e} = \frac{\delta+r}{2}$ . Eq. (13b) would not be less than Eq. (13a), or the agent will have greater utility when intrinsically motivated to adhere to the social norm when

$$\delta \geq 4\bar{e} - 2r \equiv \tilde{\delta}_{IC}. \quad (14)$$

Because the incentive rate is assumed to be non-negative, the threshold level of innate intrinsic motivation necessary to induce intrinsic motivation when incentive are offered will always be lower than the threshold required when only fixed compensation is offered; that is,  $\tilde{\delta}_{FW} \geq \tilde{\delta}_{IC}$ . Therefore, if the agent is already intrinsically motivated to work when there are no extrinsic incentives to providing higher levels of effort (i.e.,  $\delta \geq \tilde{\delta}_{FW}$ ), then intrinsic motivation would not be crowded out when incentive compensation is offered, since  $\delta \geq \tilde{\delta}_{FW}$  implies  $\delta \geq \tilde{\delta}_{IC}$  in this case.

### 3. Conclusions

This paper introduces a simple model of agent utility to illustrate how and why the introduction of incentives might result in the crowding out of intrinsic motivation. In this context, utility is interpreted in the Benthamite sense of overall satisfaction; hence, the analysis involves an examination of whether the agent will have higher utility by being intrinsically motivated or by not being intrinsically motivated when explicit incentives are introduced. The analysis shows that two factors help explain MCO – the object on which intrinsic motivation is tied and the size of the incentive and fixed compensation. These factors are consistent with cognitive evaluation theory, which posits that events perceived as controlling would induce agents not to be intrinsically motivated since controlled behavior does not provide intrinsic satisfaction.

The analysis suggests the following testable hypotheses: First, agents motivated to advance a principal's interests would be more likely to experience the MCO effect than workers intrinsically motivated to adhere to generalized norms not unique to a specific principal. For instance, the more loyal a worker is to a particular principal, other things being equal, the more likely incentives will result in MCO because incentives offered to loyal workers might be perceived as controlling. This prediction has some empirical validity (see [Barkema, 1995](#)).

Second, the introduction of relatively low incentives may not necessarily result in the MCO effect if the fixed pay offered to agents is also relatively low. For instance, “symbolic” incentives, such as peer recognition, non-monetary awards, or verbal praise, might not necessarily result in motivation crowding out (see [Frey & Jegen, 2001](#)) and could in fact increase effort because it signals appreciation or recognition for effort expended. However, when fixed compensation is large, then even small incentives could result in MCO. The key here has to do with the size of the total compensation paid to workers vis-à-vis the benefits worker efforts provide the principal and the agent’s innate level of intrinsic satisfaction. If total compensation is too large, then the salience of the extrinsic reward might be so overwhelming that the agent is rationally compelled to perceive the compensation as a mechanism of control, thus resulting in MCO. This suggests that an agent’s effort supply curve might be S-shaped in the sense that small incentives increase effort because they are not perceived as controlling, larger compensation reduces effort because of its negative effect on intrinsic motivation, but further increases in compensation increase effort in the standard economic sense.

Of course, an agent with a high innate intrinsic intensity ( $\delta$  in this paper) would not necessarily experience the MCO effect. This is consistent with some empirical evidence that finds that activities with high inherent interest motivation are not always negatively affected by extrinsic rewards (see [Cameron et al., 2001](#)). Consequently, the practical relevance of MCO depends fundamentally on how strong an agent’s innate internal motivation is and to what that motivation is tied. This implies that principals have an interest in expending resources searching for workers who have high intrinsic motivations. However, this paper has also shown that understanding the object on which an agent’s intrinsic motivation is placed is just as important in principal–agent relationships as the development of optimal incentive contracts. 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 maximize profits as enriching the owners at the expense of the firm and thus perceive incentives as a blatant form of control. 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).<sup>3</sup> The idea is that under certain circumstances firm owners could increase their residual returns by fostering inefficiencies within the organization. If workers expect this because of employer attempts to control worker effort by means of extrinsic rewards, workers may be less inclined to “trust” firm owners who claim that certain activities or programs are for the good of the firm, thus resulting in lower returns for the firm owners. This might explain why [Fehr and Gächter \(2000a\)](#) found in their experiments that principals

---

<sup>3</sup> [Holmstrom’s \(1982\)](#) impossibility theorem states that firm owners cannot simultaneously balance their budget (or pay wages totaling less than revenue) and devise an incentive plan that will induce all workers to reveal their true effort costs.



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 \(2002\)](#) argument that intrinsic motivation and incentive contracts could be complementary. The reason is that workers may perceive these actions as credible efforts by firm owners to improve organization efficiency rather than principal profits, thus encouraging agents to remain intrinsically motivated when explicit incentives are offered.

This paper advances the existing literature because it integrates the known psychology of MCO with alternative conceptions of utility to suggest that motivation crowding out might be the result of conscious choices of individuals to modify their motivational orientations when explicit incentives are introduced. The model is consistent with much of the empirical evidence, and it produces a number of testable hypotheses. Clearly, however, more work is needed in modeling and explaining this important effect from an economic perspective. For example, the model presented here examines the effects of rewards, rather than fines, on intrinsic motivation. Are the results similar for fines (e.g., does the size of fines matter)? The model is also admittedly simple, in that it treats the extrinsic and intrinsic components of agent utility as additively separable. This may not be consistent with some empirical findings that suggest the relationship between extrinsic and intrinsic motivation is interactive ([Calder & Staw, 1975](#)). Moreover, the model does not fully conform to all empirical evidence. It is well known that monetary compensation often reduces the incidence of blood donations. If one assumes that generalized norms govern the donation of blood, then according to the model developed in this paper compensation should not have a negative effect on the level of donations. Of course, it is possible that factors other than generalized norms are at work here and that the motivation to donate blood is not a generalized social norm but rather reflects an object that also supplies the rewards. But this would have to be explored in greater detail in order to understand and explain potential discrepancies between this model and any contrasting evidence. Therefore, economic models of agent motivation that more carefully describe the interrelationship between extrinsic rewards and intrinsic motivation consistent with the psychology of motivated behavior and the empirical evidence would greatly advance our understanding of motivation crowding out.

## **Acknowledgments**

I am grateful for helpful ideas offered by Kenneth Benson, Peter Klein, Mike Panik, Farhad Rassekh, Chris Starmer, Michael Sykuta, Lusine Voltmer-Darmann, and two anonymous referees on earlier versions of this paper. This research was supported in part by the Missouri Agricultural Experiment Station.

## References

- Barkema, H. G. (1995). Do job executives work harder when they are monitored? *Kyklos*, 48, 19–42.
- Calder, B. J., & Staw, B. M. (1975). Self-perception of intrinsic and extrinsic motivation. *Journal of Personality and Social Psychology*, 31(4), 599–605.
- Cameron, J., Banko, K. M., & Pierce, W. D. (2001). Pervasive negative effects of rewards on intrinsic motivation: The myth continues. *The Behavior Analyst*, 24, 1–44.
- deCharmes, R. (1968). *Personal causation: The internal affective determinants of behavior*. New York, NY: Academic Press.
- Deci, E. L. (1975). *Intrinsic motivation*. New York, NY: Plenum Press.
- Deci, E. L., Koestner, R., & Ryan, R. (1999a). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627–668.
- Deci, E. L., Koestner, R., & Ryan, R. (1999b). The undermining effect is a reality after all – extrinsic rewards, task interest, and self-determination: Reply to Eisenberger, Pierce, and Cameron (1999) and Lepper, Henderlon, and Gingras (1999). *Psychological Bulletin*, 125(6), 692–700.
- Deci, E. L., & Ryan, R. M. (1980). The empirical exploration of intrinsic motivational processes. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 13, pp. 39–80). New York, NY: Plenum Press.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum Press.
- Fehr, E., & Gächter, S. (2000a). Do incentive contracts crowd out voluntary cooperation? University of Zurich, Institute for Empirical Research in Economics, working paper no. 34.
- Fehr, E., & Gächter, S. (2000b). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, 14(3), 159–181.
- Frey, B. S. (1993). Does monitoring increase work effort? The rivalry with trust and loyalty. *Economic Inquiry*, 31(October), 663–670.
- Frey, B. S. (1994). How intrinsic motivation is crowded out and in. *Rationality and Society*, 6, 334–352.
- Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys*, 15(5), 589–611.
- Frey, B. S., & Oberholzer-Gee, F. (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *American Economic Review*, 87, 746–755.
- Frey, B. S., & Stutzer, A. (forthcoming). Beyond outcomes – measuring procedural utility. Oxford Economic Papers.
- Gneezy, U., & Rustichini, A. (2000). Pay enough or don't pay at all. *Quarterly Journal of Economics*, 115(3), 791–810.
- Holmstrom, B. (1982). Moral hazard in teams. *Bell Journal of Economics*, 13, 324–340.
- Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics*, 112(2), 375–405.
- Kreps, D. M. (1997). Intrinsic motivation and extrinsic incentives. *American Economic Review*, 87(2), 359–364.
- Lepper, M. R. (1981). Intrinsic and extrinsic motivation in children: Detrimental effects of superfluous social controls. In W. A. Collins (Ed.), *Aspects of the development of competence: The Minnesota symposium on child psychology* (Vol. 14, pp. 155–214). Hillsdale, NJ: Erlbaum.
- Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic rewards: A test of the 'overjustification' hypothesis. *Journal of Personality and Social Psychology*, 28, 129–137.
- Miller, G. (2001). Why is trust necessary in organizations? The moral hazard of profit maximization. In K. S. Cook (Ed.), *Trust in society* (pp. 307–331). New York, NY: Russell Sage Foundation.
- Murdock, K. (2002). Intrinsic motivation and optimal incentive contracts. *Rand Journal of Economics*, 33(4), 650–671.
- Newman, J., & Layton, B. D. (1984). Overjustification: A self-perception perspective. *Personality and Social Psychology Bulletin*, 10(3), 419–425.
- Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives*, 14(3), 137–158.

- Ryan, R. M., Sheldon, K. M., Kasser, T., & Deci, E. L. (1996). All goals are not created equal: An organismic perspective on the nature of goals and their regulation. In P. M. Gollwitzer et al. (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 7–26). New York, NY: The Guilford Press.
- Sorrentino, R. M. (1996). The role of conscious thought in a theory of motivation and control. In P. M. Gollwitzer et al. (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 619–644). New York, NY: The Guilford Press.