

**EXTRINSIC REWARDS AND INTRINSIC MOTIVES:  
STANDARD AND BEHAVIORAL APPROACHES  
TO AGENCY AND LABOR MARKETS**

**[WORK IN PROGRESS. COMMENTS WELCOME.]**

JAMES B. REBITZER AND LOWELL J. TAYLOR

ABSTRACT. Economists understand that mechanisms used by employers to resolve agency problems have surprising and profound consequences; when employers shape personnel policies to motivate workers, these efforts in turn affect organizational structures and such key labor market outcomes as wage inequality and involuntary unemployment. These insights emerge even in models that ascribe simple motives to employees, e.g., that workers seek to earn money with minimal effort. But worker motives are doubtless multifaceted. Human motivation also includes the desire to compare favorably to others, aspirations to contribute to intrinsically worthwhile goals, and inclinations to reciprocate generosity or exact retribution for perceived wrongs. People are driven, in ways that may be opaque even to themselves, by the need to reinforce *identity*. In this essay we explore the possibility that such behavioral features have important implications for the way we understand agency, and the role of agency plays in labor markets.

1. INTRODUCTION

Many of the most widely-discussed and contentious issues facing the U.S. economy concern the use of financial incentives to solve agency problems. These agency problems often involve the actions of individuals whose actions have great economic and social consequences. For instance, over the past year, multi-million dollar CEO contracts have been the subject of headline news. Similarly, in current debates about policy surrounding health care delivery systems in the U.S., there is debate over the merits of “pay for performance” for physicians as a means to improve patient outcomes and curtail costs. As yet another example, there are concerns that dysfunction in U.S. public education is driven in part by the misaligned incentives for administrators and teachers.

In each of debate, advocates argue that high powered incentives are necessary to get important work done efficiently. Thus, even the very large bonuses to top executives and elite

---

*Date:* November, 2009. Author affiliations: Rebitzer, Boston University and NBER; Taylor, Carnegie Mellon University.

financial engineers are “worth it” in the sense that expected gains from improved performance easily exceed the amount paid out. Critics counter that advocates for high powered incentive systems misunderstand human motivation. High powered incentives may also be unnecessary because appropriately motivated, selected and socialized agents will perform as well or better when stakes are lower. From this perspective incentives are inefficient because they generate unnecessary and potentially costly inequality within work groups or peer groups and because they needlessly divert agents’ attention away from valuable aspects of their job that are hard to monitor and reward. In extreme cases, powerful incentives can cause agents to engage in malfeasance (backdating stock options). Even more provocatively, some critics argue that the provision of extrinsic incentives actually undermines agents’ intrinsic motives and, in this way, worsen the incentive problem they are designed to solve.

This critique of incentive pay has profound implications for management and public policy. A nice example of this is the management literature on “high commitment” human resource systems. As John Roberts describes these systems:

... people work hard and cleverly in the interests of the firm, in return for good pay, empowerment, trust, and interesting, fulfilling work. They do so without any obvious, explicit formal incentives because they identify their interests and those of the firm. ... [M]utual monitoring and social pressure among employees enforce desired behaviors.

Given this conception, managers who motivate solely by calibrating the size of rewards and their link to performance measures are missing the point. The key to motivation is the process of socialization and selection of employees.

Although advocates and critics may not be fully aware of it, the public controversy about incentive pay is essentially a dispute about the appropriate specification of a work-horse economic model: the principal agent model. In its basic form, this model supports the efficiency of extrinsic rewards for motivating agents. The claims of the critics are supported, however, when more realistic—and *ad hoc*—behavioral assumptions are introduced.

Participants in current debates over agency may also be unaware that the resolution to agency problems can have profound effects on labor markets broadly, affecting gender and racial inequality, and labor market segmentation and unemployment.

In this chapter we review and analyze the principal agent model from a behavioral perspective. Our task is made simpler by the fact that conventional and behavioral principal agent models share a similar structure. In the simplest conventional models a single agent is assumed to have utility that is increasing in earnings and decreasing in the provision of effort. Given this utility function, the principal can assess how the agent will react to any given reward structure, and can often link rewards to performance in a way that induces agents to supply efficient levels of effort—even if agents are entirely self-interested and even if measures of performance are noisy and imperfect. Behavioral models employ the same structure, but modify the agent’s utility function to include additional psychological factors. For instance, the agent may take intrinsic satisfaction from the job, and this might be affected by the reward structure. Solving this new model for a new surplus-maximizing reward structure allows theorists to make statements about how psychological factors alter optimal pay arrangements.

Complicating this approach is the fact that in many labor market settings, it is not sufficient to study an isolated agent responding to the policies of a isolated principal. Principals and agents do not act in isolation from the decisions made by others—and this may lead to unexpected aggregate outcomes. For example, a market in which principals (firms) implement individually optimal pay policies, the labor market equilibrium can be characterized by unemployment and labor market segmentation. And agents rarely function in isolation. They typically work as part of larger groups within organizations and society more broadly, and this can have important implications for the design of reward structures, especially when agents have other-regarding preferences or care about inequality. Similarly, agents may be members of groups with established norms of appropriate behavior and these norms can affect optimal incentive design.

A second complication for conventional principal agent models is that pay structures often perform “double duty,” i.e., they must resolve a motivation problem and some other problem—typically a problem that involves heterogeneity or multi-tasking. Consider, for instance, the problem of heterogeneity in a labor market in which agents differ with respect to willingness to trade money for effort or in which principals differ in their ability to make

good on their promises to agents. If the heterogeneity is unobservable to market participants, then pay structures must be designed with an eye towards the kind of employees that will be attracted, and with an eye also toward the possibility about the firm's choices might reveal important information about the employer. Introducing *selection* and *signaling* into principal agent models leads to important and surprising results such as rat races in which individuals decide to work "too many hours" or firms consciously pay employees "too much."

Another situation in which the pay structure performs "double duty" is when workers must attend to multiple tasks. Pay policies that monitor and reward high performance along one dimension draw effort and attention away from other valuable dimensions of performance. The most important result in the multi-task literature is that through clever job design and task assignment employers can elicit more balanced effort with lower powered incentives when their agents are employees rather than independent contractors. This result has important implications for the theory of the firm. As we discuss, it also has implications for big public policy issues such as health care reform, where the problem of incenting independent contractor physicians is of central importance.

In both their multi-task and signaling variants, the "double duty" incentives plays an important role in behavioral principal agent models. Behavioral approaches complement the conventional multi-task story about the use of job design to elicit balanced effort with low powered incentives. Managers, however, do much more than define tasks for employees—they also persuade them of the legitimacy of the tasks and socialize them into the shared and mutually reinforcing expectations of the group. Theoretical and empirical work on reciprocity and reciprocal altruism offers important insight into this process of persuasion and socialization.

The signaling and selection variants of "double duty" incentives are similarly important for behavioral models. If agents differ in their intrinsic response to an organization's mission, then firms with especially evocative missions may design their pay structures to be sure to attract these workers. This has implications on pay structures in health care and other "caring" professions. A second behavioral application of the signaling/selection variant of "double duty" incentives is more novel. It is based on the assumption that agents don't

understand their true motives and that they learn why they do what they do through actions they take and jobs they accept. From this perspective, the double duty served by incentive systems is to send a signal to the agent herself. These models of self-signaling go farther than most other economic models in representing the underlying psychological processes that sustain motivation. They also raise the possibility that intrinsic motivation is an endogenous result of economic incentives and that an organization or market that relies on high-powered incentives may actually reduce intrinsic motivation.

Our discussion proceeds as follows. In Section 2 we present a conventional principal agent model. We begin with the simplest case of a single isolated principal and single isolated agent. We then consider the complications that arise when we insert this relationship into the context of a firm or a labor market.<sup>1</sup>

In Section 3 we introduce the problem of extrinsic rewards with “double duty” incentives. We discuss three applications: wages as a signal of firm fitness, rat races, and multi-tasking. In each case, the presence of double duty incentives greatly alters the market outcomes and employment relationships.

In Section 4, we introduce behavioral features to our agency model. To keep the discussion manageable, we focus on four issues: inequality aversion, the desire to reciprocate, behavioral norms, and identity/self-image. Many of the interesting applications in this section will focus on professional settings and will touch on professional norms as well as the effects of conflict of interest.

Section 5 considers behavioral issues in the context of double duty incentives. The most interesting question we approach here is whether extrinsic rewards might “crowd out” valuable intrinsic motivation.

We conclude the paper by highlighting what we see as promising areas for future research.

---

<sup>1</sup>As we build our model in this section and throughout the paper, we refer to relevant empirical applications from experiments and from field data.

## 2. AGENCY AND EXTRINSIC REWARDS

## 2.1. A Simple Agency Problem

There are a great many interactions in the labor market that can fruitfully be examined as a principal-agent problem—an interaction in which the principal uses a reward-based structure to motivate an agent to pursue some desired action. As a concrete example, consider a principal that seeks to maximize profit, which depends on the “effort” of an agent and the compensation given to that agent, as specified by

$$(1) \quad g(e) - w.$$

In (1),  $g(e)$  is the value produced for the principal as a consequence of an agent’s “effort”  $e$  (which can be represented as a non-negative scalar) and  $w$  is compensation given to the agent. We assume that  $g'(e) > 0$  and  $g''(e) < 0$  exist and are continuous. As for the agent, we assume simply that utility is  $w - e$ . Thus  $e$  is the money metric disutility of taking the action that benefits the principal. What makes this problem interesting is that the principal cannot directly observe effort.

Although information asymmetry is essential to our story, to set basic ideas, we ask initially what the solution would be if the principal could observe the agent’s effort, and can write a contract specifying effort and wage. The firm would then simply maximize (1), subject to the agent’s *participation constraint*, which specifies that the utility resulting from the agreed-upon wage and effort must equal or exceed the utility available to the agent elsewhere, i.e.,  $w - e \geq v$ . The principal finds it most profitable to operate with this latter constraint binding, so it immediately follows that the solution to this constrained maximization problem entails

$$(2) \quad g'(e^*) = 1.$$

This outcome is efficient: the marginal value of additional effort equals the marginal cost to the worker of supplying the effort. This agreed-upon effort level is the same as if the agent worked for herself. The resulting wage is  $w^* = v + e^*$ .

When, instead, the principal does *not* observe the agent’s effort level, the principal must find an incentive mechanism to induce the desired effort level. One possibility is that pay

can be conditioned on the value of output  $g(e)$ . In some instances, though, neither output nor effort are readily observable. We consider such a case. We suppose, instead, that the principal has a noisy signal of effort,

$$(3) \quad x = e + \epsilon,$$

where  $\epsilon$  is drawn from a differentiable, symmetric, single-peaked density  $f$  (with corresponding cumulative density  $F$ ). There are variety of possible interpretations consistent with this set-up. For example,  $x$  might be some objective measure of performance, and  $\epsilon$  is simply luck. Alternatively, we might interpret  $x$  as a principal's impression or opinion of how well a worker is performing, so  $x$  is unobservable outside parties. The  $\epsilon$  term in this latter case captures miscommunication and misinterpretation of effort.

We start with the case in which  $x$  is common knowledge, adopting the assumption that the principal can be trusted to honor commitments in which compensation  $w$  is conditioned on  $x$ . In this case, there are many incentive schemes that will do. To set the stage for results to come, we work through a particularly simple scheme: we assume that the principal commits to a policy of paying wage  $w_0$  if the observed performance  $x$  falls below some cut-off  $\bar{x}$ , and paying  $w_1 > w_0$  if  $x$  is above that cut-off.<sup>2</sup>

To be clear, we have the following timing in mind: (i) The principal announces the policy (including  $\bar{x}$ ), and posts  $w_0$  and  $w_1$ . (ii) The agent decides whether to accept the job, and if she does, takes hidden action  $e$ , which results in an observed performance outcome  $x$ . (iii) Given  $x$ , the firm pays the agreed-upon wage.

We can easily solve for the optimal wage policy,  $(w_0^*, w_1^*)$ . Conceptually, the first step is to account for the agent's *best response* to the wage policy. At effort level  $e$ , the probability of earning  $w_0$  is  $F(\bar{x} - e)$  and the probability of earning  $w_1$  is  $1 - F(\bar{x} - e)$ . So the agent wants to maximize

$$(4) \quad w_0 F(\bar{x} - e) + w_1 [1 - F(\bar{x} - e)] - e,$$

---

<sup>2</sup>For the moment we take the cut-off as given, but as will be apparent, in this particular model it is important that the cut-off be lower than the hoped-for level of performance,  $e^*$ .

which leads to the best response,  $\hat{e}$ , that solves

$$(5) \quad (w_1 - w_0)f(\bar{x} - \hat{e}) - 1 = 0.$$

From this last expression, we notice that the best response is a function of the difference between the higher wage and the lower wage, say  $b \equiv w_1 - w_0$  ( $b$  is the “bonus” that accompanies the high-performance outcome).<sup>3</sup> Thus we can write  $\hat{e}(b)$ , noting, for future reference that

$$(6) \quad \hat{e}'(b) = f(\bar{x} - \hat{e})/[bf'(\bar{x} - \hat{e})] > 0$$

(under the assumption that the second order condition holds). This makes sense; higher-powered incentives elicit greater effort.

Next, the principal must account for a *participation constraint*. The agent accepts the job only if the expected wage equals or exceeds the agent’s opportunity cost:

$$(7) \quad w_0F(\bar{x} - e) + w_1[1 - F(\bar{x} - e)] \geq v + \hat{e}(b).$$

The principal’s problem then turns out to be straightforward. Expected profit is output minus the expected wage, and given that the participation constraint binds, this is just

$$(8) \quad g(\hat{e}(b)) - [v + \hat{e}(b)].$$

The first order condition to the principal’s profit maximization problem is

$$(9) \quad [g'(\hat{e}(b^*)) - 1]\hat{e}'(b^*) = 0.$$

Above, we noted that  $\hat{e}'(b) > 0$  for any best response, so the elicited effort level,  $e^* \equiv \hat{e}(b^*)$ , described by (??) solves

$$(10) \quad g'(e^*) = 1.$$

The solution thus entails the efficient level of effort, as in (??). The principal pins down  $b^*$  using (??), which can be read as giving  $b^*$  as an implicit function of  $e^*$ . Finally, given  $b^*$  and  $e^*$ , the firm sets the base wage  $w_0^*$  to be as low as possible, while still meeting the participation constraint (??).

---

<sup>3</sup>The second order condition is  $-bf'(\bar{x} - \hat{e}) < 0$ , which is satisfied if  $\hat{e}$  exceeds the cut-off  $\bar{x}$ . Hence our observation in the previous footnote that  $\bar{x}$  must be below the hoped-for level of effort.



We have obviously chosen to work out an extremely simple case as our prototypical principal-agent model.<sup>4</sup> As simple as the model is, it is nevertheless sufficient to make the point that a solution to the agency problem entails a strategy of conditioning pay on observed productivity. This reward structure is sufficient to elicit efficient effort levels even when agents are entirely self-interested and when performance measures are noisy and imperfect.

## *2.2. Agency Matters*

In this section we demonstrate the value of thinking carefully about agency in the context of three labor market applications: (i) CEO compensation, a case in which there is a single agent and single principal, (ii) personnel policies in a firm, which involves a single principal seeking to motivate multiple agents, and (iii) unemployment and labor market segmentation that can arise in labor markets in which multiple principals compete for agents, and in which the motivation problem is addressed by the threat of dismissal. In each of these three cases, solutions to the principal agent problem are seen to have important social consequences. In each case also, empirical evidence indicates that anomalies exist that point to the importance of behavioral aspects that are not included in the standard principal agent set-up.

### *2.2.1. CEO Compensation*

In advanced economies with modern financial systems, top executives of publicly traded corporations and large financial firms play a central role in the allocation of productive resources. Thus the reward structure under which these executives operate is of considerable economic interest. The rapid increase in the pay of CEOs since the early 1980s is also one of the most striking labor market developments of the past 25 years. These pay increases have contributed in an important way to growing income inequality (Levy and Temin, 2007) and they have also been the target of intense public controversy. The rise in CEO compensation is inextricably linked to agency issues because most of the changes in pay are the result of increasing grants of stocks and stock options. For example, Hall and Liebman (1998) report the median elasticity of CEO compensation with respect to firm stock market performance more than tripled between 1980 and 1994, largely because of the rapid rise in

---

<sup>4</sup>Natural alternative conceptions would allow for risk aversion, as in the Holmström's (1979) classic paper.

stock based compensation. Bebchuck and Grinstein (2005) document a continuing rapid growth in equity-based compensation for CEOs and top five executives through 2003.

One of the great appeals of the principal agent model is that it tells us what efficient CEO reward structures ought to look like. A central prediction of the model is that the efficient reward structure for CEOs and other top executives should have higher levels of expected pay and higher incentive intensity than for other employees. As a simple formal example, suppose that the value to a firm of a particular agent's effort is  $g(e) = \theta h(e)$ , where  $h(e)$  is a concave function increasing in  $e$ , and  $\theta$  is a positive constant that differs across individuals within an organization, depending on the importance of that individual's job to the organization's profitability. CEOs and top executives will have the highest values of  $\theta$ . At the efficient level of effort,  $\theta h'(e^*) = 1$  and  $\theta h''(e^*) < 0$  (assuming that the second order condition holds), so  $\frac{de^*}{d\theta} = \frac{-h'(e^*)}{\theta h''(e^*)} > 0$ . Thus effort expectations are highest for CEOs and top executives and, because the agent's "best response" effort is increasing in the size of the bonus,  $b$  will also be highest for them. The size of the bonus,  $b$ , is likely to be very large, particular in an environment in which it is difficult to assess the CEO's absolute performance, i.e., when  $f(\bar{x} - \hat{e})$  is small (see equation (??)).<sup>5</sup>

One way of expressing the agency model presented above is that compensation should be set so that any agent becomes (at the margin and in expected value) the *residual claimant* with respect to her contributions to the firm; her own personal fortune rises or falls as a consequence of the value of the actions she undertakes on behalf of her firm,  $g(e)$ . Now in our set-up above, the firm conditions compensation on an imperfect measure of  $e$ , under an assumption about the infeasibility of measuring  $g(e)$  itself for a typical employee. But in the case of the CEO, her actions might be so consequential to the firm that her contribution essentially represents *firm profit* itself. If so, perhaps the ideal contract would make her, roughly, the residual claimant to the entire corporation. To make that happen, one would

---

<sup>5</sup>When the bonus is sufficiently high, compensation will be ideally structured so that the base pay is negative. If these is infeasible, i.e., if in our context CEO cannot be compelled to pay the firm when performance is poor, it will be necessary to modify the contract to take account of the CEO's *limited liability*. This typically entails compromises in which effort will be less than first best or a surplus transfer occurs from the firm to the CEO. In short, the necessity of having a high bonuses accorded to CEOs and other top executives will translate into high salaries generally.

want to tie CEO compensation tightly to firm profitability (i.e., stock values) and then give the CEO unlimited latitude with regard to actions she takes on behalf of the firm.

At first blush, incentives for CEOs appear match well the predictions of the bare bones principal agent model. Top executives earn multi-million dollar salaries and the bulk of their compensation comes in the form of pay linked to stock-based performance measures, as one would expect if stock markets are efficient evaluators of firm value. Indeed, empirical analysis by Gayle and Miller (forthcoming) indicates that the pattern of rising CEO pay and the rising incentive intensity of this pay over a sixty year period can be explained largely by parameters emphasized in the principal agent model: increasing losses to the firm from CEOs pursuing their own goals rather than value maximization, and the deteriorating value of stock performance as a signal of CEO effort. The former is the result largely of the increasing size of firms.

While it is clear that CEOs ought to function under higher-powered incentives than other employees, it is not clear if compensation boards are setting incentives properly. In their seminal article, "Performance Pay and Top Management Incentives," Jensen and Murphy (1990) estimate that during the period 1969 through 1983, CEO wealth increased by only \$3.25 in response to a \$1000 increase in firm value. This number would seem to offer a *prima facie* case for CEOs having inadequate incentives to increase shareholder value. Hall and Liebman (1998) present empirical evidence that in fact there is a substantially tighter link between CEO compensation and firm value, particularly when they examine more recent periods (1980-1994).<sup>6</sup>

Still, in large corporations CEOs are far from being residual claimants. As Hall and Liebman (1998) suggest, this might pose little problem for the proper alignment of some CEO actions but create large problems for others. For example, a CEO who receives \$1 in compensation per \$100 value created for a firm might be sufficiently motivated to make smart carefully-reasoned decisions about resource allocation to proposed projects. But this

---

<sup>6</sup>In particular, they estimate that during the period 1980-1994 a typical CEO whose actions caused the firm to move from the 30th percentile of annual returns to the 70th percentile enjoyed an increase in annual compensation of \$1 to 4 million dollars (1994 dollars), mostly though the increased value of the CEOs stocks and stock options. And for stellar performance the increase in CEO wealth was estimated to be much higher.

same CEO gets an effective 99% discount on a multi-million jet purchased by the firm for his own use.

This latter point is easily illustrated with a slight modification to our baseline principal agent model. Suppose, now, that firm profit is

$$(11) \quad g(e) - w_S - w_N,$$

where now  $w_S$  is the CEO's salary and  $w_N$  is the non-salary cost that results from the CEO's actions that increase the CEO's welfare at the expense of the firm (e.g., expenditures on a jet for CEO use). We now let the CEO's utility be  $w_S + u(w_N) - e$ , where  $u(w_N)$  is the money metric value to the CEO of non-salary expenditures—a function that is obviously increasing in  $w_N$ . We also expect  $u'(0) > 1$  and  $u''(w_N) < 0$ .<sup>7</sup> If the firm could observe and direct  $e$  and  $w_N$ , it would choose  $e^*$  and  $w_N^*$  so that

$$(12) \quad g'(e^*) = 1 \quad \text{and} \quad u'(w_N^*) = 1.$$

Suppose instead the firm sets the variable component of the CEO's compensation equal to the share  $s$  of the firm's profit, i.e.,  $s[g(e) - w_N]$ . The best response here will entail the CEO choosing

$$(13) \quad g'(e^{**}) = 1/s \quad \text{and} \quad u'(w_N^{**}) = s.$$

Comparison of (12) and (13) demonstrates the problem: If  $s < 1$ , we have too little CEO effort on behalf of the firm and too much squandering of resources on the CEO.

How should the corporation's compensation board respond? One argument is that  $s$  must be driven ever closer to 1, even if this entails a substantial direct surplus transfer to the CEO. An alternative might entail the judicious combination of monitoring and more-narrowly directed incentives—a process that would likely play on the hope that hard-to-observe excessive levels of  $w_N$  by the CEO would be limited by shame or a sense of obligation to share-holders. This latter strategy can only be studied in a set-up that takes such behavioral aspects into account. We return to these issues below.

---

<sup>7</sup>Concavity is natural here. The assumption  $u'(0) > 1$  simply allows for the fact that *some* expenditures on CEO well-being are efficient.

In any event, it is infeasible for CEOs to own large parts of enormous publicly traded corporations, i.e., to literally become residual claimants. Firms might try to create high-intensity incentives through the use of stock options (rather than grants of stock outright), but this can have problems of their own. If options are very far under water, their value as incentives degrades to near zero—obviously an undesirable state of affairs.<sup>8</sup> Conversely, when stock prices are just below the stock price, the payoff to small even small increases in price are huge, and this can create irresistible temptations to game the compensation system. Heron and Lie (2009), for example, estimate that 13.6% of all option grants to top executives during the period 1996-2005 were backdated or otherwise manipulated.

More surprising yet, the use of stock options is that corporations do not restrict grants of options to the top executives in a corporation. Hall and Murphy ( ) report that in S&P 500 corporations, roughly 90% of the outstanding options are granted to employees below the top five executives. This pattern is very hard to reconcile with principal agency models because of the nature of the corporate hierarchy and the vast size of publicly traded firms. The efforts of individual employees below the top five executives, in short, can have little direct influence on the price at which their company's stock trades, so a stock-based compensation of any sort is likely to have little effect on effort levels. The use of stock-based compensation for lower-level employees is even harder to understand if agents are risk averse. Here again, the simple principal agent model appears to be inadequate for explaining compensation practices, perhaps because of the omission of important behavioral aspects.

### 2.2.2. *Personnel Policies*

There have been a great many applications of the principal-agent model for the purpose of understanding compensation policies within firms more broadly. As a baseline example, consider a profit-maximizing employer whose  $n$  workers produce output according to a production function,  $Y = G(e^1, e^2, \dots, e^n)$  per period, where  $e^i$  is worker  $i$ 's effort. As above, output is increasing in effort:  $\partial G / \partial e^i > 0$ .

We continue to assume also that in a given period the worker chooses  $e$  and receives utility  $w - e$ , and that the firm cannot condition compensation on  $Y$  (or, in any event

---

<sup>8</sup>Hall and Knox estimate that at the height of the bull market in 1999, roughly a third of executive options were under water. Companies respond to this non-linearity in stock option returns by increasing option grants following stock price declines.

chooses not to). Importantly, for this application, we also assume that workers do not base their effort decisions on the effort or compensation of other workers.<sup>9</sup> Then we can treat the firm's agency problem with a given worker in terms of the function,  $g(e^i)$  which is the value of product that results from production  $G(e^i, e^{(-i)})$ , where  $e^{(-i)}$  denotes effort levels of workers other than  $i$ . We assume that  $g'(e^i)$  and  $g''(e^i)$  are continuous, and that  $g''(e^i) < 0$ .

In setting up our baseline agency model in Section 2.1 we ignored an issue that is generally germane in the workplace: The indicator of performance,  $x$  in our model, is typically observed only by the manager and by workers within a firm, and thus cannot readily be used as the basis for forming contracts that can be enforced, say, by an outside court. Workers who understand that they will have no recourse if a manager violates the implicit agreement—"pay a bonus for high observed effort"—will logically decline to accept the proposed agreement. How might a firm proceed in this case? One possibility is to set up a competition among its  $n$  workers. Suppose, for example, that the firm cannot directly condition pay on  $x$ , but *can* commit to an evaluation process at the end of the period in which (i) workers are ordered on the basis of observed performance, and then (ii) the fraction  $P$  who are lowest-performing are paid  $w_0$ , while the remaining high-performing workers are paid  $w_1 > w_0$ . The key idea here is that while individual performance is not well observed, everyone can observe the agreed-upon reward structure and see if the firm is meeting that obligation.

We can easily find a Bayesian equilibrium the which all workers supply the same level of effort in response to the competition. Suppose that worker  $i$  believes that all other workers are going to play  $\tilde{e}^{(-i)}$ . Now what is her best response? If we assume  $n$  is large, the worker can use her knowledge of  $P$  to accurately assess the cut-off value of observed performance, say  $\tilde{x}$ , which separates low- and high-performance assessments. That is, she takes note of the value  $\tilde{x}$  that solves  $F(\tilde{x} - \tilde{e}^{(-i)}) = P$ . Given that value  $\tilde{x}$ , her optimal choice is to set effort level  $e^i$  so as to maximize

$$(14) \quad w_0 F(\tilde{x} - e^i) + w_1 [1 - F(\tilde{x} - e^i)] - e.$$

---

<sup>9</sup>This last assumption follows naturally enough, given the utility function we have specified. In our discussion of behavioral models in Section 4, we allow for the possibility that workers do care directly about the effort or compensation of other workers.

But this is exactly the worker best-response problem we have just solved (compare (??) and (??)). Given this insight, it is easy to verify that the firm has a workable plan here: The firm starts by setting the “tournament prizes,”  $(w_0, w_1)$ , to be  $(w_0^*, w_1^*)$ , as derived in our baseline example in Section 2.1. Then it chooses the fraction  $P^*$  so that  $P^*w_0^* + (1 - P^*)w_1^*$  just satisfies the participation constraint (5). If worker  $i$  believes other workers are choosing effort level  $\tilde{e}^{(-i)} = e^*$ , she responds by also choosing  $e^*$ . All workers behave the same in equilibrium.<sup>10</sup>

The logic outlined in the preceding paragraph is the starting point of Malcomson’s (1984) well-known paper on hierarchy and internal labor markets. He suggests that the “tournament prize” idea can be fruitful for thinking about the internal organization of the workplace. He works with a two-period model. In the first period of one’s career, within a firm, each worker receives the same wage.<sup>11</sup> Then in the second period, the fraction  $(1 - P)$  of workers who have been most successful as junior employees are promoted to high-paying jobs, while the fraction  $P$  who have been less successful are retained in low-paying jobs (at a wage that is high enough to keep them from moving to other firms). The tournament provides an extrinsic reward designed to elicit optimal effort from young workers.<sup>12</sup>

As Malcomson (1984) notes, the simple tournament model we have just outlined is consistent with some commonly observed features of organizations, e.g., that wage structures in organizations are often “hierarchical,” with workers falling into distinct pay grades, that often workers in high-paid positions being promoted from within, that wages typically rise with seniority (perhaps by more than productivity), and that the variance of wages increases with seniority. Indeed one of the major contributions of agency theory to labor economics is

---

<sup>10</sup>It is important here that the manager actually follow through on the promise to award the higher wage to workers who have the highest realized values of  $x$ . This might be sensible, especially if realized values of  $x$  are reasonably well known by people within the firm. After all, why wouldn’t the manager want to reward to workers who have the highest performance outcomes? Having said that, if there is “favoritism” based on other criteria, the proposed incentive plan falls apart.

<sup>11</sup>This wage solves a two-period participation constraint. The first period wage is low, possibly negative.

<sup>12</sup>In Malcomson’s (1984) set-up, the agency problem is left unresolved for older workers. The point is that young workers can be motivated by the promise of future prizes (promotions, raises, etc.). Such incentives are less likely to be effective for workers nearing retirement. This is quite typical in agency-based models of “internal labor markets,” and it doesn’t substantially alter the basic insights generated in these models.

it's ability to help explain the origin of firm wage policies and hence clarify the contribution that personnel practices make to shaping the wage structure.<sup>13</sup>

As was true in its application to CEO compensation, the first-order predictions of the agency model receive considerable empirical support, but there are anomalies that suggest the model may not offer an altogether satisfactory guide to understanding the internal structure of organizations.

To begin, it is important to recognize that extrinsic incentives do matter within organizations, often in exactly the way predicted by simple models of agency: Lazear (2000) found an increase in effort when a glass installer went from fixed pay to pay-for-performance. Kahn and Sherer (1990) document the effectiveness of an evaluation-and-bonus program at a manufacturing firm. Jacob (2003) shows that high-stakes testing in the Chicago Public Schools does alter teacher behavior—intensifying effort in improvement of test-specific skills, while substituting away from low-stakes subjects like science and social studies. Important work by Theodore Groves and John McMillan and their co-authors shows that strengthened incentives led to substantial productivity increases in Chinese industry and agriculture.<sup>14</sup> And, of course, there are many other examples in the literature.

Principal agent models also require that firms are choosing pay policies in an optimal way. It is hard to find direct evidence that pay policies are chosen in this way. Indeed, much of the literature showing that “incentives work” does so by exploiting the measured consequences of poorly designed incentives. That is, they clearly demonstrate that organizations—at least in some cases—do *not* choose incentives optimally. This is clear, for example, in Oyer's (1998) work, which calls attention to the fact that salespeople seem to intensify effort at the end of the fiscal year if by doing so that can clear performance thresholds and earn a bonus. At the organizational level, Courty and Marschke (2004) similarly demonstrate that a large government organization strategically reported performance outcomes to increase earned rewards, and did so at the expense of productive activities. In work with Martin Gaynor (Gaynor, Rebitzer and Taylor, 2004), we document the effects of an HMO's incentive

---

<sup>13</sup>There are now a number of insightful overviews of the topic, including Lazear (1996), Prendergast (1999), Malcomson (1998), Gibbons (1998), and Oyer and Schaefer's chapter in this *Handbook*.

<sup>14</sup>See, e.g., McMillan, *et al.* (1989), Groves, *et al.* (1994), and Groves, *et al.* (1995).



contract designed to limit expenditures by physicians, but our identification strategy relies on the observation that a key feature of the incentive contract was implemented haphazardly. An even more extreme example is Levitt and Jacob’s (2003) demonstration that public school teachers responded to a shift toward higher-powered incentives by cheating—in their case altering questions on standardized tests taken by students.

Some of the ancillary predictions of principal agent models also lack empirical support. In the two period model we present above, compensation in period 1 ought to move inversely to expected compensation in period 2—a result that follows directly from a two-period participation constraint. In an earlier paper studying law firms (Rebitzer and Taylor, 1995a) we tested this hypothesis. We find, contrary to the predictions of our principal agent model, that large law firms with extremely high second period compensation (in the form of the high income of partnership) also pay high their starting associates very high salaries relative to smaller firms. This would seem to indicate that successful law firms use some form of “rent sharing”—a strategy that emerges when we add such behavioral features as “inequality aversion” (in Section 4.1).

A particularly jarring feature of the minimalist agency approaches to personnel practices is an “irrelevance of *ex post* inequality.” The compensation structure emerging from our model might, after all, better be termed “pay for luck” than “pay for performance.” The principal and agent(s) know that the equilibrium effort level is  $e^*$ . Even so, it is important that pay be based on the measure of observed performance so as to form the crucial extrinsic incentives. This feature—that rewards or punishments are based on an observed outcome, not on the actual behavior (even though those behaviors can be deduced)—is very common in game-theoretic approaches, including much of the work presented below. Anybody who has talked with managers (or chaired an economics department) knows that people don’t respond well when they are paid less than co-workers for what appears to be arbitrary, capricious or random reasons. This observation has been widely examined in the behavioral literature on agency, and we will discuss its implications in Section 3.

### *2.2.3. Involuntary Unemployment and Market Segmentation*

In 1984, Shapiro and Stiglitz set out an influential “efficiency wage” model that illustrates an important feature of agency models: the actions an individual firm takes to resolve an

agency problem can give rise to important social costs when adopted throughout the market. In the case of efficiency wage models, the social costs are those arising from involuntary unemployment and labor market segmentation.<sup>15</sup> The set-up we present here is a recasting of the Shapiro and Stiglitz model taken from Ritter and Taylor (forthcoming).

We consider a market in which there are a large number of identical profit-maximizing employers, each of which faces the agency problem we outlined above. Each firm in the model is assumed to behave as outlined in Section 2.1: the idea is to pay well for “good outcomes” while penalizing workers for “bad outcomes” to the maximum extent possible. *Limited liability* is invoked through the assumption that the only penalty that the firm can implement is to dismiss a poorly performing worker. The motivation problem is resolved by employers making jobs sufficiently valuable that workers will provide effort so as to prevent dismissal.

To capture the idea that jobs have value, it is necessary to set the model up in a multi-period framework. The agency model outlined in Section 2.1 is thus assumed to pertain for each period indefinitely and workers are assumed to live indefinitely with a discount rate  $\rho$ .<sup>16</sup>

#### *The Basic Set-Up*

Employees are paid  $w$  for one unit of labor per period. In each period a worker chooses  $e$ , and this produces utility  $w - e$ .

In this model, the alternative to employment is unemployment, which results in utility of  $v = 0$  in the period. The present value of being unemployed is  $V^u$  (which is not 0 because there is some prospect of being hired in the future). Hiring and termination are costless to the firm.

The model is a game between the firm and a single worker with the following order of play in each round: (1) The firm offers a wage  $w$ . (2) The worker chooses effort level  $e$ . (3) Nature plays  $x$  using the distribution  $f(\cdot)$ . (4) The firm pays  $w$ . (5) The firm decides whether to retain the worker or end the game. We focus on the perfect Bayesian equilibrium

---

<sup>15</sup>Similar points were also made in the important work of Bowles (1985).

<sup>16</sup>Long term employment relationships are central to these models and so, therefore, must be shocks to employment. The model can be enriched to allow for exogenous shocks to employment relationships, but for simplicity we omit this feature.

in which the worker is retained if and only if  $x$  exceeds an endogenous threshold  $\bar{x}$ . We assume that  $\bar{x}$  is common knowledge. (Workers can infer  $\bar{x}$  by observing the frequency of terminations.)

The solution method mimics the steps we took in the simpler model above. In particular, we first find the worker's best response. Then we see how the firm will choose its personnel policies ( $\bar{x}$  and  $w$ ) in light of the worker's best response.

*The Worker's Best Response*

Let  $\hat{e}(w, \bar{x})$  be the worker's best response. To find that best response notice that for a person who chooses  $e$  in the current period, and then reverts to  $\hat{e}$  in all future periods, lifetime utility is given by

$$(15) \quad V(e) = w - e + \frac{[F(\bar{x} - e)V^u + (1 - F(\bar{x} - e))V(\hat{e})]}{1 + \rho}.$$

The employee maximizes  $V(e)$  by choosing  $\hat{e} > 0$ . For an interior solution, the first-order condition is

$$(16) \quad \frac{f(\bar{x} - \hat{e})[V(\hat{e}) - V^u]}{1 + \rho} - 1 = 0.$$

As in our baseline agency model, the second-order condition holds when  $f'(\bar{x} - e) > 0$ . As we have noted, this incentive elicits effort because the job is valuable:  $[V(\hat{e}) - V^u] > 0$ . As is typical of models that invoke limited liability, the participation constraint does not bind.

Evaluating equation (16) at  $e = \hat{e}$  and solving for  $V(\hat{e})$ , then substituting into the first-order condition (16) produces

$$(17) \quad w = \hat{e} + \frac{\rho V^u}{1 + \rho} + \frac{\rho + F(\bar{x} - \hat{e})}{f(\bar{x} - \hat{e})}.$$

This last expression implicitly defines the worker's best response.

*Firm Profit Maximization*

Now we can turn to the firm's objective. It seeks to maximize profits, taking into account the worker's best response, i.e., to maximize

$$g(\hat{e}(w, \bar{x})) - w.$$

The solution can readily be found using the same steps we followed in solving the agency problem in Section 2.1.

In this instance the optimal employment policy again induces the socially optimal performance level regardless of  $f(\cdot)$ :

$$g'(e^*) = 1.$$

The noise in the environment does, however, affect the distribution of surplus. In particular, Ritter and Taylor (forthcoming) establish the following results: First, when the firm optimally chooses  $w$  and  $\bar{x}$ , the resulting probability of retention,  $F(\bar{x}^* - \hat{e}^*)$ , is unaffected by the level of  $\sigma^2$  (the variance of the density  $f(\cdot)$ ). Second, the optimal wage does depend on  $\sigma^2$ , as follows:

$$(18) \quad w^* = e^* + \frac{\rho V^u}{1 + \rho} + \frac{\phi(z^*)}{\phi'(z^*)} \sigma,$$

where  $z^*$  is a “standardized” random variable,  $z^* = (\bar{x}^* - e^*)/\sigma$ , and  $\phi(\cdot)$  is the “standardized p.d.f.,” i.e., the p.d.f. of  $\frac{\epsilon}{\sigma}$ . The more intractable the agency problem—the greater the value of  $\sigma$ —the higher is the wage required to achieve efficient effort and so the greater the surplus accruing to the worker.

### *Equilibrium Unemployment*

The equilibrium of the model we have just outlined generates unemployment.

Let  $V^*$  be the present value of lifetime utility of an employed individual who works at the optimal effort level. Recall that  $V^u$  is the expected lifetime utility for an individual who is unemployed. This utility level depends, clearly, on the probability of job acquisition. Let that rate be  $a$ .<sup>17</sup> Assuming that the current-period utility of an unemployment person is zero, the expected lifetime utility of an unemployed individual is

$$(19) \quad V^u = 0 + \frac{[aV^* + (1 - a)V^u]}{1 + \rho},$$

so in turn we can use (??) and (??) to solve for  $V^u$  and substitute into equation (??), giving

$$(20) \quad w = e^* + \frac{1}{\phi(z^*)} \left( a + \frac{\phi(z^*)^2}{\phi'(z^*)} \right) \sigma.$$

Equation (??) gives the locus of potential equilibrium values of  $w$  and  $a$ .

---

<sup>17</sup>Note that the rate  $a$  is a known constant to any individual, but of course is endogenous to the economy as a whole. We solve for the equilibrium rate  $a$  shortly.

Now job loss among the employed occurs with probability  $F^* \equiv F(\bar{x}^*, e^*)$ , while job finding among the unemployed occurs with probability  $a$ . So if  $u$  is the steady state unemployment rate, we must have  $au = (1 - F^*)u$ . Solving for  $a$  and substituting into (??) gives

$$(21) \quad w = e^* + \frac{1}{\phi(z^*)} \left( \rho + \frac{F^*}{u} \right) \sigma.$$

This expression shows potential equilibrium wage and unemployment levels for the labor market.

Figure 1 shows the market equilibrium when long-run labor demand is perfectly elastic, at  $w^E$ . (More general formulations are easily handled.) Equilibrium unemployment,  $u^E$ , solves

$$(22) \quad w^E = e^* + \frac{1}{\phi(z^*)} \left( \rho + \frac{F^*}{u^E} \right) \sigma.$$

Clearly  $u^E > 0$ . Also, inspection of (??) shows that an increase in  $\sigma$  increases unemployment. This outcome is intuitive. The weaker the link between the dismissal threat and employee behaviors, the stronger are the incentives required to elicit the desired effort level. In equilibrium, heightened incentives require higher unemployment.

This model of equilibrium unemployment—the Shapiro-Stiglitz model—has emerged as a workhorse for the analysis of macro-labor issues. It has proved to be useful also for evaluating policies like unemployment benefits, the public interest in regulating firms' layoff decisions (i.e., just-cause dismissal requirements, as discussed in Levine, 1991), and the potential of minimum wage policy to actually increase employment (Rebitzer and Taylor, 1995c). Having said this, economists are divided on the extent to which efficiency wages are an important source of equilibrium unemployment. Other forces, like labor market frictions, matter as well in determining equilibrium unemployment rates.<sup>18</sup> Efficiency wages are clearly not the whole story.

### *Labor Market Segmentation*

---

<sup>18</sup>Hornstein, Krusell, and Violante (2007) offer a review and discussion of models of unemployment resulting from search frictions. Search models produce both unemployment and wage dispersion, but search frictions sufficient to account for equilibrium unemployment imply far less wage inequality than is actually observed.

Although we have focused on unemployment, the lost output from unemployment may not capture the full social costs of efficiency wage personnel policies. After all, if there are some jobs in the labor market where agency issues are of little importance, workers should generally be able to find jobs there. From an efficiency perspective, finding work in these “secondary jobs” is similar to unemployment in that individuals in the secondary labor market would prefer higher-productivity “primary jobs,” but the equilibrium supply of qualified workers for these jobs exceeds the supply.

Labor market segmentation emerges if we enrich our efficiency wage model by allowing the difficulty of agency problems to vary across firms. Recall from (??), that in the Ritter and Taylor (forthcoming) version of the Shapiro-Stiglitz model firms choose to pay

$$(23) \quad w = A + B\sigma,$$

where  $A$  and  $B$  are positive constants that are independent of  $\sigma$ . Firms that have low values of  $\sigma$ , i.e., who have production processes with accurate measures of worker effort, can pay wages that are relatively low. On the other hand, firms will choose to set wages high when effort is hard monitor or, equivalently, when they face high values of  $\sigma$ .

This latter observation was emphasized in Bulow and Summers’ (1986) paper on “dual labor markets.” In their conception, firms with severe agency problems pay high wages and are said to belong in the primary sector. The strategy of paying high wages is effective because workers are motivated by the prospect of retaining valuable jobs. Thus we also expect to observe low levels of voluntary exit from such firms and efforts on the part of firms to retain workers even in a down-turn. In contrast to the primary sector, firms that have modest agency problems can pay wages that are close to the market-clearing level. These secondary sector firms will be less concerned about worker turnover. In an extension of this argument (Rebitzer and Taylor, 1991) we show that firms which employ efficiency wages as a motivating device will also be led to hoard labor, i.e., employ labor above the value-of-marginal-product curve. By taking actions to ensure future employment—perhaps by hiring contingent workers to absorb demand shocks—firms can reduce the wage needed to provide optimal motivation to workers.

The most widely examined empirical prediction of efficiency wage models of labor market segmentation is that there will be cross-firm and cross-industry wage variation resulting from firm characteristics, rather than worker characteristics. There is considerable evidence for industry and firm wage effects (including well-known work by Krueger and Summers, 1988, on industry effects, and Brown and Medoff, 1989, on firm size effects) but it is often unclear how much this is due to factors emphasized in efficiency wage models (such as monitoring difficulties) or other market imperfections such as those emerging from search frictions (e.g., Burdett and Mortensen, 1998).

One potentially helpful approach entails empirical evaluations of specific firms and industries, with an eye toward the central predictions of the model. Thus, Cappelli and Chauvin (1991) examined worker performance across plants within the same firm, examining the extent to which workers seem to choose performance on the basis of the value of their job relative to other opportunities in their local labor market. They find evidence that is generally supportive of the efficiency wage set-up. Similarly, in work with Daniel Nagin and Seth Sanders (Nagin, Rebitzer, Sanders, and Taylor, 2002), we evaluated a field experiment in which a firm manipulated monitoring rates across several work sites. In expectation with the effort-regulation model set up above, there was substantially more malfeasance in locations with low monitoring levels.

At the broadest level, the efficiency wage literature points to important social costs that emerge as a result of the strategies individual firms use to resolve agency problems. If firms indeed rely on the the fear of job loss to motivate employees, labor markets can be expected to waste human capital on a large scale through involuntary unemployment and labor market segmentation. If, however, other motivators can be mobilized to resolve agency problems, the situation may not be so grim. The costs of agency problems might be further reduced if schools can socialize children to be especially responsive to these alternative motivators. Indeed, some have speculated that such socialization may be the source of much of the social and private returns to investments in human capital. We take up some of these alternative motivators in Section 4 below. Before turning to these, however, we must first introduce another conceptual building block that is important for our story—applications in which incentives are intended to work along more than one dimension.

### 3. EXTRINSIC REWARDS AND DUAL-PURPOSE INCENTIVES

In real-world applications, compensation policies are often asked to do “double duty.” A well known and intuitive example of this is Lazear’s (2000) study of compensation practices at Safelite, a windshield installation company. As might be expected from the basic principal agent model, the introduction of an explicit piece rate system induced many workers to perform at a higher intensity level. In addition, the piece rate system had a *selection* effect: workers who disliked having to choose between lower compensation and a faster pace of work left the firm while, at the same time, the firm was able to attract workers drawn to the income-effort tradeoffs inherent in the piece rate system. In this case, incentive pay was serving a dual role: motivating and attracting employees.

At Safelite, the selection reinforced the effect of incentives on work effort. In many cases, however, there is a tension between the multiple effects of incentive pay and employers must make compromises along one dimension in order to accomplish an objective along a second dimension. In the following three sub-sections we give examples of this phenomena and illustrate how the introduction of double duty incentives helps address well known anomalies. The discussion in this section also sets up of the discussion of behavioral models that follow. The special problems posed by dual purpose financial incentives can be either ameliorated or sharpened by behavioral factors of the sort discussed in Section 4. In addition, dual purpose incentives play a key role in the models of intrinsic motivation presented in Section 5.

#### *3.1. High Wages as a Signal of Firm Fitness*

We begin by discussing a theoretical issue that is well known in the literature on efficiency wages. As we have seen, firms that pay efficiency wages must set wages above the market clearing level to elicit work effort. Effort can, however, be elicited more cheaply by a deferred compensation policy that causes employees to, in effect, post a performance bond. By judiciously back-loading pay, firms can create powerful work incentives while choosing a wage path whose discounted present value is equal to the market clearing wage. With this option available, why would employers ever select a more costly efficiency wage strategy? The practical relevance of this theoretical puzzle is sharpened by empirical work suggesting that even when very large amounts of deferred compensation were available, as is the case



in the promotion from associate to partner in large corporate law firms, firms set wages as if they were pursuing an efficiency wage strategy (Rebitzer and Taylor, 1995a).

Ritter and Taylor (1994) tackled this issue by observing that for both efficiency wages and the performance-bond incentive, the power to shape behavior depends on the likelihood that the firm will honor its future commitments to employees. All else equal, firms will more effectively solve agency problems if employees expect them to be highly reliable in honoring future wage commitments.

Ritter and Taylor build upon this insight by positing a market in which there are two types of firms: highly reliable firms (i.e., firms that are unlikely to go bankrupt or otherwise renege on commitments) and less reliable firms (firms that are more likely to become bankrupt or renege). Reliability is known by the firms but not by anyone else, though the distribution of types is common knowledge. Firms would like to resolve their agency problem as cheaply as possible, and are inclined to do so by asking workers to post bonds in the form of deferred compensation. The posted bond is forfeited if the worker is judged to be working at a sub-standard effort level but is returned, with interest, if the worker's observed performance meets the expected standard.

Under some conditions, all firms pursue the same deferred compensation strategy. In this *pooling equilibrium*, workers will require a rate of return on their bonds that reflects the aggregate level of riskiness, based on the market-wide probability a firm will fail and be unable to return the bond.<sup>19</sup> This, of course, is a good deal for low-reliability firms but a bad deal for the other type. A more interesting possibility is that an efficiency wage strategies might emerge for some firms as a *separating equilibrium*. In this equilibrium, highly reliable firms distinguish themselves by their willingness to “place bets” on future compensation outlays. A highly reliable firm that pays large up-front wages, is in effect placing a bet that it is very unlikely that it will benefit at employees' expense by reneging on its commitments in the future. Less reliable firms are, of course, not inclined to match these bets by offering equally high wages up front because they think it is rather likely that they will benefit by reneging. In this way a reliable firm that pays high wages up front offers a credible *signal*

---

<sup>19</sup>In this model, young workers have concerns about the realization of high earnings at the firm later in their career. The wage level thus depends on the degree to which firms that are judged to be unstable.

that it is a highly desirable counterparty for long-term relationships. In short, efficiency wages arise endogenously, without a recourse to limited liability arguments.<sup>20</sup>

Our primary point here concerns the use of incentives policies to do “double duty.” In the separating equilibrium, highly reliable firms use wage policies to solve an agency problem *and* to signal the fitness of the firm. In order to pursue both objectives, these firm must compromise on their use of deferred compensation and this compromise necessarily introduces distortion. Thus, in the separating equilibrium there is a surplus transfer to workers employed by highly reliable firms and employment in the stable firm sector is inefficiently low.

### 3.2. *The Rat Race*

Rat race models build on a simple observation: early in their career many successful professionals appear to be overworking. It is commonplace to find lawyers, consultants, and even assistant economics professors complaining that the hours they put in are simply “too much” and that they interfere with forming and raising a family. These strains are increased by the dramatic influx of women into professional occupations because overwork might well be most intense during prime years for family formation and child bearing. From the point of view of simple models of labor markets, this sort of overwork is anomalous. Firms are in competition for talent, and it would seem that the most successful competitors would be those who best accommodate employee preferences about work conditions—including work hours. In his famous paper on the “rat race,” Akerlof (1976) offers a potential resolution to this anomaly based on unobservable worker heterogeneity among workers.

Akerlof’s set-up focuses on a production line. At the end of the day line workers are jointly rewarded on the basis of total output. There are two types of workers—those inclined to work hard and those inclined to work less hard. To employers these workers appear identical, so they all earn the same wage. If both types of workers accept positions on the production line, this is a great deal for low-effort workers (who would earn lower pay than high-effort workers in a perfect-information world) but a bad deal for high-effort workers.

---

<sup>20</sup>In the law firm context, the term of art for paying very high wages to summer interns and associates is “paying full freight.” Perhaps law firms that are able and willing to “pay full freight” signal that the value of their partnership is high, and this in turn allows them to attract the best talent.

This is precisely the situation that might lead firms to adopt rules that will provide high-effort workers the opportunity to credibly *signal* that they are in fact high-effort workers. Thus, Akerlof's proposed solution is that the firm set the production line at a speed that is uncomfortably fast for high-effort workers but more uncomfortable yet for low-effort workers—so uncomfortable indeed that the low-effort workers will opt out of working for the firm. The rat race thereby serves the useful function of screening out the low-effort workers.

In Akerlof's model, compensation policies are, quite clearly, doing “double duty.” Compensation arrangements and work conditions are structured (i) to compensate workers at a level necessary to induce them to accept employment at the firm, and (ii) to create incentives that attract the “right kind” of worker to the firm. The distortion here is that workers are being asked to provide effort that exceeds the first-best level. In a marketplace with many employers, the market can devolve into an equilibrium in which *all* firms that hire high-effort workers ask those workers to work at uncomfortable effort levels. This “adverse selection” equilibrium occurs because any one firm failing to adopt a rat race would be swamped by low-effort workers. The equilibrium can be inefficient, in the sense that all firms would experience increased profitability if they would coordinate on a lower-effort work norm.<sup>21</sup>

Akerlof's demonstration of an over-work equilibrium was presented in a self-consciously unrealistic example, but subsequent theoretical and empirical work suggests that it is an important phenomenon in professional labor markets. For instance, in a paper with Renee Landers (Landers, Rebitzer, and Taylor, 1996), we embed Akerlof's idea into a simple tournament-partnership model designed to shed light on work practices in large U.S. law firms. In our two-period model, young lawyers accept salaried positions as “associates” for one period, and if deemed suitable are promoted to be “partners” (equity share holders) in the subsequent period. Partners equally share in firm surplus. This equal sharing rule gives

---

<sup>21</sup>Intuitively, the over-work equilibrium might persist even when there is a small number of low-effort workers. No one firm can deviate from the equilibrium without suffering harm from adverse selection. But if all firms backed away from overwork requirements, any one firm would get “stuck” with only a negligible number of low-effort workers. All firms would be better off.

incumbent partners powerful incentives to promote only highly motivated lawyers into the partnership.

We assume that there are two types of lawyers who are equally productive but have differing preferences over the hours they prefer to work: these are “short-hour” and “long-hour” workers. Now in our setting, firms have the incentive to attract workers who will be inclined to work long hours. The reason is that when workers become partners—at which point they share firm surplus—the long-hours individuals will engage in less free riding. As in Akerlof’s model, an adverse selection equilibrium emerges. In our case, associates’ willingness to put in extended hours over many years serves as a credible signal that they are long-hour individuals.

Empirical evidence for the relevance of the over-work equilibrium comes from an empirical examination of work hours and work preferences in associates at two large East Coast law firms. In a survey conducted in these firms, we find that most associates express a preferences for working shorter hours (with a correspondingly lower salary) but, importantly, their willingness to work shorter hours hinges on the work-hour norms adopted by other associates. This last finding would not be expected in a conventional labor market, but is precisely what one would expect if work hours are being used as a signal of one’s otherwise-unobservable type.

In our paper (Landers, Rebitzer and Taylor, 1996) we abstract from “career concerns,” but it is clear that over-work early in one’s career might be valuable not only as a signal within a firm but as a means of advancing one’s career elsewhere. Completion of six years as an associate at a law firm known for abusing associates with grueling hours, can thus be a valuable means of signalling an important but hard-to-observe trait to other employers in the marketplace.

The key idea that current on-the-job behaviors can affect one’s future career, through their impact on reputation, is studied in insightful papers by Holmström (1989) and Gibbons and Waldman (1999). Gicheva (2009) demonstrates how overwork early in one’s career can signal that the individual has valuable traits as a manager. Gicheva shows, further, that these sort of models help explain wage growth in a sample of workers who took the Graduate Management Admissions Test (GMAT). Specifically, she shows that among workers who

worked above-norm hours when they were young (48 or more hours per week), subsequent wage growth was positively correlated with early career hours worked. The same was *not* true for those workers who worked fewer than 48 hours; for those workers wage growth was uncorrelated with hours worked.

Signalling can be particularly dysfunction in situation in which workers can devote effort to more than one task—an issue we take up next—because it can distort effort allocation. An example is given in the work of Acemoglu, Kremer, and Mian (2008). In their model, career concerns can motivate excessive and misguided signaling by primary school teachers—misguided because some effort is devoted to improving the signal (student performance on a proficiency test), without actually improving students’ true human capital. An important conclusion in that paper is that the problem of excessive signalling can fundamentally shape the desirability of using markets *versus* the government for the provision of some services. Specifically, in a competitive market, there will be socially costly distortions as the result of excessive signalling. If teachers are employed in the public sector, the government might be able operate more efficiently because of an ability to commit to policies that eliminate this latter dysfunction.

### 3.3. *Multi-Tasking*

Perhaps the most obvious case of dual-purpose incentives occurs when a principal seeks to regulate an agent’s behaviors along more than one dimension. We have already encountered examples along these lines in our discussions above. For instance, in our examination of CEO compensation, we noted problems that arise when a compensation board seeks to create incentives for a CEO to exert effort toward increasing shareholder value *and* limiting wasteful expenditure in the executive suite. Our discussion of tournament incentives suggests a second obvious example: What happens in a tournament when each worker must be motivated to provide effort along his own assigned task *and* be motivated also to be cooperative with other workers?<sup>22</sup> A third example is Acemoglu, Kremer, and Mian’s work, mentioned in the previous paragraph, in which teachers allocate effort that improves student human capital *and* effort that merely improves a student’s test score (“teaching to the test”).

---

<sup>22</sup>A number of papers have taken up this issue, including Lazear (1989).

Holmström and Milgrom (1991) establish a number of insightful and surprising results in precisely such contexts. The central point of their paper is both simple and profound: when multiple agent tasks are involved, incentives must perform the double duty of inducing appropriately high levels of effort generally *and* inducing a desirable allocation of agents' attention across the various tasks inherent in the job.

We can get a feel for their analysis by making a simple extension to our baseline principal agent model. Let us suppose now that the agent can allocate effort along two dimensions,  $e_1 \geq 0$  and  $e_2 \geq 0$ . We suppose also that the agent's utility is now  $w + d(e_1 + e_2)$ , and, following Holmström and Milgrom, we make two key assumption about the function that gives money metric disutility of effort,  $d(\cdot)$ : (i) it is convex, and (ii) it achieves a minimum at a positive level of effort. This last assumption means that in the absence of direct incentives the worker will be happiest when putting forth some effort. The principal's objective continues to be the maximization of value added by the worker, now given by  $g(e_1, e_2) - w$ .

A principal who has reasonably accurate measures of  $e_1$  and  $e_2$ , say  $x_1$  and  $x_2$ , will be able to construct an incentive scheme in which bonuses reward each dimension of effort appropriately. Matters are more interesting when the principal has good information along one dimension of effort but not the other. To take an extreme example, suppose the firm has a subjective measure  $x_1$ , but no measure at all of  $x_2$ . The firms best strategy then will depend crucially on the nature of the production function,  $g(e_1, e_2)$ . To see how this works, we set  $d(e_1 + e_2)$  to be  $-\frac{1}{2}(e_1 + e_2 - e_B)^2$ , with  $e_B > 0$  representing the agent's "bliss" level of effort, and work out the optimal incentives for two different production functions: first, a case in which the two types of effort are perfect complements,  $g(e_1, e_2) = a_1e_1 + a_2e_2$ ; second, a case in which they are strategic complements,  $g(e_1, e_2) = e_1e_2$ .

The solution to the first case is easy to characterize. If the firm places any incentive whatsoever on the first type of effort (i.e., a bonus based on  $x_1$ ), the worker's best response will be  $\hat{e}_2 = 0$ , and with this in mind the firm can follow the steps outlined for our baseline one-dimensional principal agent model. It is easy to confirm that the result will be to elicit effort  $e_1^* = a_1 + e_B$ . Intuitively, the firm will prefer this strategy if the value of the first type of effort is high relative to the second type of effort, which is certainly true if  $a_1 \geq a_2$ .

On the other hand, if  $a_2$  is sufficiently large, the principal might decide to use no explicit incentive and instead ask (nicely!) that the worker direct all his effort along the second dimension. Given a binding participation constraint,  $w - \frac{1}{2}(e_1 + e_2 - e_B)^2 = v$ , it is easily shown that value added by the worker is

$$(24) \quad g(e_1, e_2) - w = \begin{cases} a_1 e_B + \frac{1}{2} a_1^2 - v & \text{when incentives are placed on } e_1, \text{ and} \\ a_2 e_B - v & \text{when agent effort is directed to } e_2. \end{cases}$$

As anticipated, if  $a_1$  is sufficiently high relative to  $a_2$ , the firm will simply place incentives on the observable portion of performance. This is efficient when  $a_1 \geq a_2$ , and is second-best even when  $a_1$  is moderately lower than  $a_2$ . However, when  $a_1$  is sufficiently low relative to  $a_2$ , the firm will instead try a “cooperative” strategy. No incentives are used; the worker is simply asked to direct all effort to the second dimension.

The solution to the second case, in which the two types of effort are complements, is also intuitive. When  $g(e_1, e_2) = e_1 e_2$ , the firm clearly must avoid a best response of  $\hat{e}_2 = 0$ , and so will *never* use an explicit incentive along the first effort dimension. In this case the firm instead directs the worker effort to be  $e_1 = e_2 = \frac{e_B}{2}$  and hopes that the worker complies.

Simple as this analysis is, several interesting points emerge. We notice, first of all, that there will be cases in which an employer will choose not to place an incentive on an easily-observed dimension of performance, even when that effort is highly valuable to the firm. This happens when there is a similarly-valuable second dimension of effort that is sufficiently difficult to observe and incentivize. Such an outcome is particularly likely when multiple tasks are strategic complements. In such cases the firm is best off using very low-powered incentives, i.e., simply paying a base wage.

Second, our second example illustrates that the second-best outcome, which sometimes necessarily adopted, can be quite far from the efficient outcome. For instance, when  $g(e_1, e_2) = e_1 e_2$ , it is easily confirmed that the efficient level of output is  $e_B^2$ .<sup>23</sup> The firm’s low-powered incentive scheme results, instead, in output  $\frac{1}{4} e_B^2$ . If possible the firm would very much like to find a way to make this worker a residual claimant, and indeed would be willing to suffer substantial cost along some other dimension to make this happen. In

---

<sup>23</sup>Maximize  $e_1 e_2 - w$  subject to  $w - \frac{1}{2}(e_1 + e_2 - e_B)^2 \geq v$ .

short, a strong motive exists here to outsource the task at to an independent contractor if this can be made workable.<sup>24</sup>

Third, when a firm cannot use independent contracting, or for some compelling reason chooses not to, a central goal of the firm often entails structuring activities to improve management’s ability to closely monitor and supervise key contributions by employees. Put another way, the issue of multi-tasking can matter a great deal for the organization of firm production.

Finally, and most important for our purposes, the multi-tasking approach developed by Holmström and Milgrom, and illustrated with the simple example above, clearly opens the door for “behavioral factors” to play a central role. For example, in our model, the “bliss” level of effort ( $e_B$ ) is taken to be an exogenous constant. It is unlikely that this simplification holds in typical real-world settings. Rather, that parameter likely varies across individuals: Some people will work hard in the absence of explicit motivation, while other will work hardly at all. Also, even people who are typically inclined to work hard might feel de-motivated on any given day. Presumably such features must be captured by deeper psychological conceptions.

Also, the default level of effort ( $e_B$ ) is in many contexts likely to be driven by multiple factors, some of which are shaped by the context and some of which might be, at some level, influenced by the firm, either directly or through its organizational structure. These are the sorts of issues that lead one into the territory of behavioral economics—concerns revolving around reference-group comparisons, dedication, confidence, self-doubt, attribution, etc.

#### 4. BEHAVIORAL APPROACHES TO AGENCY AND MOTIVATION

In this section of the paper, we turn to explicitly “behavioral considerations”—the introduction of elements intended to shore up the psychological foundation of agency models. There are are a many ideas from behavioral economics that are potentially applicable: These

---

<sup>24</sup>From a legal perspective, employees are distinguished from independent contractors by the extent of control and supervision the principal exerts over the actions of the agent. A large literature focuses on the forces that drive firm boundaries, focusing on such issues as the direction of employee activities (e.g., Coase, 1937, and Simon, 1951), and firm ownership of assets (e.g., Williamson, 1985, and Grossman and Hart, 1986). The relationship between these issues, and the agency problem—particularly in the multi-tasking setting—is developed most clearly in Holmström and Milgrom (1994).



might include approaches that recognize human limitations in pursuing intended goals, e.g., the tendency to use heuristics to solve problems and condition behaviors, difficulties in making probabilistic judgements (e.g., systematic errors in the use of Bayesian updating), or insufficiency of will-power. We might similarly explore novel ways of thinking about preferences over uncertain outcomes or intertemporal choice. For the most part, though, we will restrict attention to modifications of standard preferences to include such factors as the intrinsic value of one's work or concerns about one's standing relative to others.<sup>25</sup>

Economists have long understood that utility depends on more than selfish inputs, as is typically depicted in agency models.<sup>26</sup> Here we begin our investigation by considering the potential relevance of “social preferences,” which arise when individuals compare their own payoffs and sacrifices to others in a reference group. We proceed to consider the properties of the solutions to agency problems in employment relationships when social preferences are important. An important theme in our discussion is that social preferences matter, but the way in which they matter depends on whether one is considering the agency problem in isolation or in a competitive labor market.

---

<sup>25</sup>Typically we will be studying models that make modest modifications to the standard agency models discussed above. Camerer and Loewenstein (2004) provide a nice justification for this approach: “Theories in behavioral economics . . . strive for generality—e.g., by adding only one or two parameters to standard models. Particular parameter values then often reduce the behavioral model to the standard one, and the behavioral model can be pitted against the standard model by estimating parameter values. Once parameter values are pinned down, the behavioral model can be applied just as widely as the standard one” (Camerer and Loewenstein, 2004, p. 4).

More generally, Camerer and Loewenstein's paper provides an insightful introduction to a broad and rich set of ideas in behavioral economics, including observations about the origins of modern behavioral economics, and suggestions about future directions for the field. DellaVigna (2009) gives a good recent review of behavioral economics, focusing on evidence drawn from the field.

<sup>26</sup>For example, several decades back Gary Becker initiated important strands of inquiry in economics by positing preferences that incorporate such factors as “altruism” (Becker, 1981) or “distaste” for interaction with people of a different race (Becker, 1957).

One way to think about social preferences is that they establish reference points for assessing pay and work effort.<sup>27</sup> Importantly, under social preferences, reference points emerge endogenously as a result of the choices made by employers and employees.

#### *4.1. Pay Status: Financial Incentives and Inequality Aversion Within Firms*

People dislike inequality—especially when they have drawn the short straw. Indeed, there is substantial indirect evidence that wellbeing is shaped in large measure by comparisons with others.<sup>28</sup> This idea can matter within organizations because people are likely to compare themselves with others around them in the workplace. In turn this can be an important determinant in shaping firm compensation policies.

The idea that interpersonal comparisons matter to agents can easily be captured by including “asymmetric inequality aversion” into utility functions. Utility is increasing in one’s own income, of course, but decreasing in the income of other relatively-wealthier comparison individuals. The “asymmetry” refers to a presumption that agents suffer more from inequality that is to their material disadvantage than they gain from inequality that is to their material advantage (see, e.g., Fehr and Schmidt, 1999, and Encinosa, Gaynor and Rebitzer, 2007).<sup>29</sup>

---

<sup>27</sup>Assessments based on reference points play an important role in behavioral economics generally, included features germane to labor economics. Kahneman and Tversky’s “prospect theory” of decision making under uncertainty argues that individuals are loss averse and that they calculate gains and losses relative to (potentially manipulable) reference points (Rabin and Thaler, 2001). Sometimes reference points are determined by the status quo or by inertia (e.g., Thaler and Sunstein, 2008, and Genesove and Mayer). Reference points also play a role in behavioral models of labor supply. Camerer, Babcock, Loewenstein, and Thaler (1997), for instance, argue that the labor supply of taxi drivers is best understood if the drivers evaluate their daily income relative to a daily target. (See also Farber, 2005 and 2008, for additional evidence, some of it to the contrary, and DellaVigna, 2009, for a clarifying discussion.) Fehr and Goette (2007) provide evidence for a field experiment suggesting that loss aversion and reference points may be important in determining work intensity.

<sup>28</sup>This is the basis, for example, of the well-known “Easterlin paradox”—the paradoxical results that (i) individuals who are low in a nation’s income distribution report themselves to be unhappy, (ii) the average level of unhappiness does not much vary across nations with different levels of aggregate income, and (iii) countries do not get much happier as they get richer. In short, while there is evidence to suggest that absolute income matter for happiness (e.g., Stevenson and Wolfers, 2008), it seems clear that one’s standing on the economic totem pole matters as well. Frank’s (1985) well-known book provides an interesting and wide-ranging discussion on the human inclination for social comparison.

<sup>29</sup>An assertion of utility based on inequality aversion represents a “stripped down” way of characterizing the behavioral phenomena under study. Specifically, this treatment makes no distinction about agent attributions concerning the nature of inequality (e.g., what the inequality might say about the principal’s intentions or other agents’ intentions). We consider more sophisticated approaches below.

As an example of how asymmetric inequality aversion can affect our agency models, recall the tournament model, as set out in Section 2.2.2. In that model, workers had utility given by  $w - e$ , and made a net contribution of  $g(e) - w$ . The firm conditioned pay on an imperfect measure of  $e$ : it paid a base wage  $w_0$  to all workers, and in addition gave a bonus  $b$  to the fraction  $(1 - P)$  of workers who had the highest observed performance. (The “bonus” in this case would typically be a promotion to a higher-paying job.) Given worker’s best-response effort choices,  $\hat{e}(b)$ , we saw that the firm’s first-order condition was

$$(25) \quad [g'(\hat{e}(b^*)) - 1]\hat{e}'(b^*) = 0,$$

and given that  $e'(b) > 0$ , we had the efficient outcome,  $g'(e^*) = 1$ .

Suppose we take that same model but now introduce asymmetric inequality aversion by letting

$$(26) \quad \text{utility} = \begin{cases} w_0 + b - \delta_W b - e & \text{for “winners,” and} \\ w_0 - \delta_L b - e & \text{for “losers.”} \end{cases}$$

Here,  $\delta_W \geq 0$  reflects the possibility that winners feel empathy for losers, proportional to the inequality generated (but of course people do like to win, so  $\delta_W < 1$ ).  $\delta_L > \delta_W$  reflects the fact that workers who do *not* win the bonus suffer an even large utility loss due to inequality aversion. Repeating steps outlined in Section 2.2.2, we can show that the principal’s solution now has a first-order condition

$$(27) \quad [g'(\hat{e}(b^{**})) - 1]\hat{e}'(b^{**}) = (\delta_L - \delta_W)P > 0,$$

and therefore we have  $g'(e^{**}) > 1$ , which in turn means that the firm here settles for a second-best effort level,  $e^{**} < e^*$ .

Inequality aversion causes the incentive pay parameter to do the “double duty” of eliciting work effort and determining the extent of expected pay inequality in the firm. As a result the firm must compromise along an important dimension by lowering incentive pay,  $b$ , and reducing the effort level elicited from workers. Equation (??) also shows that in this setting the firm will want to be careful how it sets its promotion rate. Here the firm would like (all else equal) to set  $P$  near 0, which would allow effort to approach first-best. Intuitively, the

cost to the firm of inequality is lowest when there are relatively few people who are affected by the inequality, i.e., when the promotion rate,  $1 - P$ , is close to 1.<sup>30</sup>

The logic of our model of income comparisons underlies Frank’s (1984) seminal article on inequality aversion in labor markets. In his treatment, an employee gains in utility from being high in the firm’s pay hierarchy and loses utility from having a low position. Just as in our model, these concerns cause firms to operate with lower-powered incentives. Frank presents evidence drawn from many different types of organizations; he sees, for example, a dampening effect in commissions paid to car salesman and realtors, pay compression among college professors.

Encinosa, Gaynor and Rebitzer (2007) present a similar analysis of these issues in the context of incentive pay within medical partnerships. In these professional organizations, physicians determine incentive intensity by choosing how broadly they wish to share the income they generate with others in the practice. For example, groups often choose to share income equally across physicians—in a practice with  $n$  physicians, each physician keeps  $1/n$  of profits—which minimizes inequality. The practice of equal sharing rules has the potential disadvantage, though, of offering the lowest possible level of incentive intensity to partners. The model of inequality aversion set out by Encinosa, *et al.* shows that the tension between these forces makes sharing rules less attractive in large partnerships than in smaller partnerships—a result supported by the available data. The authors also present evidence consistent the notion that physicians compare effort as well as income. We take up effort comparisons in the next section.<sup>31</sup>

As a second example of the potential impact of inequality aversion in a principal agent setting, consider the multi-tasking model we examined in Section 3.3. Recall that in that

---

<sup>30</sup>Of course, issues of limited liability might limit the size of  $b$ , and in turn prevent the firm from setting  $P$  as low as it would otherwise prefer.

<sup>31</sup>As for tests of inequality aversion, there is a challenge in distinguishing between muted incentives arising from equity concerns and muted incentives due to risk aversion. For example, if risk averse agents contract with a risk neutral firm, efficiency requires that the firm absorb some of the natural variation in income generated by the environment. This occurs even at the expense of forgoing a bit of extra work intensity. (This idea—concerning long-term implicit contracts—is an old one in labor economics. See, e.g., Bailly, 1974.) Analogously, inequality averse employees suffer asymmetric losses: the disutility of having realized pay  $X$  below their reference group is greater than the gain from having pay  $X$  above their reference group level. Thus firms can profit by offering “inequality” insurance in the form of lower pay and lower-powered financial incentives.

model, we assumed an agent utility function is  $w - \frac{1}{2}(e_1 + e_2 - e_B)^2$ , and we assumed further that the principal had a good signal for  $e_1$  but not  $e_2$ . So when the principal's payoff is  $e_1 e_2$ , the best the principal can do is pay the agent a fixed wage  $w$  that meets the participation constraint (i.e., so that  $w = v$ ) and then direct the agent (nicely!) to allocate effort so that  $e_1 = e_2 = \frac{e_B}{2}$ . Now suppose that the agent is inequality averse, just as in (??), but that in this case her comparison is the *principal's* income,  $\pi$ . So for the agent, utility is

$$(28) \quad w - \frac{1}{2}(e_1 + e_2 - e_B)^2 - \delta(\pi - w), \quad \text{if } w < \pi,$$

where  $\delta > 0$  reflects the extent to which the agent is inequality averse. The important point here is that the agent can always costlessly enforce perfect equality by simply adjusting effort allocation (keeping total effort at  $e_B$ ); she can reduce the principal's profits, while causing no harm to herself. If principal sets the wage to  $v$ , the agent will adjust effort so that  $\pi$  also equals  $v$ . So the principal typically finds it profitable to increase  $w$  above the participation; the principal uses "rent sharing." Reducing wages to the level of the employee's outside option would be self-defeating in this context because it risks that the agent will become disgruntled and take steps to "even the score."

The idea that inequity aversion supports rent sharing has been extensively explored in laboratory and field experiments involving variations on the Ultimatum Game. In this bargaining game a proposer offers to divide a fixed amount of money between himself and a responder. The responder can accept or reject this offer. If the offer is accepted, the money is divided according to the offer. If the offer is rejected, however, neither the responder nor the proposer gets any money. The conventional game theoretic solution to this bargaining problem for entirely selfish players is for the proposer to make an offer in which he keeps all, or nearly all, the money while the responder accepts any offer.

It turns out, however, that the participants in these games don't behave as entirely selfish players. Proposers routinely make offers close to an equal division of the pie and responders routinely reject low offers. These anomalies can be resolved, of course, by introducing equity concerns into players' utility functions. Inequality averse proposers get less utility than entirely selfish players do by proposing a division of the pie that greatly favors them. Conversely, inequality averse responders can credibly threaten to destroy the surplus if

highly unequal divisions are proposed. In equilibrium, reasonably egalitarian offers are made and accepted.

Most employment relationships exist in the context of competitive labor markets. Thus it is not sufficient to demonstrate that individuals prefer more equitable pay practices. Economists must also establish that these preferences matter for equilibria that emerge in labor markets. Fehr and Schmidt (1999) examine this central issue by considering whether the egalitarian rent sharing observed in the Ultimatum Game survives in an environment in which there are many proposers and the responder must accept or reject the best offer received. They find that competition between proposers leads to less egalitarian outcomes. To see the logic, consider a situation in which  $n$  proposers each offered a 50 percent to a responder, leaving each proposer with a  $1/n$  chance of having his offer accepted. An individual proposer could clearly do better by simply offering a 51 percent share to the responder, thereby insuring that his offer was selected (the probability increases from  $1/n$  to 1). But all proposers are driven by this same logic, and in the end, in equilibrium at least two proposers will offer the responder 100 percent shares. This, of course, is the outcome we would observe if proposers had no inequality aversion.

The irrelevance of inequality aversion stems from the fact that with many competitors, no single player can prevent an inequitable outcome. If no individual action can reduce the inequality of the equilibrium outcome, then inequality aversion cannot an important determinant of behavior. Fehr and Schmidt conclude that things would be different if individual players had a way to impose a cost on the counterparty to a highly unequal offer (as does the agent in the example we consider with equation (30)). Specifically,

... competition renders fairness considerations irrelevant if and only if none of the competing players can punish the monopolist by destroying some of the surplus and enforcing a more equitable outcome. This suggests that fairness plays a smaller role in most markets for goods than in labor markets. This follows from the fact that, in addition to the rejection of low wage offers, workers have some discretion over their work effort. By varying their effort, they can exert a direct impact on the relative material payoff of the employer (Fehr and Schmidt, 1999).

In short, agency problems of the sort depicted in our model above (that allows equity concerns), can survive market competition.

The Fehr-Schmidt conjecture has been examined experimentally by Fischbacher, Fong and Fehr (2008).<sup>32</sup> They find that increasing proposer competition in the Ultimatum Game, by adding extra proposers, causes a large increase in mean accepted offers. Similarly, increasing responder competition causes a reduction in mean accepted offers. Both of these results suggest that competition undercuts the influence of equity norms on bargaining outcomes, although in each case the increase in inequality of outcomes is less than one would predict on the basis of competition alone.

As for the idea that workers will exercise retribution when treated unfairly, Bewley's (1998) qualitative interview study indicates that managers and other labor market participants believe that there is a connection between morale and performance. Bewley focuses on why employers do not cut wages in recessions. He found that employers averse to cutting wages because of the fear of hurting morale, and causing a backlash from employees.

Empirical evidence of actual employee backlash, of the sort described by participants in Bewley's surveys, can be found in recent work by Mas and co-authors. In a remarkable study, Mas (2006) finds that when New Jersey police officers lose in final offer arbitration, so that the wage they receive is lower than the requested wage, arrest rates and average sentence lengths decline, while crime reports rise. This evidence is consistent with the idea that workers are less inclined to provide effort when the wage falls below a salient reference wage. Krueger and Mas (2004) report evidence that a long and contentious strike and the hiring of replacement workers in a Bridgestone/Firestone plant contributed to the production of defective tires. Mas (2008) finds Caterpillar plants that underwent contract disputes experienced reduced workmanship and reduced product quality. In this latter study he estimates that the contract dispute was associated with at least \$400 million in lost service flows due to inferior quality.

The papers cited in the previous paragraph study workers in unionized environments. The presence of a union likely facilitates the sort of collective retaliation that punishes employers who take morale lowering actions. The central idea—perceptions of unfairness can damage effort—may well be important more generally in labor markets. Evidence along these lines appears, for example, in our field experiment with Nagin and Sanders (Nagin,

---

<sup>32</sup>That paper also provides reference to a large relevant experimental literature more generally.

*et al.*, 2002), which manipulated monitoring levels at call centers collecting donations for charitable causes. In that study, conducted in a non-union environment, we had access to a direct measure of malfeasance on the part of individual employees as well as direct measures of individual employee perceptions about the employer. The measure of employee malfeasance is the rate at which employees artificially inflated their level of sales in order to earn extra commissions at the expense of the firm. The employer could catch some, but not all, of this activity through costly monitoring of a random sample of calls. When employees worked in centers with very low apparent rates of monitoring, opportunistic behavior increased. However, this increase was observed only for a subset of employees, and, importantly, increased opportunism was most prevalent among among workers who had expressed feelings that the employer treated them unfairly, did not care about them, and provided a bad place to work.<sup>33</sup> Employees who perceived themselves to be unfairly treated struck back at the employer (and added to their own income) when the opportunity to do so arose.

Inequality aversion on the part of employees has a number of interesting ancillary predictions about the way employment relationships are organized. Firms that rely on hiring employees in both very highly compensated occupations and less highly compensated occupations must go to some length to be sure that employees in the low-wage occupation do not include the high-wage occupation in their reference group. Failure to make this separation can lead to pressures to either pay employees in the low-wage occupation too much or employees in the high-wage occupation too little. Indeed, it would not be at all surprising to see firms choosing to choose outsourcing to other firms to avoid just these sorts of invidious comparisons.

Similarly, if employees respond to perceived inequities by retaliating along important, but hard to monitor dimensions of work effort and quality, firms that engage in highly unequal pay practices ought to seek out ways to reduce the perceived level of inequality. Secrecy regarding pay is a common human resource practice and it obviously makes invidious pay comparisons more difficult. Some companies, such as Walmart and Lincoln Electric,

---

<sup>33</sup>These attitude questions were collected in an anonymous survey of employees conducted before the field study began.



famously go to great lengths to discourage ostentatious executive perks and the depressing effects on morale they might encourage.

For publicly traded corporations the compensation of top executives is a matter of public record. In practice, however, these companies adopt complex and opaque compensation practices that make it difficult to understand exactly how much and in what ways top executives are paid. Bebchuck and Fried (2003) argue that anomalous features of executive compensation—such as the reliance on “at the money” stock options rather than stock grants—are best understood as efforts to camouflage pay and so avoid “outrage” from shareholders, employees, regulators and other interested parties.

Levy and Temin (2007) examine these outrage costs from an institutional and historical perspective. They argue that the Federal government enforced a set of informal yet egalitarian social norms on executive pay from the post World War II era through the early 1970’s. These norms were part of a larger set of institutional arrangements that included powerful unions, high minimum wages, and high marginal tax rates for high earners. With actions taken by Reagan administration in the 1980s (notably the firing of the air-traffic controllers), the federal government signaled that it was leaving such decisions as CEO pay strictly to market forces. The degree to which income norms can be shaped by national institutions and economic policy is an important question that would benefit from more rigorous empirical and historical research.

#### *4.2. Effort Norms*

Fehr and Schmidt’s key idea is that equity concerns constrain firm behavior even in competitive labor markets, because of behavioral features in agency problems inherent in employment relationships. If, as appears to be the case, employees respond to unequal or unfair treatment by taking actions that punish the employer, it is a short further step to presume that morale enhancing activities ought to motivate employees to take actions in the interest of their employer. This is the premise of Akerlof’s (1984) gift exchange model of efficiency wages.

Akerlof’s model plays a central role in behavioral labor economics because it relies on very different sociological and psychological mechanisms than the standard agency approach presented in Section 2 above. Instead of engaging in calculations about the costs and benefits

of working harder, employees in Akerlof's model are motivated by norms governing behavior in the exchange of gifts. If the employer pays employees a wage higher than some reference wage, the employee perceives himself to have received a gift from the employer. This gift creates an obligation to give something valuable in return. In the employment context, the obvious way to reciprocate is to provide the firm with more than the minimally acceptable level of work effort and attention.

Akerlof's approach to the problem of agency rests critically on the concept of effort norms, i.e., on the idea that individuals are motivated to provide effort in ways that enable them to conform to their self image or social identity. Decent people, so the reasoning might go, return kindness with kindness and so, wishing to preserve the self image of decency, the employee responds to a high wage by returning the favor in the form of high effort to the employer.<sup>34</sup>

Experimental investigations suggest that reciprocity of the sort identified in Akerlof can survive in competitive environments. For example, Fehr, *et al.* (1998) report results from a laboratory experiment in which sellers have the opportunity to select quality levels above the minimum level enforceable by buyers. In treatments where sellers have the opportunity to do so, they reciprocate high prices with high quality levels. Anticipating this behavior, buyers profit by offering high prices far in excess of the seller's reservation prices. In treatments where sellers do not have the opportunity to reciprocate, buyers offer lower prices.

If employee effort responds to the perceived "fairness" of wage offers, then policy makers must pay special attention to policies that might shift perceptions of the "fairness" of a wage offer. Policy may be especially likely to affect fairness if individuals care about employer intentions as well as outcomes. A wage of  $X$  in the absence of a minimum wage might be perceived to be quite fair because employers could have offered a good deal less but chose not to. If, on the other hand, the minimum wage was set to be  $X$ , then the employer might have to offer a wage above  $X$  to demonstrate their good intentions.

---

<sup>34</sup>This idea is modeled in an insightful way in an important paper by Rabin (1993). Charness and Rabin (2002) provide a clear statement of the ideas, and give reference to subsequent literature. They also give compelling evidence from laboratory experiments on reciprocity. For a discussion of additional experimental evidence, see the chapter by Kuhn and Charness in this *Handbook*.

In an important paper, Falk, Fehr and Zehnder (2006) investigate whether minimum wage laws influence the perceived fairness of wage offers. They set up an experimental labor market in which individual employees (students paid to participate in the experiment) have to decide whether or not to accept a job offered by a firm. Contrary to the conventional self-interest model, but consistent with a fairness-concerns model, individuals had reservation wages significantly above zero. There was also considerable heterogeneity in reservation wages, giving individual firms an upwardly sloping supply curve for labor. Introducing a minimum wage in this labor market had the effect of increasing individual reservation wages considerably—a result consistent with the hypothesis that the perceived intentions of the firm matter in determining fairness. Surprisingly, there appeared to be hysteresis in the effect of minimum wage laws on reservation wages: subjects exposed to the laws after participating in labor markets with no minimum wage laws increased their reservation wages, but subjects who first participated in labor markets with minimum wages did not revise reservation wages downwards when the laws were removed. These results, if they hold outside the laboratory, have implications that extend far beyond the issue of minimum wage laws. Labor market regulations that influence employer scope of action must take into account how these regulations are likely to affect employee perceptions of employer intentions. More provocatively, the hysteresis result also raises the possibility that regulators might not be able to “undo” the effects of policy simply by reversing previous decisions.

In strong form, well-functioning norms can have considerable social value generally. They can serve to reduce the problems created by agency in many contexts, including employment relationships within firms.

Considering their considerable economic value, it is important to understand the social processes that generate and sustain socially valuable effort norms. In a path-breaking economic analysis, Frank (1988) emphasized the role of emotions in resolving the “commitment problem,” i.e., the problem of eliciting a commitment to constructive cooperation. He argues that rational calculation is often not sufficient to sustain cooperation because by the time the misbehavior occurs, the benefit of punishing the bad actor has often already passed. Emotions, in contrast, can be the foundation of much more powerful sanctions because the commitment to follow through on the action is rooted in the primitive reward structure of

the brain. Thus, “cross me and you’ll never work in this town again” is a weak deterrent when uttered by a rational calculator who may decide after the fact that it is not worth the effort to punish the double-crosser. It is a strong deterrent, however, when uttered by someone who gets visceral satisfaction in carrying out his threat regardless of the cost to himself.

From a psychological perspective, emotions emerge from a genetically determined neurological reward system. The triggers of this reward system, however, are shaped by an intense and costly socialization process that trains individuals to have a “conscience,” i.e., to feel strong emotions when they lie, cheat or otherwise disappoint others expectations. Evolutionary considerations led Frank to expect that these efforts at socialization will not be entirely successful. Society will be composed of mixture of types: opportunists who take advantage of chances to free-ride and reciprocators who will devote resources to monitoring the behavior of their counterparties and cooperate so long as they perceive others doing the same.<sup>35</sup>

The idea that populations contain a mixture of opportunists and reciprocators is supported by experiments involving games involving contributions to public goods. Fehr and Gaetcher (2000) study such games and contrast two treatments. In the “no punishment” treatment, anonymous individuals are randomly allocated to groups of four and are given the opportunity to make contributions towards a public good. Payoffs are such that the dominant strategy is to make no contributions towards a public good. In the “punishment” treatment a second stage is added which gives each individual the opportunity to punish others by subtracting from their payout. Punishing poor contributors is costly, however, and no one interested in maximizing their monetary payoff will choose to punish after the damage is done.<sup>36</sup> For this reason one would expect the dominant strategy to be one of

---

<sup>35</sup>In a population entirely composed of cooperators there will be little reason to devote resources to monitoring the actions of others. This is an environment in which opportunists will thrive. Conversely in an environment with many opportunists, cooperators will enjoy an advantage so long as they devote resources to monitoring their counterparts. Evolutionarily stable equilibria will therefore involve a mix of opportunists and cooperators with the latter spending resources seeking to weed out or punish the former. See, e.g., Gintis, *et al.* (2003).

<sup>36</sup>The fact that some individuals will punish opportunists even when it is not in their direct material interest suggests that punishment is supported by psychological reward mechanisms rather than rational calculation. Consistent with this view, recent brain imaging studies taken during an economic experiment involving trust and retaliation suggest that punishment of individuals who violate trust activates a brain

“no contribution” in both punishment and no-punishment treatments. This prediction is confirmed in the “no punishment” game: average contributions converge to almost complete free-riding. In contrast, in games with the option to punish in the second stage, individuals do make substantial contributions to the public good and these contributions do not fall over time. Consistent with our effort norms model, subjects are more heavily punished the more his or her contribution falls below the average contribution of other group members. Individuals also exhibit heterogeneous tendencies to free-ride and punish. Depending on the definition, the authors estimate that between 20 and 53 percent of subjects in their study were free riders. Heterogeneity in the tendency to behave opportunistically has important implications for labor markets and personnel practices that we explore further in Section 5.

Emotions can support pro-social behavior in ways other than sustaining irrationally high levels of retaliation against defectors. Eckman (2001), for example, argues that emotional states can be read from the facial expressions of individuals. It follows from this that lying and other opportunistic activities that can elicit strong emotions are harder to sustain during face to face interactions. Valley, *et al.* (1998) investigate this hypothesis in a bargaining experiment which requires negotiators to elicit private information about the true value of an underlying asset when the incentives in the experiment do not support revealing this information truthfully. The study finds that face to face negotiations are more likely to reach efficient solutions than negotiations conducted over the phone or in writing.

The feelings of guilt and shame that support truth telling and honesty are similar to the emotions that support effort norms, and these emotions are generally thought to be strengthened by physical proximity and face to face interactions (Sally, 2002). A nice laboratory experiment by Falk and Ichino (2006) provides evidence along these lines. In particular, in that study the authors observed “peer effects” in subjects who would otherwise have provided low effort were motivated to increase effort when physically paired with high-productivity workers.<sup>37</sup>

---

region, the Caudate, that is involved in actions motivated by anticipated rewards (de Quervain, *et al.*, 2004). High Caudate activation likely reflects the anticipated satisfaction from punishing defectors.

<sup>37</sup>On net, Falk and Ichino estimate a positive impact on output due to these peer effects similar to estimate of peer effects on absenteeism behavior found in Ichino and Maggi’s (2000) study of workers in different branches of a large Italian bank.

In a remarkable study of cashiers at a national supermarket chain, Mas and Moretti (2008) find that substituting a worker with below average productivity with a worker with above average productivity is associated with a 1 percent increase in the effort of other workers on the same shift. Low productivity workers are especially responsive to the composition of their co-workers and this peer effect occurs only for low productivity workers who are in the line of vision of the high productivity workers. The effect of high output peers on the productivity of others declines with distance and with the frequency of interaction as measured by the degree to which shifts overlap.

If effort norms indeed require close proximity and frequent interactions within a work group, it is natural to ask whether these motives can operate in large organizations. Very little empirical work has focused on this important issue.<sup>38</sup> A closely related question is whether these norms can persist in cross-functional work groups where the social distance between members of the group might be large. Cross functional work groups composed of employees with widely different levels of income, status and education play an important role in our health care system where it would not be unusual to have surgeons, nurses and high school graduate technicians working together to improve processes. The failure of individuals to operate successfully in these settings likely contributes to inefficiencies in our health care system (as described, e.g., in Cebul, *et al.*, 2008, and other papers referenced therein).

As for implications for labor markets generally, a small but potentially important literature looks at the broader implications of gift exchange relationships. As Akerlof and Yellon argued in their 1990 paper, if the reference wage in a gift exchange sits at or somewhat above the reservation wage, a natural consequence can be unemployment and labor market segmentation. The central logic is much the same as in the Shapiro-Stiglitz efficiency wage model discussed in Section 2.2.3 above.

---

<sup>38</sup>An exception is Knez and Simester (2001). This case study of Continental Airlines finds evidence of the apparent success a firm-wide incentive scheme that paid out a modest sum of money to almost all employees if airline's aggregate on-time departure statistics cleared a certain threshold. The authors argue that effort norms, enforced by the relatively small and homogeneous ramp and ground crews at each airport, could and did augment the low-powered financial incentives inherent in the bonus plan.

[Need two or three paragraphs on aggregate labor market consequences of “gift exchange” models, perhaps drawing on the work of Fehr, Goette, Zehnder, and others. Bewley. What can these models tell us about wage rigidity?]

### 4.3. Professional Norms

In such professions as law and medicine, the principal agent problem takes on a special importance. Professionals are in theory the agents of their clients, but professionals enjoy advantages of education, credentials, status and specialized knowledge that make their clients especially vulnerable to exploitation. In order to protect clients from abuse, professions go to great lengths to inculcate norms of professional conduct. This makes professions an especially important venue for analyzing the effect of norms.

In our analysis of physician incentives in a managed care organization with Martin Gaynor (Gaynor, Rebitzer and Taylor, 2004), we develop a simple model of professional norms that we adapt here. The model follows the approach used throughout this essay—we modify the agent’s utility function, in this instance to include physicians’ regard for their patients. In this case, we posit that the socialization of physicians causes them to experience disutility when the practice style they deliver medical services are lower than the level that the patient would select for themselves (if they were as well informed as the physician). Think of this level of services,  $m_B$ , as the level (measured here in dollars) that results when the physician incorporates a patient’s own preferences into his utility function.

We write the utility of a physician treating  $i = 1, \dots, n$  patients as a function of income earned,  $w$ , the deviation of medical services,  $m_i$  from the ideal level:

$$(29) \quad w + \sum_{i=1}^n \mu_i d(m_i - m_B^i),$$

where  $d(\ )$  is a convex function that achieves a maximum when  $m_i$  is equal to a subjective “best” levels of care,  $m_B^i$ , and the  $\mu_i$  parameters indicate the weight the physician places on each patient’s well-being. Thus physician utility is increasing in both income and services provided when they adopt a practice style with  $m_i < m_B^i$ . In a well designed fee-for-service environment where insurers don’t try to “manage” the care physicians provide, one would expect physicians to deliver care at or close to  $m_B^i$ .

Managed care organizations, such as Health Maintenance Organizations (HMOs), often try to influence physician practice styles through the use of financial incentives. The managed care organization we studied, for example, adopted a simple policy of extrinsic motivation designed to restrict utilization without substantially harming patients: the principal (HMO) offered agents (physicians) a bonus  $b$  if total annual medical expenditures fell below a target  $\bar{m}$ . The probability that a physician's expenditures on behalf of patients fell below this threshold depended of course on the decisions made on behalf of each patient and also luck (because of the impossibility of anticipating all possible expenditures), so the probability of earning the bonus, given expenditures  $m_i$  and target  $\bar{m}$ , is given by the c.d.f.  $F(\bar{m} - \sum m_i)$ .

Giving the physician responsibility for the allocation of resources to a panel of patients in this way makes sense when physicians have practice norms of the sort characterize in (??). We can illustrate the idea easily with the case in which disutility is  $d(m_i - m_B^i) = -\frac{1}{2}(m_i - m_B^i)^2$ . Then the physician's best response to a policy,  $b$  and  $\bar{m}$  is found by maximizing

$$(30) \quad bF(\bar{m} - \sum m_i) - \frac{\mu_i}{2}(m_i - m_B^i)^2.$$

This leads to the best response function for the treatment of each patient ( $i = 1, \dots, n$ ):

$$(31) \quad \hat{m}_i(b) = m_B^i - \frac{b}{\mu_i} f(\bar{m} - \sum m_i).$$

The extrinsic reward induces the physician to conserve resources on behalf of the HMO, and if  $\mu_i$  is the same across all patients, say  $\mu$  (i.e., there is no favoritism), the physician does so in a sensible way.<sup>39</sup> Also, it is easy to confirm that if the second order condition holds, increases in the bonus increase the agent's inclination to reduce patients expenditures,

$$(32) \quad \hat{m}'_i(b) = \frac{f(\bar{m} - \sum m_i)}{bf'(\bar{m} - \sum m_i) - \mu_i} < 0.$$

From (??) and (??) we can see that the intrinsic value the physician places on patients (represented by  $\mu_i$ ) governs the level of expenditures chosen for patients, as well as the power of the explicit incentive to alter chosen expenditures.

In our empirical analysis of internal records in an HMO (Gaynor, Rebitzer, and Taylor, 2004), we found results consistent with the prediction in (??); increased incentive intensity

---

<sup>39</sup>Indeed, as we note below, the outcome is potentially efficient.



led physicians to adopt medical practices entailing lower costs. We found also that physicians appeared to cut costs most for outpatient and elective procedures and not at all for in-patient procedures. This suggests that physicians cut costs most in settings where the consequences for patient welfare were least likely to be severe.

Several implications merit mention. Society surely does want physicians, in their resource allocation capacity, to limit expenditures in a reasonable way, because ultimately we pay for those resources. This presumably is the role of HMOs. Fortunately, given the preferences we posit, HMOs have a reasonable way to do this. By rewarding the careful stewardship of resources by physicians, the HMO can improve efficiency.<sup>40</sup> There are clearly merits to the common practice of giving physicians substantial discretion over resource allocation to patients, as long as physicians at least partially internalize patients' preferences. This result hinges in an important way on the assumption that physicians place intrinsic value on patient well-being, and that intrinsic motivation is allocated roughly evenly across patients (i.e., physicians engage in no taste-based discrimination). Indeed, the viability of cost-containment incentives depends crucially on physicians' intrinsic motivations—motivations are doubtless driven by a variety of factors at an individual, organizational, and society level. Quite clearly, for excessively selfish physicians (those with low values of  $\mu$ ) cost containment in the form of extrinsic bonuses may be “too effective,” i.e., lead to undertreatment. By the same token, as we will discuss at length in Section 5, extrinsic rewards can undermine intrinsic norms. This issue would seem to be potentially germane in the debate of health care reform (much of which focuses on cost-containment incentives).

An additional concern arises to the extent that physicians' extrinsic incentives are shaped in ways that reward resource or effort allocation in dysfunctional ways, as when physicians

---

<sup>40</sup>Given that  $d(\cdot)$  in (??) reflects physician disutility due to harming patient utility, it is possible that a well-chosen level of  $b$  might result in an efficient allocation. This is most likely if patients in the pool have similar preferences and income, and therefore have roughly the same demand for medical treatment intensity. With income (or preference) heterogeneity, consumers differ in the extent to which they wish to make tradeoffs in the price of health care and the extent to which their physician limits medical expenditures. Cooper and Rebitzer (2006) construct a model in which HMOs compete for consumers on the basis of price and the size of physician networks they offer. Physicians operating under norms of the sort we described above prefer working with lower-powered cost-cutting incentives. Thus health plans wishing to offer consumers large physician networks must operate with lower-powered cost containment incentives. The resulting high costs require insurers to charge high premiums. Conversely, an HMO wishing to appeal to consumers preferring low premiums must impose more stringent incentives and therefore operate with smaller physician networks.

rewards are greater for the treatment of one patient as compared to another. Strong treatment norms (e.g., high values of  $\mu$  in our model) reduce the damage these incentives might create. More troublesome yet, a physician might face a situation in which the choice of treatment benefits the physician at the expense of a patient. Many professions have rules and regulations designed to guard against these conflicts, but refinements of some the behavioral economic ideas we have discussed so far suggest that regulating these conflicts might be far more difficult than would otherwise appear.

Dana and Loewenstein (2003) offer an especially insightful discussion of these issues in the context of gifts to physicians from pharmaceutical companies. Their discussion relies on the notion of “self serving biases.” Self serving biases occur when individuals evaluate an ambiguous set of facts in their favor.<sup>41</sup> Drug companies famously give gifts to physicians, and there has been considerable concern that these gifts bias physicians in favor of one or another pharmaceutical (Avorn, 2004). Recent guidelines therefore define gifts of more than nominal value as inappropriate but small gifts may not be. Dana and Loewenstein (2003) argue, however, that even small gifts trigger a norm of reciprocity that induces self serving biases in the interpretation of clinical trials and hence in the prescribing behavior of physicians. The claim that “gift exchange” induces feelings of reciprocity that bias the judgment of highly trained professionals has provocative implications for the many important economic settings where clients hire professionals to work as their agents in navigating complex and unfamiliar transactions.<sup>42</sup>

---

<sup>41</sup>For example, in the context of personnel policies, employees might compare their own compensation to that of their peers in the belief that fairness requires equal pay for equal performance. But performance can often be measured in multiple ways. If self serving biases operate, individuals choose the performance measure that most advantages themselves.

Along these same lines, Babcock, Wang and Loewenstein (1996) find that in collective bargaining between teachers unions and school districts, each side tended to chose for comparison, local districts for whom the comparison was most favorable to their interests. Similarly, Babcock, Loewenstein, Issacharoff and Camerer (1995) conduct a series of studies in which lawyers and law students were randomly assigned a role in a law suit (plaintiff and defendant) and were then asked to evaluate the facts of the case. Subjects consistently interpreted the facts of the case in favor of the role they were assigned, and these biases had an impact on their willingness to settle cases in simulated negotiations. This work establishes that it may be quite difficult to persuade participants that they are victims of self-serving biases. In short, such biases are *not* consciously adopted. (See Babcock and Loewenstein, 1997, for a nice discussion.)

<sup>42</sup>Moore and Loewenstein (2004) provide an insightful discussion of the psychological mechanisms that make conflict of interest a difficult problem.

#### 4.4. *Identity*

Our discussion of professional norms focused on the idea that physicians might experience disutility—perhaps profound feelings of discomfort or anxiety—if they deviate from proscribed behaviors with respect to their clients or patients. This approach to economic sociology is discussed at length in the work of Akerlof and Kranton (2000, 2005) on “identity.”

Here is the key idea:

The term *identity* is used to describe a person’s social category—a person is a man or a woman, a black or a white, a manager or a worker. The term identity is also used to describe a person’s self-image. It captures how people feel about themselves, as well as how those feeling depend upon their actions. In a model of utility, then, a person’s identity describes gains and losses in utility from behavior that conforms or departs from the norms for particular social categories in particular situations.

This concept of utility is a break with traditional economics, where utility functions are not situation-dependent, but fixed. In our conception, utility functions can change, because norms of appropriate and inappropriate behavior differ across space and time. Indeed, norms are taught—by parents, teachers, professors, priests, to name just a few. Psychologists say that people can *internalize* norms; the norms become their own and guide their behavior (Akerlof and Kranton, 2005).

The idea that “category” and “situation” can be fundamental elements in preferences enormously expands the range of principal agent models. For instance, to the extent that identity can be manipulated within an organization, identity-based incentives might substitute for extrinsic rewards.<sup>43</sup> Just as families and religious communities undertake important and costly investments to ensure that their children internalize a set of values and practices consistent with passing on the family or group’s social identity, so organizations might make investments in practices that persuade employees to adopt goals of the organization, and so mitigate agency problems. These investments are likely to be greatest where financial rewards are most costly to the organization, e.g. when performance measures are especially noisy and where high effort (or high effort at peak times) is critical to the organization’s success. Investment in identity incentives will also be greatest where inculcating identity is cheap, and it is reasonable to suppose that imparting identity is cheapest when agents are

---

<sup>43</sup>For example, practices at West Point are designed to “inculcate non-economic motives in the cadets to that they have the same goals as the U.S. Army” (Akerlof and Kranton, 2005), and firm or workgroup loyalty can be found more generally in many organizations.

young and/or when highly motivated individuals self-select into the organization—an issue to which we return in Section 5.

The great virtue of identity models is that they are highly flexible and therefore able to account for behavior that is anomalous from the perspective of simple agency models. This virtue is a curse, however, when it comes to generating falsifiable hypotheses for testing identity models themselves. One way around this problem might be to focus on a particular relevant social category and seek to understand key norms that can be studied systematically and characterized in a parsimonious way.

A template for this latter approach can be found in a series of careful and nuanced investigations of psychological factors that generate *gender* differences in economic behaviors. For example, work by Babcock and her co-authors, demonstrates a profound gender-based difference in the inclination to initiate negotiation; “women don’t ask.”<sup>44</sup> A simple and clear demonstration emerges in an experimental study in which subjects are asked to complete a simple task, and are then put in a position in which there is ambiguity with regard to the payment. In a typical experiment, subjects were told in advance that the payment would be between three and ten dollars. Then, at the conclusion of the session, the experimenter says, “Here’s three dollars. Is three dollars okay?” Eight times as many men as women asked for more money in this experiment. Even in a variant of the experiment in which the experimenter provides cues to signal the social acceptability of negotiation (e.g., with a prompt, “the exact payment is negotiable”), far more men than women take up the opportunity (Small, Gelfand, Babcock, and Bettman, 2007).

In laboratory and field experiments, this disinclination by women to “ask” affects outcomes in negotiated settlements, leading women to do less well than men. Importantly, though, when a woman advocates on behalf of *someone else*, she is typically more successful than when she negotiates for herself, and indeed is generally more effective than men in this capacity (Bowles, Babcock, and McGinn, 2006). Part of the reluctance to “ask,” it appears, comes from a desire to avoid self promotion.

---

<sup>44</sup>Babcock and Laschever (2003) provide an engaging and wide-ranging discussion. The authors present a great deal of real-world evidence about women’s general disinclination to ask, and they includes observations about the implications for gender inequality.

Along these same lines, Gneezy, Niederle, and Rustichini (2003) find that women respond differently than men to tournament style incentives when these contests involve both men and women. When paid by piece rate or when competing in single sex tournaments, women's performance is similar to those of men. Niederle and Vesterlund (2007) provide experimental evidence that in comparison to men, women generally shy away from incentives schemes that involve tournament competition.

*Gender identity*, in short, matters in economically important ways. It is tempting to assert, indeed, that female identity includes a component that guides women to shy away from competition with men and to reject self promotion. However, it is important to understand that this might not be the whole story, or even the most important part of the story, when explaining gender differences in behavior. There is considerable evidence in psychology that a "kinder, gentler image" is expected of women (to use the expression in Rudman and Glick's 1999 article on the topic). Women who violate that norm by engaging in self-promotion face the potential of backlash, which can entail psychological and material costs (as when a woman is bypassed for promotion because she is seen as "inappropriately assertive"). Thus, even a woman who feels no particular disinclination for self promotion might find it in her self interest to adopt the expected "kinder, gentler" norm (Bowles, Babcock, and Lai, 2006).<sup>45</sup>

Standard agency models, discussed in Sections 2 and 3, do have implications for gender in labor markets.<sup>46</sup> The recent work by Babcock, Niederle, Vesterlund, and their co-authors, discussed above, adds a new and promising perspective for understanding the role of gender in organizations and labor markets. Babcock and Lashever (2003) provide extensive evidence that women's reluctance to ask often includes an unwillingness to negotiate their own salaries. It follows logically that in labor markets in which there is rent sharing,

<sup>45</sup>The point here is that individuals are not passive carriers of their social identities and, as Akerlof and Kranton note, there are many instances in which identities are supported by sometimes severe social sanctions meted out to those whose behavior deviates from proscribed behaviors.

<sup>46</sup>For example, as Bulow and Summers (1984) note, if women have generally have lower labor market attachment than men (perhaps because they are more likely to withdraw from the market for bearing and raising children or elder care), efficiency wages will be a less effective in motivating women than in motivating men. This leads to an equilibrium in which a higher proportion of women than men will end up in the "secondary sector." As a second example, long-hours work norms that emerge in rat race models, such as those of Landers, Rebitzer, and Taylor (1996) and Gicheva (2009) might be particularly disadvantageous to women. See Landers, Rebitzer, and Taylor (1997) of a discussion of this latter issue.

this psychological phenomenon contributes to male-female wage and income gaps. On the other hand, the title to Neiderle and Vesterlund’s 2007 paper—“Do Women Shy Away from Competition? Do Men Compete Too Much?”—suggests an important point: Cooperation, and the willingness to work hard on the behalf of others, are valuable traits, which should receive positive value in the labor markets. An important research agenda going forward is the incorporation of new findings on gender from psychology and behavioral economics into models of organizations and labor market equilibrium for the purpose of investigating these very issues.<sup>47</sup>

There are certainly other important identity categories that deserve attention from by behavioral economists who study organizations and labor markets. *Ethnicity* and *sexual orientation* are additional identity categories that are important in many contexts, including, quite possibly, the labor market.<sup>48</sup> Berman’s (2000) economic analysis of ultra-orthodox Jewish groups indicates a strong behavioral impact of *religious identity*, which induces many Israeli ultra-orthodox men to engage in full-time yeshiva study into their early 40s, thereby impoverishing themselves and their families.

#### 4.5. *Miscommunication and Race*

Another identity category of indisputable importance is *race*. There is little theoretical work in economics that explores the role of race in organizational form and compensation

---

<sup>47</sup>Our focus on interactions within organizations follows from the focus of this essay—behavioral approaches to agency. Gender identity and changing norms are likely important for women’s labor supply as well. In an intriguing and ingenious study, Fernandez, Fogli, and Olivetti (2004), find that married women are more likely to work outside of the home if they are married to a man whose mother worked outside the home. A causal link is established by the use IV estimation, with cross-State variation in male World War II mobilization rates as the instrument.

<sup>48</sup>In the broad social sphere, Bisin and Verdier’s (2000) analysis of *ethnic* identity and intermarriage makes a strong prediction that if families value homogamous matches (matches between men and women in the same ethnic group), minority families will make greater investments in identity-preserving activities than majority families, because there is a greater chance that their children will enter heterogamous matches. In a field study of one workplace, Bandiera, Barankay, and Rasul (2006) document workplace favoritism based on nationality (presumably because of social connections between those who share language and national origin) that is costly to the firm. As for *sexual orientation*, it seems possible that when gay individuals take the (possibly very costly) break from powerful expectations to adopt a heterosexual identity and norm, this reduces costs for deviation from traditional norms along other dimensions, such as occupational choice. Along these lines, Black, *et al.* (2000) show that during the Korean War era (1950–1954), military service rates were 12 times higher for lesbian women than other women, and Black, Sanders, and Taylor (2007) show that lesbian college graduates sort into traditionally male majors at substantially higher rates than other women. We know of no work in economics that explores implications for organizations and labor markets.

practices. A very important exception is Lang’s (1986) “language theory” of statistical discrimination, which focuses squarely on agency and performance within organizations. The starting point of Lang’s analysis is the observation the misunderstanding and misinterpretation are common workplace problems. Lang draws on a wide body of literature in psychology and linguistics to argue that these problems are exacerbated when managers and workers are from different “cultural” or “linguistic” group than workers who are not.<sup>49</sup>

[Need to explain the Lang model, and provide an overview on recent evidence from the literature in psychology. Any work in behavioral economics? How does Austen-Smith and Fryer (2005) fit in?]

Following Lang’s lead, Ritter and Taylor (forthcoming) consider a labor market in which there is potential for race-based workplace misunderstandings. Their focus is the possibility that this force contributes to black-white gaps in unemployment. The model of unemployment is the agency-based efficiency wage model outlined in Section 2.2.3.

Suppose that most supervisors in the U.S. are white (perhaps because capital is disproportionately in the hands of whites in the U.S.), and that these managers are more successful evaluating the performance of white employees than black employees. Now recall that in the efficiency wage model, set out above,  $\sigma$  (the standard deviation of the noise) reflect the precision with which managers evaluate workers. The logic of Lang’s arguments leads to the conclusion that  $\sigma$  is relatively higher when white managers evaluate black workers. If so, the unemployment rate will be higher for blacks than for whites.

To see the logic of the Ritter-Taylor result, recall, from (??), that in an economy with homogenous workers, an efficiency wage strategy of worker motivation leads to the following relationship between the equilibrium wage ( $w^E$ ) and unemployment rate ( $u^E$ ):

$$(33) \quad w^E = e^* + \frac{1}{\phi(z^*)} \left( \rho + \frac{F^*}{u^E} \right) \sigma.$$

Now suppose that black and white workers are equally productive and thus in equilibrium must be paid the same wage. Suppose also that, as discussed in the last paragraph, there

---

<sup>49</sup>See also the excellent discussions by Bradford and Welch (1996). The idea that minority individuals might be more difficult to assess than non-minority workers is of course also at the root of the classic work on statistical discrimination. (See, e.g., Arrow, 1998, for references to earlier literature, and a thoughtful discussion.)

is more noise in the evaluation of black workers than white workers:  $\sigma^B > \sigma^W$ . Then the following must hold in equilibrium:

$$(34) \quad e^* + \frac{1}{\phi(z^*)} \left( \rho + \frac{F^*}{u^B} \right) \sigma^B = e^* + \frac{1}{\phi(z^*)} \left( \rho + \frac{F^*}{u^W} \right) \sigma^W,$$

where  $u^B$  is the unemployment rate for black workers and  $u^W$  is the unemployment rate for white individuals. Clearly,  $u^B > u^W$ . We therefore have a potential explanation for racial differences in unemployment rates.

This model is thus consistent with evidence, such as Neal's (2006), that among men, black-white gaps in the *wage* are small when one conditions on a measurement of human capital taken when the men were youths (the AFQT), but black-white gaps in *unemployment* are large. Ritter and Taylor (forthcoming) show that black-white unemployment gaps persist when one conditions on the AFQT, and that unemployment rates are highest for black men who attended high schools in which other students were mostly black. If these same men find on-the-job interactions with their boss difficult, this evidence would then be consistent with the model of racial differences in unemployment we have just outlined.<sup>50</sup>

## 5. DUAL-PURPOSE INCENTIVES: CAN PAY DESTROY INTRINSIC MOTIVATION?

In Section 3 we discussed dual-purpose incentives as they arise in conventional models of extrinsic rewards, noting that they arise in many contexts, including the use of compensation practices (i) to signal some otherwise-unobserved characteristic of the firm, (ii) to avoid adverse selection of workers along some otherwise-unobserved characteristic (as when firms use overwork as a means of selecting workers inclined toward hard work), and (iii) to deal with multi-tasking. We return in this section to the study of incentives that must do “double duty,” but now in the context of intrinsic motives.

---

<sup>50</sup>Grogger (2009) provides another piece of evidence consistent with the idea that impediments to black-white interactions to spill into the labor market. Even when he controls for skill and family background, blacks with speech patterns that sound distinctively black (according to anonymous listeners) are found to be relatively less successful in the labor markets.

Even so, the “language discrimination” story we have outline is likely a modest part of the profound racial divide in the U.S., as indicated by the black-white gap in unemployment and labor force participation, as well as many other economic and social dimensions.



Perhaps the most intriguing possibility here is that high-powered pay systems can destroy intrinsic motivation. As we will see, this possibility emerges in many cases that take behavioral factors into play.

### 5.1. *Pay and Selection on Dedication*

To set the stage, we begin with an extremely simple model in which increasing pay induces adverse selection along an important dimension—dedication to the job. Our case considers a group of workers who are qualified for a particular occupation that for many is a “calling” or “vocation.” There are many examples of dedication: religious ministry, policy advocacy, nursing, early childhood education, occupations that seek to alleviate poverty, public-interest law, etc. The “calling” in this case is a potentially important form of intrinsic motivation. In models that we consider the sub-sections that follow, this form of intrinsic motivation will be potentially shaped, at the individual level, by one’s environment. For example, an agent’s dedication to her job might be reinforced or eroded by the behavior of a principal or by the behavior of peers. Similarly, people might take pride in their dedication and wish to signal it to others. Here, though, we simply take a vocational inclination to be an unobserved feature of preferences.

Specifically, we consider the model set up by Heyes (2005) and analyzed further by one of us (Taylor, 2007). The analysis, which uses the market for nurses as the focus of discussion, begins with simple behavioral assumptions. There are  $L$  qualified nurses, each of whom falls into two categories: (i) A proportion  $\pi$  views nursing as a “vocation.” They provide high-quality nursing along an unobservable dimension, the value of which is given by  $q_H$ . They also find their work fulfilling, and thus earn money metric utility  $m$  beyond the earned wage  $w$ . (ii) The remaining proportion  $1 - \pi$  view nursing as simply a job. The value of their work is  $q_L < q_H$ , and they receive no utility beyond their pay from working as a nurse.

Each individual has a reservation wage  $r$  which is drawn from a log concave p.d.f.,  $f(r)$ , that has a corresponding c.d.f.,  $F(r)$ . The function  $f(\cdot)$  is assumed to be the same for both types of worker. Thus, at wage  $w$ , the quantity of nursing labor supplied is

$$(35) \quad \tilde{L}(w) = [\pi F(w + m) + (1 - \pi)F(w)]L,$$

and the average quality is of nursing care is

$$(36) \quad \tilde{q}(w) = \theta q_H + (1 - \theta)q_L,$$

where  $\theta$  is the proportion of employed nurses for whom nursing is a vocation, i.e.,

$$(37) \quad \theta = \frac{\pi H(w + m)}{\pi H(w + m) + (1 - \pi)H(w)}.$$

Heyes' key insight is that this latter proportion is declining in the wage.<sup>51</sup> Thus, the higher the wage, the lower will be the quality of services provided.

Consider an employer acting in isolation, e.g., a monopsonistic National Health Service (NHS), that hires nurses. (In different markets one could think of the Roman Catholic church setting wages for priests or nuns, or Habitat for Humanity setting wages for professional builders who take part-time positions constructing affordable housing.) Heyes shows that an employer who understands the adverse selection properties of high wages will set the wage to be lower than would otherwise be chosen. Thus, an NHS that maximizes surplus generated by nurses will operate with an apparent "shortage" of nursing, in the sense that the expected net value of product will be positive for the marginal nurse.

It is possible indeed that the principal will be driven to a corner solution, with pay set to zero. Thus, Habitat for Humanity has the well-known policy of using unpaid volunteers for many key tasks. Organization that seek to remedy injustice in the legal system often rely on *pro bono* attorney services. Historically, many religious workers take a "vow of poverty," accepting compensation at near-subsistence levels. The idea, of course, is that the lower the pay, the higher will be the dedication level of individuals willing to adopt the vocation.

The analysis in Taylor (2007) shows that a monopsonist that seeks to maximize surplus generated by workers will always set the wage *lower* than the socially efficient level. The reason is that the monopsonist fails to take account of the surplus generated to those individuals who view their work as a vocation. Because wages are too low, too few vocationally-oriented workers end up in an occupation in which they create the greatest social value.

On the other hand, a parallel analysis in Taylor (2007) shows that if the labor market is perfectly competitive, the equilibrium wage will be inefficiently *high*. To see how this

---

<sup>51</sup> To see that point, take the derivative of  $\theta$  with respect to  $w$ . The derivative has the same sign as  $\frac{H(w+m)}{h(w+m)} - \frac{H(w)}{h(w)}$ , which is negative for a log concave function.

happens, notice that under the assumption that all workers must be paid the same wage, social planner would want to maximize

$$(38) \quad \pi \left[ (q_H + m)F(w_H + m) + \int_{w_H + m}^{\bar{r}} r f(r) dr \right] + (1 - \pi) \left[ q_L F(w_L) + \int_{w_L}^{\bar{r}} r f(r) dr \right],$$

where  $r$  is taken to be the value of the worker in some other capacity (with  $\bar{r}$  being the highest value in the distribution). Maximization of (??) leads to the wage being set to be the weighted sum,

$$(39) \quad w^* = \hat{\theta} q_H + (1 - \hat{\theta}) q_L, \quad \text{with} \quad \hat{\theta} = \left[ \frac{\pi f(w + m)}{\pi f(w + m) + (1 - \pi) f(w)} \right].$$

Next notice that in a competitive market the wage instead will equal average productivity,

$$(40) \quad w^c = \tilde{\theta} q_H + (1 - \tilde{\theta}) q_L, \quad \text{with} \quad \tilde{\theta} = \left[ \frac{\pi F(w + m)}{\pi F(w + m) + (1 - \pi) F(w)} \right].$$

Using the property of log concavity given in footnote (??), we can compare wages in (??) and (??), finding that  $w^c > w^*$ . The problem with the competitive market is that each firm makes hiring decisions on the basis of *average* market productivity. A social planner would instead make decisions on the basis of the productivity of the *marginal* worker, i.e., would take account of the fact that as the wage increases in the market, the productivity of the marginal worker declines.

This simple model serves as a first illustration of an important point that reappears throughout this section of our essay: Pay policies can affect intrinsic motivation, often in surprising ways. Here, high pay *reduces* intrinsic motivation in a workforce in a particularly transparent way. Low-pay environments attract workers for whom the job is a vocation—workers who have an intrinsic inclination to provide high-quality service. The higher the pay, the greater will be the proportion of workers for whom the job is simply a job, i.e., workers who will provide lower-quality service.

A particularly interesting feature of this simple behavioral model is that markets can lead to wages being either too high or too low relative to an efficient benchmark, depending on the market's structure. To see the logic of this point, consider this question: “If you were a falsely convicted death-row inmate, would you rather be in a State in which you must rely on an organization that reviews cases using *pro bono* attorney services, or in a State that

purchases legal services on the competitive market?” In the State that relies on *pro bono* services, attorneys who work on death-penalty cases will be highly dedicated to justice, and will provide excellent legal aid, but that aid will be in short supply. In contrast, in a State that purchases legal services for death-row inmates, access to attorneys may be greater, but those attorneys will have a lower expected level of dedication. Our model shows that in the State that relies on *pro bono* attorneys, the wage is too low and the quality level too high relative to the efficient benchmark. But in the State that uses the competitive market, the wage is too high and the quality too low. We cannot know which is the socially preferred second-best outcome.

Recently, a number of papers have examined models in which agents differ in their level of intrinsic motivation. Delfgaauw and Dur (2007), for example, have a wage-posting model in which a monopsonist faces the same tension discussed above: the higher the posted wage, the higher the probability of filling a vacancy, but the lower the expected motivation level of workers who apply. In their model, workers have private information about their utility—information which they may wish to signal to or conceal from an employer. Besley and Ghatak (2005) and Delfgaauw and Dur (2008) study public sector employment under the assumption that some agents have a “public service motivation” that takes the form of intrinsic value derived from making a contribution to one’s organization.<sup>52</sup>

The model we have examined in this section omits, obviously, several relevant issues that merit further consideration. For instance, the set-up abstracts from the core problem of agency; workers are simply assumed to supply effort on the basis of their internal intrinsic values. Second, workers are assumed to be *steadfast*; their motivation is not affected by the actions of those around them. Thus, a worker who is inclined to provide high quality service is not de-motivated when she is surrounded by others who provide low quality service. In short, the analysis abstracts from “social preferences” of the sort discussed in Section 4. Third, the model does not take account of the possibility that a worker’s motivation can

---

<sup>52</sup>See also Prendergast (2007), who sets out a model in which there is variation in the degree to which agents care about the outcome of some action they might take, as when bureaucrats vary in the extent to which they have altruism and empathy for individuals they are serving. One example he develops concerns social workers hired to determine eligibility for public assistance programs. While a “client-serving ethic” is important for this occupation generally, that same trait may be an impediment for the bureaucratic task at hand.

be affected by attributions the agent might place on the intentions of the principal. It is easy to see that such attributions might be germane, though, in the case of the “service motivations” we have been discussing.<sup>53</sup> Finally, the worker’s intrinsic motivation ( $m$  is not undermined when pay for the job increases). We turn next to models that take up such issues.

### 5.2. *Social Preferences, Conformism, and the Principal’s Use of Extrinsic Rewards*

Recent work by Sliwka (2007) considers the question of agency in a model that draws on a “social preference framework,” i.e., allows for agents’ motivation to be shaped in part on the behavior of those around them. As in the model set out in the last section, there is heterogeneity in worker motivation. There are two type of steadfast agents. One type is “strictly selfish;” these are agents who care only about their own payoff. The second type is “fair” in the sense that they care about the well-being of others; specifically, their utility is increasing in the principal’s payoff. The key innovation is to assume that there is yet a third group, “conformists,” whose inclination toward fairness depends the values of those around them. To keep matters simple, Sliwka assumes that when a conformist learns what agent type is in the majority, the conformist behaves like the majority-type agent.

In this set-up, a principal who understands that his workforce is comprised predominantly of fair agents might be able to use compensation policies as a credible *signal* to his workers, inducing “conformists” to believe that others around them are “fair.”

To see how this works, we set up a simple example similar to that developed by Sliwka. Suppose the principal’s profit earned by employing an agent is that agent’s output, which just equals the agent’s effort  $e$ , minus compensation,  $w(e)$ , which is written to be a function of  $e$  (as effort is assumed to be observable here, and thus can be used to condition the wage). The principal wishes to maximize profit,

$$(41) \quad \pi = e - w(e).$$

---

<sup>53</sup>For example, when an organization hires motivated agents to pursue some jointly-shared social goal, agents must believe that they indeed are advancing that goal. Presumably, religiously-oriented individuals will be demoralized if they discover that are working for a corrupt church. See, e.g., Besley and Ghatak (2005) for a discussion along these lines.

An agent's money-metric utility is increasing in the wage but decreasing in effort, and, for a fair worker depends also on the principal's profit as follows:

$$(42) \quad \text{utility} = \begin{cases} w(e) - \frac{e^2}{2} & \text{for a steadfast selfish agent, and} \\ w(e) - \frac{e^2}{2} + \mu\pi & \text{for a steadfast fair agent,} \end{cases}$$

where  $\mu$  reflects a fair agent's level of identification with the principal's objective (with  $0 < \mu < 1$ ). Finally, suppose that the principal sets compensation to have a fixed component and a "bonus" that is a linear function of effort,  $w(e) = w_0 + \beta e$ . It is easy to see that for a given announced compensation policy, an agent's best response is

$$(43) \quad \hat{e}(\beta) = \begin{cases} \beta & \text{for a steadfast selfish agent, and} \\ (1 - \mu)\beta + \mu & \text{for a steadfast fair agent.} \end{cases}$$

With all this in mind, consider a firm that must post a wage-bonus policy that pertains to all workers. Clearly, effort is increasing in the incentive intensity,  $\hat{e}'(\beta) > 0$  for both types of steadfast agent. Sufficiently high effort-contingent bonuses would seem to be in order.<sup>54</sup> Remarkably, it might nonetheless be in the principal's best interest instead to set  $\beta$  equal to 0 and increase the baseline wage, i.e., to rely solely on low-powered incentives.

The key is the emergence of a separating equilibrium in which conformists become convinced that most steadfast agents are *fair*. Recall that the firm has private information about the proportion of steadfast agents who belong to each type, which for simplicity is taken to have a *low* value or a *high* value. It is a matter of simple algebra to confirm that there are parameter values for which the following holds: If the firm has a *low* number of fair agents, it pays a high bonus,  $\beta > 0$ , and a low wage. If the firm has a *high* number of fair agents, it pays no bonus,  $\beta = 0$ , and a relatively higher wage, i.e., it uses low-powered incentives. Here low-powered incentives serve as a credible signal, so conformists follow suit and behave like fair agents. This makes sense, because a principal who has a *high* fraction of steadfast fair agents will incur a smaller loss than a principal with a *low* fraction of fair agents when it sets the bonus to 0. A willingness of a principal to raise the fixed wage

---

<sup>54</sup>With perfect information, the bonus would be  $\beta_S = \frac{1}{2}$  for a selfish agent and  $\beta_F = \frac{1}{2} - \frac{\mu}{2(1-\mu)} < \beta_S$  for a fair agent. Ideally, the principal prefers a larger bonus for a selfish agent, but will want to have a positive bonus here as long as  $\mu < \frac{1}{2}$ . Notice also that fair agents have "intrinsic motives," and provide effort even when the bonus is 0.

further strengthens the signal. Conformists, in response to this credible signal, behave like fair agents. Profit for this firm is higher than if it used higher-powered incentives.

The structure of the model is very much like the Taylor-Ritter model discussed in Section 3.1, in which the firm (principal) uses compensation policy to signal hidden information about itself (i.e., the firm's financial fitness). In that model, low-powered incentives signal relatively good financial fitness, which allows the firm increased profitability.<sup>55</sup> Here, the principal's hidden information is the mix of worker type. There is a cost to low-powered incentives, of course, as effort is set by all agents to be less than first-best. But the low-powered incentive persuades some workers—the “conformists”—to behave in an altruistic fashion, when they would otherwise have not.

Sliwka interprets his model as generating “trust as a signal of a social norm.” In Sliwka's setting, the principal observes effort and can, if he chooses, condition rewards on effort. By setting no explicit incentives, the principal expresses trust in his workers. This trust directs a social norm. Some workers are more generous in their efforts than they would have been if they perceived a different norm.

An important feature of Sliwka's model is that monetary incentives can crowd out intrinsic incentives. If a firm moves from a “high trust” low-powered incentive scheme to a “low trust” high-powered incentive scheme, the firm shifts the norm and undermines the intrinsic portion of worker's incentives (the “social component” in the utility of a worker who would otherwise behave as a “fair agent”). Sliwka develops his theory further by looking at employee self-selection into firms. Here again, low-powered incentives can serve to attract workers with high intrinsic motives (fair agents), which serves to reinforce the positive work norms that influence those who conform to others.

The key behavioral underpinning of the Sliwka model is the observation that many people seem to want to conform to those around them. As we have noted, there is considerable evidence about the key component of this story—that people many people are influenced by

---

<sup>55</sup>Similar logic pertains in Spier's (1992) model, in which a principal knows more about the profitability and riskiness of a project than does an agent, and in Allen and Gale (1992), in which a supplier has superior information about his ability to distort a signal of production costs.

*norms*. For example, some individuals feel bad about a particular action only in situations in which they think others would experience remorse for that same action.<sup>56</sup>

In previous sections of this paper, we cited empirical work supportive of the social forces that create norms, e.g., studies by Ichino and Maggi (2000), in which worker absenteeism in an Italian bank was affected by the absenteeism of those around them, and Mas and Moretti (2009), in which effort by supermarket checkout workers was affected by other similar workers in their sightline. Yet another study, by Bandiera, Barankay, and Rasul (2009), shows that the productivity of farm workers is affected by the productivity of friends on the job. Jackson and Bruegmann (2009) document peer learning for teachers, which might be read as providing additional evidence conformity to norms. Of course, considerably more empirical work will be required to know if conformism plays a sufficiently strong role to generate in real-world organizations the “crowding out” or intrinsic motivation predicted by the Sliwka model.

### *5.3. Extrinsic Incentives when Agents Want to Impress the Principal*

Ellingsen and Johannesson (2008) present a model that, like Sliwka’s, relies on social preferences. Also, like Sliwka’s, their model opens up the possibility that extrinsic incentives can undermine valuable intrinsic motivations. The behavioral foundations, though, are quite different.

[A couple of paragraphs are needed here to tell the story. The punch line is that trust reinforces desired behavior by agents. I think I can show this with a very simple example.]

### *5.4. Extrinsic Rewards and an Individual’s Intrinsic Motivation*

The Sliwka (2007) and Ellingsen and Johannesson (2008) approaches provide plausible mechanisms whereby high-powered extrinsic rewards can undermine the average level of intrinsic motivation at a group level. In both cases, the psychological mechanisms hinge on agents learning about the characteristics of other relevant agents. In Sliwka’s model, some agents (the “conformists”) care about the preferences of other agents. In Ellingsen

---

<sup>56</sup>There is a great deal of empirical work across disciplines on norms. One particularly evocative story is told by Fisman and Miguel (2007): When United Nations diplomats in New York were given immunity for parking violations, and were therefore constrained only by cultural norms, violations were highest among diplomats from countries that had high levels of corruption (independently measured). At a theoretical level, work by Bernheim (1994) is important. Fischer and Huddart (2008) discusses the role of endogenous social norms on organizational design.



and Johannesson’s model, agents care about the principal’s esteem. Thus in these models, agents’ behaviors are shaped in fundamental ways by the actions of other around them—or at least by what the agents perceive them to be. But psychologists’ arguments that extrinsic rewards undermine intrinsic motives are not limited to setting in which agents make decisions by looking at the behaviors of other agents.

Can crowding out of intrinsic motives work also at an individual level? In an important paper by Bénabou and Tirole (2003) the answer is “yes.” In that paper, Bénabou and Tirole take a novel approach to the agency problem—one that ultimately yields great insight. They set up a problem in which a principal seeks to motivate an agent to provide effort toward a task that will create value to the principal, say  $W$ , if the task is successfully completed. Undertaking the task comes at a cost to the agent, but successful completion of the task creates intrinsic value for the agent as well. The key is that the *principal* has better information than does the agent about some crucial aspect of the task—for example, the cost the agent will incur if she undertakes the task, the personal satisfaction she will experience in the task is completed successfully (i.e., the *intrinsic value* to the agent), or the likelihood that the agent will in fact be successful in the task at hand.<sup>57</sup>

To keep matters simple, suppose the agent can adopt one of two effort levels,  $e = 0$  or 1. If the effort is 0, the proposed project fails and creates 0 value to the principal and 0 value to the agent. On the other hand, if the agent exerts high effort the project succeeds with probability  $\theta$  (where  $0 < \theta \leq 1$ ) in which case the agent receives intrinsic value  $V$ . The cost of effort is  $c$ . The principal’s compensation policy is to pay a performance-based bonus,  $b < W$ , if the task is successfully completed. As a baseline observation, note that if the agent observed  $V$ ,  $c$ , and  $\theta$  he would choose the high effort level if and only if

$$(44) \quad \theta(V + b) \geq c.$$

---

<sup>57</sup>Many signalling models assume a firm has hidden information about *itself*, e.g., its own technology, level of risk, likelihood of survival, or workforce. For example, above we mention Ritter and Taylor (1994), Sliwka (2007), Spier (1992), and Allen and Gale (1992), all of whom present models in which allow a firm’s offered compensation to signal some hidden firm characteristic. In the Bénabou-Tirole model we consider here, in contrast, the principal has hidden information about the *agent* (e.g., the agent’s fitness for the task at hand).

Suppose, instead, that the principal is better informed than the agent about some feature of the task. Specifically, suppose initially that the principal observes  $c$  but the agent observes only a signal, say  $\sigma$  that allows him to make inferences about the likely value of  $c$ . This signal is informative; it has a conditional distribution  $G(\sigma|c)$  with density  $g(\sigma|c)$ , and  $g(\cdot)$  has the property that for any signals  $\sigma_1$  and  $\sigma_2$  for which  $\sigma_1 > \sigma_2$ ,  $g(\sigma_1|c)/g(\sigma_2|c)$  is decreasing in  $c$ .<sup>58</sup> Now, of course, the agent's effort decision will depend on his *expectation* about the value of  $c$ , which depends on his signal, of course, but also, quite possibly, on the bonus  $b$  being offered. In short, the agent's best response is to provide effort if and only if

$$(45) \quad \theta(V + b) \geq E[c|\sigma, b].$$

Clearly, the effort decision now depends on how “good the news is” about the likely cost of completing the task; for any given  $b$ , there is a threshold signal, say  $\sigma^*(b)$ , such that high effort level is provided when the signal exceeds that threshold.

Given the agent's best response, how will the principal choose the intrinsic motivator  $b$ ? Note, first of all, that the principal's profit, for any choice of  $b$  is

$$(46) \quad \theta[1 - G(\sigma^*(b)|c)](W - b).$$

Bénabou and Tirole focus on a Perfect Bayesian equilibrium in this two-stage game that accords nicely with intuition. To see the logic, consider a principal who understands that the cost  $c$  an agent will face is going to be very high. The principal anticipates therefore that the signal  $\sigma$  will likely be low; the news the agent learns about the task is likely to be bad. If the news is indeed, the agent will be de-motivated. High-powered incentives will be necessary; the principal sets  $b$  to be relatively high. Unfortunately, from the principal's perspective, the agent will use the principal's actions—that the bonus is set to be high—to further reinforce the agent's view that the task is likely to be costly. That feature of the high-powered scheme is therefore de-motivating. A high-powered extrinsic reward undermines the intrinsic motivation to provide effort.<sup>59</sup>

<sup>58</sup>Thus, high values of  $\sigma$  are “good news;” they tell the agent that  $c$  is likely to be low.

<sup>59</sup>Formally, analysis shows: (1) If a principal offers  $b_1 < b_2$ , then for the agent  $\sigma^*(b_1) > \sigma^*(b_2)$ . A principal will offer a relatively high bonus only when the principal knows that the agent will respond by providing effort over a wider range of the signal. (2) Large bonus offers are “bad news” in the sense that they confirm to the agent that  $c$  must be relatively high. Thus, if  $b_1$  is offered when the principal knows that the agent's

We can ask, next, what happens if the cost of the activity  $c$  is known to both parties, but the principal has superior information to the agent about the intrinsic value the agent will earn upon successful completion of the task (e.g., the principal knows  $V$  but the agent only has a signal of  $V$ ). Results are in this case the same as when  $c$  is unknown. A high extrinsic reward in the form of a large bonus signals the principal’s assessment that the realization of the task will provide little in the way of intrinsic rewards.

These results illustrate what Bénabou and Tirole call the “trust effect.” As the authors note, this effect “arises when the principal’s private information concerns a parameter, such as the cost or pleasure of accomplishing the task, that directly enters only in the *agent’s incentive problem*—as envisioned by the principal.” A principal who trusts that the agent will be sufficiently motivated to work with few extrinsic rewards (once the agent’s makes his own observations about the task), would be wise to offer little in the way of extrinsic rewards.

Suppose, finally, that  $V$  and  $c$  are common knowledge, but the productivity of the worker is better known by the principal than the agent. In particular, suppose that productivity can take on one of two values,  $\theta_L$  or  $\theta_H > \theta_L$ , and that the principal knows the type of the agent, while the agent does not. Bénabou and Tirole show in this case that increased extrinsic incentive intensity leads  $\theta_L$ -agents to exert effort when they would otherwise not have done so. These increased extrinsic rewards are also “bad news” to agents in the sense that they signal the agent that he a low-productivity agent. As Bénabou and Tirole put it, for an agent, this news “damages self-confidence.” To the contrary, a low-powered incentive scheme can be thought of as a means of “confidence management.”

This analysis provides a clear and interesting way to evaluate the use of extrinsic rewards and intrinsic motives. When there is perfect information about the effort and likely payoffs involved, elements that affect the agent’s intrinsic well-being (reflected by  $V$  and  $c$ ) can be cleanly separated from the extrinsic reward ( $b$ ). If the agent is unsure about the likely net intrinsic elements of a task at hand, though, his assessment of those intrinsic payoffs can be affected by the intensity of the extrinsic rewards offered by the principal. Similarly,

---

cost is  $c_1$ , and  $b_2$  is offered when the principal knows that the cost is  $c_2$ , then  $b_2 \geq b_1$ . (3) The offered bonus undermines the agent’s assessment of difficulty of the task they are being asked to undertake. In particular, for any signals,  $\sigma_1$  and  $\sigma_2$ , and equilibrium bonus offers,  $b_1$  and  $b_2$ ,  $E(c|\sigma_1, b_1) < E(c|\sigma_2, b_2)$ .

there are circumstances in which a principal will provide higher-powered incentives only when agents have low ability, in which case the offer of such incentives signals to the agent that the principal has a low assessment of the agent's ability. Bénabou and Tirole offer a nice précis of the concerns about potentially demotivating extrinsic rewards: “To sum up, before worrying about the negative impact of rewards, one should first check that the reward provider has private information about the task or the agent's talent . . . One should then, as the agent does, think through the provider's ulterior motivation and how her payoff from giving a contingent reward is affected by her knowledge” (Bénabou and Tirole, 2003, p. 505).

High-powered incentives clearly work quite differently in the Bénabou-Tirole model than in the standard principal-agent model as specified in Section 2.1. To see how, consider a principal who seeks to motivate an agent in two consecutive periods, and suppose the principal tries two different levels of extrinsic incentive intensity, i.e., two different values of  $b$ , in the two different periods (but with the same agent). The standard model of course leads the agent to provide more effort when the bonus is higher. In the Bénabou-Tirole model, however, matters are much trickier. Suppose for instance that the principal offers the agent a large bonus in the first period. This leads the agent to understand that the task at hand is more difficult or less satisfying than he would have otherwise expected, or it leads the agent to lower his own view about how effective he will generally be at the task. This information is irreversible. Thus, if the principal tries a smaller bonus in the second period, the agent's motivation will typically be lower than it would have been if that same bonus had been offered in the first period. This latter idea is altogether anomalous in standard agency models, but common in psychological models—that the introduction, and then withdrawal, of high-powered incentives can permanently undermine intrinsic motivation.

### *5.5. Incentives and Identity*

[This section covers the Bénabou-Tirole (2006) on “self signalling.” A short example should suffice to make the point. The main point is that agents like taking actions—like being generous—that reinforce their own views of themselves.]

### 5.6. *Empirical Evidence on “Crowing Out” Intrinsic Motives*

This section gives a very brief overview of discussions from psychology and elsewhere on the topic. Deci. Fehr and Falk. Frey. Gneezy and Rustichini (2000).

## 6. CONCLUSIONS

[What are the key ideas? What are areas of future work? What have we missed in our discussions?]

## REFERENCES

- Acemoglu, Daron (1996), "A Microfoundation for Social Increasing Returns in Human Capital Accumulation," *Quarterly Journal of Economics*, 111(3): 779-804.
- Acemoglu, Daron, Michael Kremer, and Atif Mian (2008), "Incentives in Markets, Firms, and Governments," *Journal of Law, Economics, and Organization*, 24(2): 273-306.
- Akerlof, George A. (1976), "The Economics of Caste and of the Rat Race and Other Woeful Tales," *Quarterly Journal of Economics*, 90(4): 599-617.
- Akerlof, George A. (1982), "Labor Contracts as Partial Gift Exchange," *Quarterly Journal of Economics*, 97(4): 543-569.
- Akerlof, George A. (1984), "Gift Exchange and Efficiency-Wage Theory: Four Views," *American Economic Review*, 74(2): 79-83.
- Akerlof, George A. and William T. Dickens (1982), "The Economic Consequences of Cognitive Dissonance," *American Economic Review*, 72(3): 307-319.
- Akerlof, George A. and Rachel Kranton (2000), "Economics and Identity," *Quarterly Journal of Economics*, 115(3): 715-53.
- Akerlof, George A. and Rachel Kranton (2005), "Identity and the Economics of Organizations," *Journal of Economic Perspectives*, 19(1): 9-32.
- Akerlof, George A. and Janet L. Yellen (1990), "The Fair Wage-Effort Hypothesis and Unemployment," *Quarterly Journal of Economics*, 105(2): 255-283.
- Allen, Franklin and Douglas Gale (1992), "Measurement Distortion and Missing Contingencies in Optimal Contracts," *Economic Theory*, 2(1): 1-26.
- Andreoni, James, William Harbaugh and Lise Vesterlund (2003), "The Carrot or the Stick: Rewards, Punishments, and Cooperation," *American Economic Review*, 93(3): 893-902.

- Arrow, Kenneth J. (1998), "What Has Economics to Say about Racial Discrimination?" *Journal of Economic Perspectives*, 12(2): 91-100.
- Austen-Smith, David and Roland G. Fryer, Jr. (2005), "An Economic Analysis of 'Acting White,'" *Quarterly Journal of Economics*, 120(2): 551-83.
- Avorn, Jerry (2004), *Powerful Medicines: The Benefits, Risks and Costs of Prescription Drugs*. New York: Alfred A. Knopf.
- Babcock, Linda and Sara Laschever (2003), *Women Don't Ask: Negotiation and the Gender Divide*. Princeton and Oxford: Princeton University Press.
- Babcock, Linda and George Loewenstein (1997), "Explaining Bargaining Impasse: The Role of Self-Serving Biases," *Journal of Economic Perspectives*, 11(1): 109-26.
- Babcock, Linda, George Loewenstein, Samuel Issacharoff and Colin Camerer (1995), "Biased Judgments of Fairness in Bargaining," *American Economic Review*, 85(5): 1337-43.
- Babcock, Linda, Xianghong Wang, and George Loewenstein (1996), "Choosing the Wrong Pond: Social Comparisons in Negotiations That Reflect a Self-Serving Bias," *Quarterly Journal of Economics*, 111(1): 1-19.
- Baily, Martin Neil (1974), "Wages and Employment under Uncertain Demand," *Review of Economic Studies*, 41(1): 37-50.
- Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2005), "Social Preferences and the Response to Incentives: Evidence from Personnel Data," *Quarterly Journal of Economics*, 120(3): 917-962
- Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2007), "Incentives for Managers and Inequality Among Workers: Evidence From a Firm-Level Experiment," *Quarterly Journal of Economics*, 122(2): 729-773
- Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2006), "Social Connections and Incentives in the Workplace: Evidence from Personnel Data," Working Paper.

- Bartel, Ann, and Casey Ichniowski, and Kathryn Shaw (2004), "Using 'Insider Econometrics' to Study Productivity," *American Economic Review, Papers and Proceedings*, 94(2): 217-23.
- Bebchuk, Lucian and Jesse M. Fried (2003), "Executive Compensation as an Agency Problem," *Journal of Economic Perspectives*, 17(3): 71-92.
- Bebchuk, Lucian and Yaniv Grinstein (2005), "The Growth of Executive Pay," *Oxford Review of Economic Policy*, 21(2): 283-303.
- Bebchuk, Lucian, Yaniv Grinstein, and Urs Peyer (2006), "Lucky CEOs," Harvard Law School Working Paper.
- Becker, Gary, S. (1957), *The Economics of Discrimination*, Chicago: University of Chicago Press.
- Becker, Gary S. (1981), *A Treatise on the Family*, NBER Books.
- Bénabou, Roland, and Jean Tirole (2002), "Self-Confidence and Personal Motivation," *Quarterly Journal of Economics*, 117(3): 871-915.
- Bénabou, Roland, and Jean Tirole (2003), "Intrinsic and Extrinsic Motivation," *Review of Economic Studies*, 70(3): 489-520.
- Bénabou, Roland, and Jean Tirole (2006), "Incentives and Prosocial Behavior," *American Economic Review*, 96(5): 1652-1678.
- Berman, Eli (2000), "Sect, Subsidy, and Sacrifice: An Economist's View of Ultra-Orthodox Jews," *Quarterly Journal of Economics*, 115(3): 905-53.
- Besley, Timothy and Maitreesh Ghatak (2005), "Competition and Incentives with Motivated Agents," *American Economic Review*, 95(3): 616-36.
- Bewley, Truman F. (1998), "Why Not Cut Pay?" *European Economic Review*, 42(3-5): 459-90.



- Bisin, Alberto and Thierry Verdier (2000), “‘Beyond the Melting Pot’: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits,” *Quarterly Journal of Economics*, 115(3): 955-88.
- Black, Dan A. (1995), “Discrimination in an Equilibrium Search Model,” *Journal of Labor Economics*, 13(2): 309-34.
- Black, Dan A. Gary Gates, Seth Sanders, and Lowell J. Taylor (2000), “Demographics of the Gay and Lesbian Population in the United States: Evidence from Available Systematic Evidence,” *Demography*, 37(2): 139-54.
- Black, Dan A. Seth Sanders, and Lowell J. Taylor (2007), “The Economics of Lesbian and Gay Families,” *Journal of Economic Perspectives*, 21(2): 53-70.
- Bowles, Hannah Riley, Linda Babcock, and Lei Lai (2006), “Social Incentives for Gender Differences in the Propensity to Initiate Negotiations: Sometimes it Does Hurt to Ask,” *Organizational Behavior and Human Decision Processes*, 103(1): 84-103.
- Bowles, Hannah Riley, Linda Babcock, and Kathleen L. McGinn (2005), “Constraints and Triggers: Situational Mechanics of Gender in Negotiation,” *Journal of Personality and Social Psychology*, 89(6): 951-65.
- Bowles, Samuel (1985), “The Production Process in a Competitive Economy: Walrasian, Neo-Hobbesian, and Marxian Models,” *American Economic Review*, 75(1): 16-36.
- Bowles, Samuel and Herbert Gintis (1986), *Democracy and Capitalism: Property, Community and the Contradictions of Modern Social Thought*. New York: Basic Books.
- Brown, Charles and James Medoff (1989), “The Employer Size-Wage Effect,” *Journal of Political Economy*, 97(5): 1027-1059.
- Bulow, Jeremy I., and Lawrence H. Summers (1986), “A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination, and Keynesian Unemployment,” *Journal of Labor Economics*, 4(3): 376-414.

- Burdett, Kenneth and Dale T. Mortensen (1998), "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 39(2): 257-273.
- Cain, Daylian M., George Loewenstein, and Don A. Moore (2005), "The Dirt on Coming Clean: Perverse Effects of Disclosing Conflicts of Interest," *Journal of Legal Studies*, 34(1): 1-25.
- Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler (1997), "Labor Supply of New York City Cabdrivers: One Day At A Time," *Quarterly Journal of Economics*, 112(2): 407-41.
- Cappelli, Peter and Keith Chauvin (1991), "An Interplant Test of the Efficiency Wage Hypothesis," *Quarterly Journal of Economics*, 106(3): 769-787.
- Cebul, Randall, Ray Herschman, James Rebitzer, Lowell Taylor and Mark Votruba (2008), "Organizational Fragmentation in the U.S. Health Care System," *Journal of Economic Perspectives*, 22(4).
- Charness, Gary and Matthew Rabin (2002), "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, 117(3): 817-69.
- Coase, Ronald H. (1937), "The Nature of the Firm," *Economica*, 4(16): 386-405.
- Cooper, David J. and Rebitzer, James B. (2006), "Managed Care and Physician Incentives: The Effects of Competition on the Cost and Quality of Care," *B.E. Journals in Economic Analysis and Policy: Contributions to Economic Analysis and Policy*, 5(1): 1-30.
- Cornell, Bradford and Ivo Welch (1996), "Culture, Information, and Screening Discrimination," *Journal of Political Economy*, 104(3): 542-71.
- Courty, Pascal and Gerald Marschke (2004), "An Empirical Investigation of Gaming Responses to Explicit Performance Incentives," *Journal of Labor Economics*, 22(1): 23-56.

- Dana, Jason and Loewenstein, George (2003), "A Social Science Perspective on Gifts to Physicians from Industry," *Journal of the American Medical Association*, 290(2): 252-55.
- Delfgaauw, Josse and Robert Dur (2007), "Signaling and Screening of Worker Motivation," *Journal of Economic Behavior and Organization*, 62: 605-24.
- Delfgaauw, Josse and Robert Dur (2008), "Incentives and Workers' Motivation in the Public Sector," *Economic Journal*, 118: 171-91.
- DellaVigna, Stefano (2009), "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47(2): 315-72.
- deQuervain, Dominique J. F., *et al.* (2004), "The Neural Basis of Altruistic Punishment," *Science*, 305(5688): 1254-58.
- Dixit, Avinash (2002), "Incentives and Organizations in the Public Sector: An Interpretative Review," *Journal of Human Resources*, 37(4): 696-727.
- Ekman, Paul (2001), *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*, 3rd ed. New York City: W. W. Norton and Company.
- Ellingsen, Tore and Magnus Johannesson (2008), "Pride and Prejudice: The Human Side of Incentive Theory," *American Economic Review*, 98(3): 990-1008.
- Encinosa, William E., III, Martin Gaynor, and James B. Rebitzer (2007), "The Sociology of Groups and the Economics of Incentives: Theory and Evidence on Compensation Systems," *Journal of Economic Behavior and Organization*, 62(2): 187-214.
- Fairlie, Robert and William Sundstrom (1999), "The Emergence, Persistence, and Recent Widening of the Racial Unemployment Gap," *Industrial and Labor Relations Review*, 52(2): 252-70.
- Falk, Armin, Ernst Fehr, and Christian Zehnder (2006), "Fairness Perceptions and Reservation Wages—The Behavioral Effects of Minimum Wage Laws," *Quarterly Journal of Economics*, 121(4): 1347-81.

- Falk, Armin and Andrea Ichino (2006), "Clean Evidence on Peer Effects," *Journal of Labor Economics*, 24(1): 39-57.
- Farber, Henry S. (2005), "Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers," *Journal of Political Economy*, 113(1): 46-82.
- Farber, Henry S. (2008), "Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers," *American Economic Review*, 98(3): 1069-82.
- Fehr, Ernst, Martin Brown, and Christian Zehnder (2009), "On Reputation: A Microfoundation of Contract Enforcement and Price Rigidity," *Economic Journal*, 119: 333-53.
- Fehr, Ernst, and John A. List (2004), "The Hidden Costs and Returns of Incentives—Trust and Trustworthiness Among CEOs," *Journal of the European Economic Association*, 2(5): 743-771.
- Fehr, Ernst and Simon Gaechter (2000), "Cooperation and Punishment in Public Goods Experiments," *American Economic Review*, 90(4).
- Fehr, Ernst and Klaus M. Schmidt (1999), "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114(3): 817-68.
- Fehr, Ernst and Lorenz Goette (2007), "Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment," *American Economic Review*, 97(1): 298-317.
- Fehr, Ernst, Georg Kirchsteiger, Aron Riedl (1998), "Gift Exchange and Reciprocity in Competitive Experimental Markets," *European Economic Review*, 42(1): 1-34.
- Fernandez, Raquel, Alessandra Fogli, and Claudia Olivetti (2004), "Mothers and Sons: Preference Formation and Female Labor Force Dynamics," *Quarterly Journal of Economics*, 119(4): 1249-99.
- Fischbacher, Uris, Christina M. Fong, and Ernst Fehr (forthcoming), "Fairness, Errors and the Power of Competition," *Journal of Economic Behavior and Organization*.

- Fischer and Huddart (2008), "Optimal Contracting with Endogenous Social," *American Economic Review*, 98(4), 1459-75.
- Fisman, Raymond and Edward Miguel (2007), "Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets," *Journal of Political Economy*, 115(6), 1020-48.
- Frank, Robert H. (1984), "Are Workers Paid their Marginal Products?" *American Economic Review*, 74(4), 549-71.
- Frank, Robert H. (1985), *Choosing the Right Pond: Human Behavior and the Quest for Status*. New York: Oxford University Press.
- Frank, Robert H. (1988), *Passions Within Reason: The Strategic Role of the Emotions*. New York City: W. W. Norton and Company.
- Frey, Bruno S. (1997), *Not Just for the Money: An Economic Theory of Personal Motivation*. Cheltenham: Edward Elgar.
- Frey, Bruno S. and Reto Jegen (1999), "Motivation Crowding Theory," *Journal of Economic Surveys*, 15(5): 589-611.
- Gayle, George-Levi, and Robert A. Miller (forthcoming), "Has Moral Hazard Become a More Important Factor in Managerial Compensation?" *American Economic Review*.
- Gaynor, Martin, James Rebitzer, and Lowell J. Taylor (2004), "Physician Incentives in Health Maintenance Organizations," *Journal of Political Economy*, 112(4).
- Genesove, David and Christopher Mayer (2001), "Loss Aversion and Seller Behavior: Evidence from the Housing Market," *Quarterly Journal of Economics*, 116(4): 1233-60.
- Gibbons, Robert (1998), "Incentives in Organizations," *Journal of Economic Perspectives*, 12(4): 115-32.
- Gicheva, Dora (2009), "Working Long Hours and Career Wage Growth," Working Paper, Yale University.

- Gintis, Herbert, Samuel Bowles, Robert Boyd and Ernst Fehr (2003), "Explaining Altruistic Behavior in Humans," *Evolution and Human Behavior*, 24(3): 153-72.
- Gneezy, Uri, and John A. List (2006), "Putting Behavioral Economics to Work: Field Evidence of Gift Exchange," *Econometrica*, 74(5): 1365-84.
- Gneezy, Uri, Niederle, Muriel, and Rustichini, Aldo (2003), "Performance in Competitive Environments: Gender Differences," *Quarterly Journal of Economics*, 118(3): 1049-74.
- Gneezy, Uri, and Aldo Rustichini (2000), "Pay Enough or Don't Pay at All," *Quarterly Journal of Economics*, 115(3): 791-810.
- Grossman, Sanford J. and Oliver D. Hart (1986), "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94(4), 691-719.
- Groves, Theodore, Yongmiao Hong, John McMillan, and Barry Naughton (1994), "Autonomy and Incentives in Chinese State Enterprises," *Quarterly Journal of Economics*, 109(1): 183-209.
- Groves, Theodore, Yongmiao Hong, John McMillan and Barry Naughton (1995), "China's Evolving Managerial Labor Market," *Journal of Political Economy*, 103(4): 873-92.
- Hall, Brian and Jeffrey Liebman (1998), "Are CEOs Really Paid Like Bureaucrats?" *Quarterly Journal of Economics*, 113(3): 653-91.
- Harris, Milton and Bengt Holmström (1982), "A Theory of Wage Dynamics," *Review of Economic Studies*, 49(3): 315-333.
- Heyes, Anthony (2007), "The Economics of Vocation or Why is a Badly Paid Nurse a Good Nurse?" *Journal of Health Economics*, 24, 561-569.
- Holmström, Bengt (1979), "Moral Hazard and Observability," *Bell Journal of Economics*, 10(1), 74-91.

- Holmström, Bengt (1999), "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 66(1): 169-82.
- Holmström, Bengt and Paul Milgrom (1991), "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, 7, Special Issue (Papers from the Conference on the New Science of Organization): 24-52.
- Holmström, Bengt and Paul Milgrom (1994), "The Firm as an Incentive System," *American Economic Review*, 84(4), 972-991.
- Hornstein, Andreas, Per Krusell, and Giovanni Violante (2007), "Frictional Wage Dispersion in Search Models: A Quantitative Assessment," NBER Working Paper.
- Ichino, Andrea, and Giovanni Maggi (2000), "Work Environment and Individual Background: Explaining Regional Shirking Differentials in a Large Italian Firm," *Quarterly Journal of Economics*, 115(3): 1057-90.
- Jackson, Howell E. (2008), "The Trilateral Dilemma in Financial Regulation," in Anna Maria Lusardi, ed., *Improving the Effectiveness of Financial Education and Savings Programs*. Chicago: University of Chicago Press.
- Jackson, Kirabo C. and Elias Bruegmann (2009), "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers," Manuscript.
- Jacob, Brian A. (2005), "Accountability, Incentives and Behavior: the Impact of High-Stakes Testing in the Chicago Public Schools," *Journal of Public Economics Volume* 89(5-6), 761-796.
- Jacob, Brian A. and Steven D. Levitt (2003), "Rotten Apples: An Investigation of The Prevalence and Predictors of Teacher Cheating," *Quarterly Journal of Economics*, 118(3), 843-877.
- Jensen, Michael, and Kevin J. Murphy (1990), "Performance Pay and Top-Management Incentives," *Journal of Political Economy*, 98: 225-64.

- Kahn, Lawrence M. and Peter D. Sherer (1999), "Contingent Pay and Managerial Performance," *Industrial and Labor Relations Review*, 43(3): 107S-120S.
- Kahneman, Daniel, David Schkade, and Cass R. Sunstein (1998), "Shared Outrage and Erratic Awards: The Psychology of Punitive Damages," *Journal of Risk and Uncertainty*, 16(1): 49-86.
- Kandel, Eugene and Edward P. Lazear (1992), "Peer Pressure and Partnerships," *Journal of Political Economy*, 100(4): 801-17.
- Knez, Marc and Simester, Duncan (2001), "Firm-Wide Incentives and Mutual Monitoring at Continental Airlines," *Journal of Labor Economics*, 19(4): 743-72.
- Kreps, David, Paul Milgrom, John Roberts, and Robert Wilson (1982), "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory*, 27(2): 253-79.
- Krueger, Alan B. and Alexandre Mas (2004), "Strikes, Scabs, and Tread Separations: Labor Strife and the Production of Defective Bridgestone/Firestone Tires," *Journal of Political Economy*, 112(2): 253-89.
- Krueger, Alan B. and Lawrence H. Summers (1988), "Efficiency Wages and the Inter-Industry Wage Structure," *Econometrica*, 56(2): 259-293.
- Landers, Renee, James Rebitzer, and Lowell J. Taylor (1996), "Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms," *American Economic Review*, 86(3).
- Lang, Kevin (1986), "A Language Theory of Discrimination," *Quarterly Journal of Economics*, 101: 363-82.
- Lang, Kevin, Michael Manove, and William T. Dickens (2005), "Racial Discrimination in labor Markets with Posted Wage Offers," *American Economic Review*, 95(4): 1327-40.
- Lazear, Edward P. (1989), "Pay Equality and Industrial Politics," *Journal of Political Economy*, 97(3): 561-80.



- Lazear, Edward P. *Personnel Economics for Managers*, 1998, New York: John Wiley and Sons.
- Lazear, Edward P. (1999), "Culture and Language," *Journal of Political Economics*, 107(6, part 2): S95-S126.
- Lazear, Edward P. and Sherwin Rosen (1981), "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89(5): 841-64.
- Levine, David I. (1991), "Just-Cause Employment Policies in the Presence of Worker Adverse Selection," *Journal of Labor Economics*, 9(3): 294-305.
- Levy, Frank and Peter Temin (2007) "Inequality and Institutions in 20th Century America," MIT Department of Economics Working Paper.
- Loury, Glenn C. (2002), *The Anatomy of Racial Inequality*. Cambridge, MA: Harvard University Press.
- MacLeod, W. Bentley and James Malcomson (1989), "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica*, 57: 447-80.
- MacLeod, W. Bentley and Daniel Parent (1999), "Job Characteristics and the Form of Compensation," *Research in Labor Economics*, 18: 177-242.
- Malcomson, James M. (1984), "Work Incentives, Hierarchy, and Internal Labor Markets," *Journal of Political Economy*, 92(3): 486-507.
- Mas, Alexandre (2006), "Pay, Reference Points, and Police Performance," *Quarterly Journal of Economics*, 121(3): 783-821.
- Mas, Alexandre (2008), "Labour Unrest and the Quality of Production: Evidence from the Construction Equipment Resale Market," *Review of Economic Studies*, 75(1): 229-58.
- Mas, Alexandre and Enrico Moretti (2009), "Peers at Work," *American Economic Review*, 99(1): 112-45.

- McMillan, John, John Whalley and Lijing Zhu (1989), "The Impact of China's Economic Reforms on Agricultural Productivity Growth," *Journal of Political Economy*, 97(4): 781-807.
- Moore, Don A. and George Loewenstein (2004), "Self-Interest, Automaticity, and the Psychology of Conflict of Interest," *Social Justice Research*, 17(2): 189-202.
- Nagin, Daniel, James Rebitzer, Seth Sanders, and Lowell Taylor (2002), "Monitoring and Motivation: An Analysis of a Field Experiment" *American Economic Review*, 92(4).
- Malcomson, James M. (1984), "Work Incentives, Hierarchy, and Internal Labor Markets," *Journal of Political Economy*, 92(3): 486-507.
- Milgrom, Paul R. (1981), "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 2(2), 380-91.
- Neal, Derek (2006), "Why Has Black-White Convergence Stopped?" *Handbook of the Economics of Education*, vol. 1, edited by Eric Hanushek and Finis Welch. Amsterdam: Elsevier.
- Niederle, Muriel and Lise Vesterlund (2007), "Do Women Shy Away from Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics*, 122(3), 1067-1101.
- Oettinger, Gerald S. (1999), "An Empirical Analysis of the Daily Labor Supply of Stadium Vendors," *Journal of Political Economy*, 107(2): 360-92.
- Oyer, Paul (1998), "The Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality," *Quarterly Journal of Economics*, 113(1), 14985.
- Prendergast, Canice (1999), "The Provision of Incentives in Firms," *Journal of Economic Literature*, 37(1): 7-63.
- Prendergast, Canice (2007), "The Motivation and Bias of Bureaucrats," *American Economic Review*, 97(1): 180-96.

- Prendergast, Canice and Robert H. Topel (1996), "Favoritism in Organizations," *Journal of Political Economy*, 104(5): 958-78.
- Rabin, Matthew (1993), "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, 83: 1281-1302.
- Rabin, Matthew and Richard H. Thaler (2001), "Anomalies: Risk Aversion," *Journal of Economic Perspectives*, 15(1): 13.
- Rebitzer, James B., and Lowell J. Taylor (1991), "A Model of Dual Labor Markets When Product Demand is Uncertain," *Quarterly Journal of Economics*, 106(4).
- Rebitzer, James B., and Lowell J. Taylor (1995a), "Efficiency Wages and Employment Rents: The Employer Size Wage Effect in the Job Market for Lawyers," *Journal of Labor Economics*, 13(4).
- Rebitzer, James B., and Lowell J. Taylor (1995b), "Do Labor Markets Provide Enough Short-Hour Jobs? An Analysis of Work Hours and Work Incentives," *Economic Inquiry*, 33(2).
- Rebitzer, James B., and Lowell J. Taylor (1995c), "The Consequences of Minimum Wage Laws: Some New Theoretical Ideas," *Journal of Public Economics*, 56(2).
- Ritter, Joseph A., and Lowell J. Taylor (1994), "Workers as Creditors: Efficiency Wages and Performance Bonds," *American Economic Review*, 84(3).
- Ritter, Joseph A., and Lowell J. Taylor (forthcoming), "Racial Disparity in Unemployment," *Review of Economics and Statistics*.
- Ritter, Joseph A., and Lowell J. Taylor (2009), "Low-Powered Incentives and the Motivation of Critical Workers," manuscript.
- Rosen, Asa (2003), "Search, Bargaining, and Employer Discrimination," *Journal of Labor Economics*, 21(4): 807-30.

- Sally, David (2002), "Two Economic Applications of Sympathy," *Journal of Law, Economics, and Organization*, 18(2): 455-87.
- Shapiro, C. and J. Stiglitz (1984), "Involuntary Unemployment as a Worker Discipline Device," *American Economic Review*, 74(3): 433-44.
- Simon, Herbert A. (1951), "A Formal Theory of the Employment Relationship," *Econometrica*, 19(3), 293-305.
- Sliwka, Dirk (2007), "Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes," *American Economic Review*, 97(3): 999-1012.
- Small, Deborah, Michele Gelfand, Linda Babcock, and Hilary Gettman (2007), "Who Goes to the Bargaining Table? The Influence of Gender and Framing on the Initiation of Negotiation," *Journal of Personality and Social Psychology*, 93(4): 600-13.
- Spier Kathryn E. (1992), "Incomplete Contracts and Signalling," *RAND Journal of Economics*, 23(3) 432-43.
- Stevenson, Betsy and Justin Wolfers (2008), "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox," *Brookings Papers on Economic Activity*.
- Taylor, Lowell J. (2007), "Optimal Wages in the Market for Nurses: An Analysis Based on Heyes' Model," *Journal of Health Economics*, 26(5), 1027-30.
- Thaler, Richard H. and Cass R. Sunstein (2008), *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press.
- Williamson, Oliver E. (1985), *The Economic Institutions of Capitalism*. New York: Free Press.
- Valley, Kathleen L., Joseph Moag, and Max H. Bazerman (1998), "'A Matter of Trust:' Effects of Communication on the Efficiency and Distribution of Outcomes," *Journal of Economic Behavior and Organization*, 34(2): 211-38.