

**TRUST AND INCENTIVES
IN PRINCIPAL-AGENT NEGOTIATIONS:
THE “INSURANCE/INCENTIVE TRADE-OFF”**

Gary J. Miller
Department of Political Science
Washington University
One Brookings Drive
CB 1063
St. Louis, MO 63130-4899
Email: gjmiller@artsci.wustl.edu

Andrew B. Whitford
Department of Political Science
University of Kansas
and
RWJF Scholars in Health Policy Research Program*
University of Michigan
School of Public Health
Department of Health Management and Policy
109 S. Observatory, SPH-II
Ann Arbor, MI 48109-2029
734-763-0387
Fax: 734-936-9813
Email: whitford@sph.umich.edu

* Address for correspondence (until 7/31). We thank Dan Carpenter and participants of the Scientific Study of Bureaucracy conference in College Station, Texas for their helpful comments. All remaining errors are our own.

**TRUST AND INCENTIVES
IN PRINCIPAL-AGENT NEGOTIATIONS:
THE “INSURANCE/INCENTIVE TRADE-OFF”**

Abstract

The canonical principal-agent problem involves a risk-neutral principal who must use incentives to motivate a risk-averse agent to take a costly, unobservable action that improves the principal’s payoff. The standard solution requires an inefficient shifting of risk to the agent. This paper summarizes some experimental research that throws doubt on the validity of this conclusion. Experimental subjects were routinely able to achieve efficiency in agent effort levels *without* inefficient risk-sharing. This is precisely the kind of efficient outcome that principal-agency theory says is unavailable.

These experimental outcomes, while anomalous from the standpoint of principal-agency theory, are quite consistent with other experimental data testing notions of trust-based implicit contracting. Such contracting within a hierarchy may allow an outcome preferred, by both principal and agent, to that deemed possible by principal-agency theory.

If this is true, then the lessons to be learned from principal-agency theory are all the wrong ones. Concentrating on incentives can crowd out the very qualities in a relationship that make social efficiency possible.

The connections to the literature on public bureaucracy are numerous. The focus on principal-agency theory and incentives has caused us to lose track of the research agenda that once defined public bureaucracy – such as bureaucratic “types”, cooperation in a hierarchy, and professionalism and its obligations.

More than half a century ago, Berle and Means (1932) argued that many diffuse shareholders could not hope to constrain corporate managers to act in the shareholders' interest. Corporate managers were, therefore, free to pursue their own interests. Since the interests of managers were quite distinct from those of shareholders, managers were presumably free to use the resources of the firm in their own interests. The legal rights of the stockholders were not matched, they argued, by effective control of the firm.

Not coincidentally, Berle and Means wrote at the dawn of the New Deal, a period in which the evolving power of unelected federal bureaucrats raised similar concerns about managerial accountability in the public sphere. The problem of public bureaucratic accountability was in some ways analogous to that of private managerial accountability. The legal rights of citizens in a democracy are held by a large diffuse constituency; the necessity of energetic and effective management of large organizations seemed at times to create a real mismatch between citizenship (or ownership) and control of the organizational infrastructure of the state (or firm).

In the 1970s, concern with the problem of corporate accountability generated a body of research known as principal-agency theory (Fama 1980; Mirrlees 1976; Ross 1973; Shavell 1979). This literature recognizes an information asymmetry. Shareholders (the principal) cannot monitor corporate managers directly; therefore, they need to induce an agent to take costly actions in the interests of the principal, without monitoring. The early literature found that, despite the information asymmetry, principals could find a (constrained) optimal solution that induces the greatest possible agent effort even though the principal can never directly observe the effort.

As popularly interpreted, the message of principal-agency theory has been that Berle and Means underestimated the ability of principals to shape agents' behavior; even in the presence of an information asymmetry, shareholders may use incentives to control the behavior of managers. In academics, this has caused an explosion of research on incentive contracts for corporate managers, sales staff, and all other positions in the firm (Milgrom and Roberts 1992). It has also fueled an expansion in the use of such incentives in the world of corporate America.

Principal-agency theory has had enormous influence in the study of public agencies as well as private firms. Part of the attraction has been the use of principal-agency theory to study accountability in hierarchical relationships where monitoring costs make direct control infeasible. For example, the literature on congressional oversight of bureaucracy had traditionally lamented the scarcity of direct monitoring by congressional committees, and Fiorina (1981) had lamented a "Mismatch of Incentives and Capabilities" in congressional control of the bureaucracy. However, Weingast (1984) used principal-agency theory to argue that Congress's lack of direct oversight could be a manifestation of effective use of incentives, which allow Congress to control agencies *without* expensive oversight procedures.

The purpose of this paper is to re-examine canonical principal-agency theory and its implications for public bureaucracy. We argue that the primary message of the theory is really a negative one – efficiency is not easily to be obtained. We further argue that in the case of public bureaucracy the costs of using incentives are likely to be high. We also argue that empirical research on the canonical principal-agency problem casts doubt on the predictive power of the theory, and that those doubts are again especially applicable in the case of public bureaucracy. Furthermore, if the predictions of principal-agent theory turn out to be the focus on incentives

brought about by the interest in principal-agency theory may lead to a set of managerial prescriptions that are counter-productive.

Because of these issues, the concentration on incentive plans as a solution to the problems facing private and public bureaucracies is misplaced. The real concern should be on traditional concerns of the public administration literature: internal motivation, selection, organization, and bureaucratic politics.

In the next section, we review the tradeoff between insurance and incentives implied by principal-agency theory. The subsequent section addresses an experiment meant to test the principal-agency solution to this tradeoff. We then turn to reciprocity as an explanation for the anomalies observed in principal-agent exchanges, and then to the crowding out alternative motivations by a focus on incentives. Last we offer a set of illustrations of the power of reciprocity in work environments, with special emphasis on public management.

INSURANCE VERSUS INCENTIVES: THE LESSONS OF PRINCIPAL-AGENCY

To induce the agent to work hard, we will have to give up some of the efficiency that is obtained by putting all the risk on the principal. The question is, How can we do this as efficiently as possible? (Kreps 1990, 584)

The basic principal-agency problem is very simple. There is one principal and one agent. The principal can not observe the agent's efforts, but can clearly measure outcomes that affect her wellbeing – in fact there is only one outcome that matters. In Dixit and Nalebuff's highly accessible illustration, that outcome is the success or failure of the principal's computer game (1991).¹

If the outcome in question were dependent *only* on the agent's actions, then the principal could work backwards from the observed outcome to deduce the agent's actions. For example, if the computer program's success were completely determined by the sales agent's efforts, then failure could only be caused by low effort. This deducibility makes the principal's problem just as trivial as if the agent's actions were directly observable: pay the agent only for the desired outcome. In this case, there would be no control loss by the principal. The outcome would be perfectly efficient, and the principal could extract all the profits from the program beyond the agent's opportunity cost (which would be determined by the market for the agent's labor).

But in general, the outcome is determined *in part* by some exogenous factor that appears as a random variable in the production function. In the example above, the success of the computer game is normally determined in large part by a random variable known as *demand*, which is the result of the economy, shifting tastes, changing technologies, the prices of other substitute and complement goods, all of which are out of the control of the sales agent. This random variable creates variability, or risk, in the outcome, and in the principal's wellbeing.

Assume that the principal and the agent are risk-neutral. If the program turns out to be a success, it will generate \$200,000 in revenue. The chances for success depend on the effort of the computer programmer. The programmer could apply either a *routine* or *high* effort. A routine effort would cost the programmer \$50,000 worth of effort cost, and create a $p = 60\%$ chance of success. If the programmer put in a high effort, the effort cost would increase to \$70,000, and the chance of success would increase to $q = 80\%$ (Dixit and Nalebuff 1991, 306). These are the same parameters used in the experiments described later in the paper.

If the principal could cheaply monitor the agent's efforts, then the principal could be sure of making \$90,000 in expected profits. To do this, she would contract to pay the agent \$70,000

for a high effort, monitor to make sure that effort was provided, and receive in return an 80% chance at \$200,000 in profits. Can the principal do just as well if she *cannot* monitor the agent?

The answer of principal-agency theory is that the principal can still net an expected \$90,000, as long as the agent is risk-neutral. The trick is to find some way to motivate the agent to make a high effort, even when she knows that effort is not monitorable or directly rewardable. The *marginal* cost of a high effort to the programmer is a crucial variable we will denote by **E**. It is the difference between the cost of a high effort and a low effort. In this example, **E** is \$70,000 - \$50,000 or \$20,000.

A flat wage cannot provide the motivation to take the high effort, because the flat wage would by definition be paid whether or not the programmer supplies the high effort. The agent would always prefer low to high effort if the principal has chosen to pay a flat wage, even if the flat wage were much greater than that required to cover the high effort cost.

Consequently, the owner must pay the programmer an outcome-dependent (and therefore risky) *bonus* **B**. The bonus will be paid in the event of a success even if the programmer shirked. And it will *not* be paid in the event of a failure, even if the programmer provided a high effort. As a result, a risk-neutral programmer will pay the additional effort cost **E** only if the expected value is greater than the expected value of providing low effort. The agent's expected payoff for high effort is $q\mathbf{B} - \$70,000$, while the expected payoff for low effort is $p\mathbf{B} - \$50,000$. If $q\mathbf{B} - \$70,000 > p\mathbf{B} - \$50,000$, then it will be worthwhile for the programmer to provide the additional effort. This requirement can be stated as $\mathbf{B} > \mathbf{E}/(q-p)$.

We can think of $q-p$ as a measure of efficacy. In this example, the agent's efficacy is 20%, because a high effort instead of a low effort increases the probability of success by that proportion. If $q-p$ were equal to one, then a high effort by the agent completely determines

success, and the bonus need only cover the marginal effort cost to induce a high effort. If $q-p$ is small, it means that extra effort by the agent has little impact on the probability of success, and B must be potentially *much* larger than the effort cost to induce effort. For our example, the necessary bonus that solves the principal's problem is $E/(q-p) = \$20,000/0.2$, or \$100,000. In other words, a 20% increase in the probability of getting a bonus of \$100,000 just compensates the programmer for an extra \$20,000 in effort cost. Any bonus less than \$100,000, in Dixit and Nalebuff's example and in our subsequent experiments, should be insufficient to justify a high level of effort by a rational agent.

If the agent and the principal were both risk-neutral, a bonus payment of \$100,000 would completely compensate for the information asymmetry. That is, with a bonus of \$100,000, the owner could be just as confident she was getting a high effort level from her programmer as if she were monitoring it directly. As Dixit and Nalebuff point out, a bonus of \$100,000 (with a \$200,000 success) is equivalent to granting the risk-neutral agent a 50% ownership stake in the firm in order to guarantee high effort. For this 50% ownership, the principal could even *charge* the agent \$10,000, plus his labor. The agent's expected payoff would be \$70,000 (an 80% chance at \$100,000, less the \$10,000 upfront charge to the agent); this would exactly compensate the risk-neutral agent for the cost of supplying a high effort. By granting a bonus of \$100,000, the principal is left with an 80% chance at \$100,000 (valued at \$80,000), plus the \$10,000 charge to the agent, for a net expected profits of \$90,000 – exactly the same net profits that were obtainable *with no information asymmetry*. Thus, there is no efficiency loss, and no loss of profit to the owner, due to the information asymmetry. This is the solution to the principal's problem.

As Dixit and Nalebuff point out, “The incentive system has done a perfect job; the unobservability of effort hasn’t made any difference” (1991, 305). This message, *that incentives can efficiently counteract information asymmetry*, is often the message that is carried over to public administration. However, as principal-agency theorists are careful to note, this is not the whole story, for there *must be* an efficiency loss when information asymmetry is combined with *risk aversion* – and bureaucrats are notably risk averse.

Risk Aversion

Risk neutrality on the part of the *principal* is a reasonable assumption.² Because of the modern stock market, entrepreneurship is not just for risk-takers. After all, every American capitalist has the efficiency of the capital market on which to manage risk efficiently and to hedge against particularly threatening risks. A risk-averse investor can invest in a way that minimizes risk by balancing investments that may do badly in some circumstances with investments that will do well in those same circumstances. Overall, each investor achieves the most wealth if the managers of each firm in her portfolio act as risk-neutral profit-maximizers. If each of her firms risks \$1000 on a 1/1000 chance at \$1.5 million, she will maximize her wealth, even though most of the firms go broke! Thus, modern corporate finance assumes that stockholders act as if they are risk neutral with respect to each investment.

Each firm’s managers are more risk averse. No manager of a firm will want to take such long shots at big profits, even if it maximizes the wealth of the firm’s shareholders. Thus, one essential aspect of principal-agency theory is the presumed discrepancy in risk preferences between owners and managers. The problem for shareholders is to coerce managers to take risks that they don’t want to take (Fama 1980).

In the Dixit-Nalebuff example, the owner is presumed to be risk neutral. The programmer, on the other hand, may rely on his labor for all of his income. In the case of a failure, a bonus incentive system would result in no salary whatsoever, which would have serious consequences for him and his family. For this reason, principal-agency theory normally assumes the agent is more risk-averse than the principal. It is this difference in risk-aversion that creates an efficiency requirement that is *incompatible* with the necessity of imposing risky incentives on the agent.

Whenever two people have different risk preferences, both could be made better off if the more risk-averse person buys insurance from the less risk-averse. Both the driver and the insurance company are made better off by a transaction in which the driver pays a premium and the insurance company absorbs the risk of an auto accident. If the driver and the insurance company were prevented from making that transaction, the result would be inefficiency in the market for risk. One thing that produces these inefficiencies is *moral hazard*, the insurance company's recognition that some drivers are more likely to drive recklessly insured than uninsured.

Similarly, moral hazard on the part of the agent prevents efficiency in his relationships with the principal. The risk-averse programmer may require a bonus of *more than \$100,000* to compensate for the extra \$20,000 effort cost imposed by a choice of high effort. For example, a risk-averse programmer may require a bonus of \$110,000 to compensate for the risk he is taking. Otherwise, he could "play it safe" by simply supplying low effort, which puts less of his own effort costs at risk.

Does the required extra bonus represent a redistribution from the agent to the principal, or an actual efficiency loss? It turns out that it represents an actual efficiency loss. To see this,

consider two schemes, one in which the owner pays a flat wage of \$85,000, and one in which the owner pays a bonus of \$110,00 for success. A bonus of \$110,000 has an expected payout of $0.8 * \$110,000$, or \$88,000, so it is more costly than a flat wage of \$85,000. If it could be assured of generating a high effort, the owner would then prefer the flat wage of \$85,000. The risk-averse agent also prefers a sure \$85,000 to the 20% possibility of a loss of \$70,000 of effort with bonus-based compensation. (See Technical Appendix A for a sample utility function that justifies this statement.) But the owner's recognition that a flat wage, no matter how generous, does nothing to reduce the agent's moral hazard (incentive to shirk) prevents the owner from using the flat wage.

Even if the risk-averse agent (like the risk-neutral agent) could be induced to "buy into" the relationship with an upfront fee of \$10,000, the principal will see a reduction in expected profits to \$82,000. Both principal and agent are made worse off by the inefficiency caused by the necessity of compensating a risk-averse agent with a risky outcome-based compensation. As Kreps notes, "To induce the agent to work hard, we will have to give up some of the efficiency that is obtained by putting all the risk on the principal" (1990, 584).

The Lesson of Principal-Agency Theory

The underlying lesson of principal-agency theory, then, is a negative one. As long as agents are risk-averse, principals may succeed in inducing effort even in the presence of an information asymmetry, but only at a personal and social cost.

This efficiency loss may be sufficient to justify an abandonment of incentives in favor of paying for direct monitoring and supervision. For example, we notice that, in practice, incentive plans that are laden with risk often run into serious problems in practice. At DuPont, an incentive plan went into effect in good economic times, and it initially delivered good-sized

incentive bonuses; but many employees were worried about their prospects during bad economic times, which they pointed out would be no fault of theirs. DuPont eventually had to abandon the incentive plan as they recognized the level of employee risk aversion, and the demotivating consequences of that risk aversion (Wall Street Journal 1988). As the employees of high tech Internet firms have discovered that stock options are indeed a risky form of compensation, their employers have increasingly shifted toward salaries to attract workers (Wilcox 2000).

In a public bureaucracy setting, how likely is it that agency costs will be significant? In bureaucracies such as a school, a police department, or the proverbial state road crew, the effort cost of providing a routine effort (following the rules, going along with accepted norms) are likely to be small. The *additional* effort costs of being really committed to success (**E**) are likely to be extremely large. As any teacher knows, a commitment to excellence in classroom instruction is extremely costly. Blau (1964) provided a classic account of employment counselors who practiced “creaming”, working with easy-to-place clients—a practice which minimized their effort costs while diminishing the value of the service provided, since the clients thus served were the ones who could most easily find jobs for themselves. Brehm and Gates (1997) show that there is a significant variation in the level of effort by police; one can easily infer the difference in effort costs that motivates some police officers to linger in doughnut shops or seek nap opportunities on late shifts. Overall, Lipsky (1980) argues that the job of the street-level bureaucrat is a very difficult one to do correctly, made more difficult by inadequate resources, hostile clients, and unclear performance standards. The net result is that many such bureaucrats accommodate to the job by hiding behind red tape and routine; the effort required to increase the probability of success with a given client can be costly. So **E** is likely to be large in many public bureaucracies.

But the bonus necessary to induce these costs is a factor *larger than the effort cost*, because E must be divided by the probability that an individual will actually make a difference to the success of the organization. In Missouri, third-grade teachers are judged by whether or not a certain proportion of their students pass a standardized state exam – a relatively concrete goal for public bureaucrats. In a certain elementary school in the St. Louis area, the principal regularly harangues the third-grade teachers (but not the first-grade or second-grade teachers) about the necessity of meeting that goal. The third-grade teachers, however, can only work with the students who come to them at the beginning of the year; their ability to pass the third-grade exam is largely determined by the success of the first and second grade teachers and, before that, the students' preschool experience and the commitment of the parents. This exact situation is played out across the country on a regular basis, as a recent series on testing in the New York state educational system makes abundantly clear (Goodnough 2001).

As a result, the third-grade teachers feel that $(q-p)$ is relatively small. As a result, $E/(p-q)$ is enormous. Realistically, the size of the financial bonus that would be required to make it in the self-interest of third-grade teachers to commit to a maximal effort to meet the goal must be several times the large marginal cost of high effort by an elementary teacher in Missouri or New York.

Consequently, it is probably fair to say that principal-agency theory (and outcome-based incentives) provide limited help in understanding why any third-grade teacher would really commit to anything more than a routine effort, in the case where such an effort would be not directly observable. Incentives based on outcomes are insufficient to motivate high effort in situations of low individual efficacy and a high marginal cost of transcendent effort.

THE EMPIRICAL VALIDITY OF PRINCIPAL-AGENCY

The gift of the firm to the worker (in return for the worker's gift of hard work for the firm) consists in part of a wage that is fair in terms of the norms of this gift giving (Akerlof 1982, 555).

In the previous section, we argue that principal-agency theory gives us little hope of understanding high effort levels on the part of risk-averse agents who feel that the marginal cost of high effort is great and their personal effect on success is low. In this section, we ask a related question: do the empirical predictions of principal-agency theory fit available data? That is, can principals and agents negotiate contracts that provide the incentive bonuses that are presumed to be necessary for success? Even more basically, are incentive bonuses necessary for high effort in the presence of information asymmetry?

Existing studies provide mixed evidence on these questions. In an early study, Eisenstadt (1988) found that routine sales jobs like operating a cash register were paid with fixed salaries, while sales positions that involved establishing a close relationship with the customer were paid with a commission. This is consistent with principal-agency theory in that employers bore the risk for those positions in which employees could be cheaply and easily monitored; but risk was shifted to those employees from whom it was necessary to replace monitoring with output-based incentives. Eisenstadt did not establish whether those firms that deviated from this pattern were less efficient.

While a predictable share of the sales force is paid with outcome-based incentives, on the whole, incentives are simply not used as frequently as one might expect. Lawler (1971, 158) concludes that most employers create very weak ties between performance and outcome. Baker Jensen, and Murphy (1988, 594) find this to be a striking example of a variety of compensation

practices that they conclude are “inconsistent with our traditional economic theories”. They speculate that managers may find it difficult to link pay to employee, and that managers themselves are not compensated in a way that provides an incentive for this task. They confirm this conjecture in their review of the use of incentive compensation for bosses, where “the lack of strong pay-for-performance incentives for CEOs indicated by our evidence is puzzling” (Baker, Jensen and Murphy 1990, 262).

The narrower question is, are incentives effective in enhancing performance when used? Or, conversely, does the *lack* of incentives in fact discourage performance compared to their use? Homans (1954) studied the productivity of employees posting cash payments for a utility company and found that on average they exceeded work standards by an average of 15%, with no expectation of a bonus or promotion. Whyte (1955) documents the countervailing social effects that limit the efficacy of piece-rate plans. Rothe (1970) tracked the productivity of welders after the elimination of an incentive plan and found that after an initial reduction, performance levels without incentives returned to the levels they had attained with incentives. Rich and Larson (1987) were unable to find any evidence that incentive plans for top executives resulted in performance improvements that benefited shareholders. These studies open the possibility that other forms of motivation may be as important, or even more effective, than outcome-based compensation.

Controlled laboratory experiments may be necessary to see the causal effects of outcome-based compensation. McLean Parks and Conlon (1995) provide a controlled laboratory study with owner/employee dyads who produced solutions to a mathematical problem. The employee’s contribution was to spend money for computer-generated solutions that stochastically increased in accuracy with the agent’s expenditures. The pairs negotiated

compensation contracts composed in part of a flat wage and in part of an outcome-based bonus. Each dyad participated in twenty trials, negotiating a contract and completing the exercise in each trial. In half of the dyads, the agent's expenditure could be monitored by the owner in the next period, and in half, the agent's expenditure could never be realized. Another, independent treatment was the profitability to the owner.

In profitable environments, dyads agreed to much more use of outcome-contingent bonuses when monitoring was not possible. This result is entirely analogous to Eisenhardt's observation that commissions are used more frequently for difficult-to-monitor sales assignments. However, in unprofitable environments, dyads used contingent bonuses *less* frequently.

Most interestingly, employees tended to *over-produce* in all settings. In the profitable environments, employees spent on average 2.6 times the optimal amount on information, regardless of monitoring. In the unprofitable treatments, the ratio was 1.7 or 5.6, depending on whether monitoring was or was not possible (McLean Parks and Conlon 1995, 826). The strikingly super-optimal level of expenditures indicates that some other motivation (e.g., competition with other subjects or an inner desire to solve the math problem) was driving the experimental subjects to provide the overly large expenditures that they did.

A recent experiment by Guth, *et al.* (1998) is virtually unique in its attempt to examine the effect of compensation contracts on employee effort levels. In these repeated games, the principal offers from one to three contracts consisting of a profit share and fixed wage; the agent chooses one. They found that employers did tend to offer increasing fixed wages across trials, and that agents' costly (and private) effort levels increase across trials. The effect of profit-share on effort is negligible early in a relationship, but increases over time. Contrary to principal-agent

models, the fixed wage had a sizable independent effect on agent's choice of work effort. In fact, the size of the fixed wage increased across periods, and the responsiveness of the agent's choice to the fixed wage both increased.

The Guth results are very suggestive. As with the McLean Parks and Conlon, repeated play provides a mechanism to reach greater social efficiency in principal-agent negotiations. If replicable, they imply that *in a repeated game*, principal-agent dyads can approach a more cooperative equilibrium in which inefficient risk sharing is *not* required for efficient incentives. Yet, in a repeated game, the Folk Theorem implies that an infinite range of equilibria is supportable.

Our experimental results deviate from the approach these two studies bring to understanding the principal-agent problem. Instead of concentrating on the power of the paradigm in the context of repeated play, we report results from a pilot experiment where subjects interacted in one-shot negotiation experiments. Because of the power of the Folk Theorem, this approach helps answer the fundamental question in principal-agent: what is the underlying mechanism by which interacting subjects are able to reach beyond the outcomes the prescriptions of principal-agent say are possible? Is repeated play necessary for subjects to reach even higher levels of social efficiency?

In this pilot experiment, 116 MBA students at a top-20 business school were randomly assigned to pairs in order to negotiate an employment contract. In each pair, one subject was assigned to a "principal" role, and one to an "agent". They faced parameters equal to those in the Dixit and Nalebuff problem. The contract would consist of two variables: a flat wage, and a bonus to be given to the employee in case of a success. Either payment could be zero or any positive number. The students were playing for "participation points" that would count toward

their class grade, in a highly motivated environment (Kormendi and Plott 1982). It was one of a series of such exercises that they participated in during the semester class, and in each they demonstrated a high level of intensity and commitment to earning points.

In this exercise, the points they earned were based on how much profits or net benefit they earned as owner, or programmer, respectively, as a result of the negotiation. If the program was a success, the profit earned by the owner was \$200,000 less flat wage and bonus; if the program was a failure, profits were negative: the owner still had to pay whatever flat wage was promised. The programmer earned the bonus plus the flat wage less the effort cost in the case of a success; he earned only the flat wage less effort cost in case of a failure. The effort cost was \$70,000 for a high effort, and \$50,000 for a routine effort. At the end of the exercise, the programmer, acting alone and in secret, had a chance to select a high effort or a routine effort.

The owner never found out what choice the programmer had made.

The probability of success was 60% with a routine effort and 80% with a high effort, just as in the original Dixit and Nalebuff example. To implement this, the owner picked six numbers from one to 10; the programmer selected two more.

To determine the success of each pair, a number from one to 10 was secretly selected after the exercise was concluded. If the number was one of the 6 numbers selected by the owner, then the program was a success, regardless of the programmer's effort level; the owner got \$200,000 less the flat wage and bonus paid to the programmer. If the number selected was one of the two numbers selected by the programmer, *and if the programmer had selected high effort*, then the program was also a success; otherwise, the program was a failure. If the number was not one of those selected by either the owner or the programmer, then the program was a failure regardless of the programmer's effort level. This operationalization captured the 60% success

rate with a low effort and the 80% success rate with high effort that were promised to the subjects. See Technical Appendix **B** for complete instructions for our pilot experiment.

The MBA students found this a challenging exercise. They had had several previous negotiation exercises, in which they were pitted against each other in pairs in various strategic situations, and the degree of competitiveness in this exercise matched that of previous exercises, at least to the observer.

Risk Aversion

It was clear that the programmers were quite risk averse. They were concerned about providing either a routine or high effort cost, and then losing the bonus in the event that the program was not a success. This meant that they were quite concerned to negotiate a flat wage, which was a form of insurance against a program failure. However, the mean flat wage negotiated was \$47,500. While this did not cover the programmers against the loss of a routine effort cost (\$50,000) or a high effort cost (\$70,000), it was a substantial amount of insurance. This is especially the case, since that \$47,500 was a pure loss to the owners in the event of a program failure. It was also substantial considering that the flat wage had *no incentive effect*; it did nothing to encourage programmers to provide the unobserved high effort cost that the owner needed to guarantee a higher probability of success.

One reason that the programmers did so well in negotiating the effort cost could well be differential risk preferences. Recent research has demonstrated that negatively-framed subjects are more risk acceptant than positively framed subjects in negotiation settings (Bottom 1998). The owners were negatively framed, since their concern was with *how much of the \$200,000 potential profits they could keep*. This “loss avoidance” orientation systematically predicts more risk-acceptant behavior. The programmers were programmed to see how much compensation

they could negotiate, compared with a starting point of zero. This “gain-seeking” orientation is systematically associated with more risk aversion. Putting a risk-acceptant and a risk-averse player together, then the negotiation is no longer a constant-sum exercise: both could be made better off by shifting risk to the more risk-acceptant owner. This is precisely what happened with the large flat wages that were negotiated. From the standpoint of experimental design, this is advantageous, since principal-agency theory assumes that the owner is more risk-acceptant than the programmer. The framing of the subjects thus serves to satisfy the risk-preference assumptions of the theory.

Incentives

From the standpoint of principal-agency theory, however, the owner’s real concern should be with the incentive bonus, not the flat wage. Negotiating a large enough bonus to increase the chance of success from 60% to 80% is essential to efficiency. The marginal cost of the high effort is \$20,000. The extra effort cost generates an extra 20% chance of \$200,000, for a marginal benefit of \$40,000. Thus, efficiency requires a high effort, and (according to principal-agency theory) high effort requires a bonus of \$100,000.

[Insert Table 1 about here.]

In fact, the mean bonus negotiated by the MBA students was only a bit more than \$70,000. This means that the average programmer, with a flat wage of \$47,500 and a bonus of \$70,000, earned an expected compensation of \$89,500 for routine effort—well worth the trouble of a routine effort. With a routine effort, the average programmer had an expected net benefit of \$39,500. But this compensation package did not create incentives for high levels of agent effort. The average programmer would make himself *worse off* with a high effort, since the expected gain (an extra 20% chance at \$70,000) is less than the marginal cost.

In other words, the MBAs systematically failed to negotiate a package of financial inducements that would motivate high effort. Only fifteen pairs, or 27.6%, of the MBA students negotiated a bonus of \$100,000. The failure of most pairs to shift a sufficient amount of risk to reward high effort constitutes a challenge to the empirical validity of principal-agency theory.

Of those fifteen pairs, all but one of the programmers did supply a high effort level. The effort level requirement for efficiency was met. However, there was a significant efficiency cost to negotiating large bonuses, assuming that the programmer was in fact more risk-averse than the owner. The efficiency loss was due to *inefficient bearing of risk*. We observe this cost in the fifteen pairs of MBA students who did ensure a high effort from the programmer by shifting sufficient risk from the owner.

Avoiding Moral Hazard

Forty-three out of the fifty-eight pairs negotiated a bonus of less than \$100,000, leaving them with no incentive to supply a high effort. *Yet more than 83% of these agents supplied a high effort*, despite the fact that their decision was a secret that was never revealed to their principals. Each of these forty-three agents chose a smaller over a larger expected net payoff, immediately after a round of intense, apparently self-interested negotiation. How can we make sense of this?

This pilot result is first of all a profound challenge to principal-agency theory. A primary assumption of principal-agency theory is that incentives are *necessary* to get agents to act in the interests of the principal.

Nor did these programmers do so out of simple altruism. The purest altruism would have been evidenced in a willingness to work hard for zero wages. The fact is that they gave every

evidence of a strong commitment to their own “financial” self-interest in the negotiations that preceded their effort decision.

One explanation is cognitive error. Subjects must be able to do an expected value calculation, and to understand the concept of marginal cost and marginal benefit. However, these are topics that the students (all graduating second-year MBA students) should have covered adequately in a highly technical first-year curriculum that covered microeconomics and decision-making in detail. While some of the students could have been quite confused on the subject, it is unlikely that this explains the behavior of a majority of the students.

Moreover, these results occurred in a single-shot interaction. Unlike the repeated play results detailed above, these pilot experiments were single-shot negotiations.

In fact, looking at the top row in Table 2, there seems to be a substitutability of flat wages for incentive bonuses, in inducing high effort levels. The group at the upper left negotiated high flat wages and significant bonuses – the latter averaging a little more than half the theoretical level that should induce high effort. The group at the upper right negotiated low flat wages and higher bonuses.

[Insert Table 2 about here.]

This outcome, achieved by those who first negotiated a high flat wage and then delivered a high effort level, in fact *dominated the one predicted by principal-agency theory*.³ The programmers, “insured” by compensation that was primarily in the form of a flat wage, nevertheless supplied optimal levels of effort.

We confirm these results by estimating the following equation:

$$\text{ROUTINE} = \beta_0 + \beta_1 * \text{FLATWAGE} + \beta_2 * \text{BONUS} + \varepsilon_i$$

Here, we estimate the likelihood of a pair's interaction resulting in routine effort (as opposed to a high effort), as a function of both the flat wage and bonus set. Clearly, the revealed likelihood of routine effort in this game is low, suggesting that most pairs resulted in high effort regardless of the "flat wage-bonus" combination of incentives. We account for the relative paucity of routine effort in this binary dependent variable by estimating this model as a logit specification, with the bias corrections suggested by King and Zeger (*forthcoming*); this method also holds power in small sample environments and includes robust variance calculations. Because the "flat wage-bonus" combination was negotiated simultaneously, we include both in a single estimation equation.

Table 3 provides the results for this model. This model provides no evidence for a direct role for either a flat wage or a bonus in increasing the likelihood of a routine effort on the part of the programmer. While the signs are in the correct direction (increasing either the flat wage or the bonus is expected to decrease the likelihood of a routine effort), both effects fail to attain conventional significance levels. In fact, the computed probability (absolute risk) of seeing a routine effort is smaller (0.131) given the covariates than is its incidence in the sample (0.138) unconditional on any covariates. The 95% confidence interval for this probability is 0.059 and 0.275, indicating that 97.5% of the estimated distribution lies below a 0.275 chance of routine effort. Generally, this statistical model – given information about the wage contract negotiated – is unable to beat the null model incorporating only a constant term, which suggests that a uniform factor is a better predictor for the pairs involved in this experiment. Together these results call into question the critical claim of principal-agency: that routine effort is the norm unless incentives are employed. In this pilot experiment, incentives were not causal *and* high effort was the norm.

[Insert Table 3 about here.]

RECIPROCITY IN PRINCIPAL-AGENT RELATIONSHIPS

A considerable part of our morality and our lives themselves are still permeated with this same atmosphere of the gift, where obligation and liberty intermingle (Mauss 1950, 65).

Let us repeat; the results could not be more of a challenge to principal-agency theory. The message of principal-agency theory is that incentives are necessary to induce costly effort from agents in a context of information asymmetry. But in these experiments, *self-interested risk-averse agents supplied efficient levels despite the absence of the “minimal necessary” level of incentives*. As surprising as these results are, they turn out to be quite consistent with a different experimental literature on reciprocity and trust.⁴

Reciprocal Gift Exchange

An increasingly persuasive model of how social motivation can resolve inefficiencies (including those of the principal-agent game) is a model of reciprocity, or gift exchange (Akerlof 1982; for its most mathematically developed form, see Rabin 1993). In this model, people are assumed to respond to kindness with kindness, and to harmful acts with revenge.

Evidence for reciprocity as a motivator of human behavior is widespread. In anthropology, reciprocity is seen as a norm in many cultures. In hunter-gatherer cultures, much like those humans must have evolved in for 100,000 years or more, reciprocity serves as an efficient method of insurance in a risk-filled environment. A hunter may be successful only one day out of 10; but in a band of a half dozen hunters, some hunter in the band will be successful much more frequently, with more food than he can eat before it spoils. Reciprocity in sharing

the results of the hunt smooths out the consumption levels for everyone in the group (Mauss 1950).

Experiments on Trust

Evidence for reciprocity is increasingly available from controlled laboratory experiments, as well, as ethnographic studies. One of the most compelling recent laboratory experiments speaks directly to the anomalous transcendence of principal-agency's risk-sharing/incentive trade-off. Berg, Dickhaut, and McCabe (1995) ran experiments in which subjects in Room A were anonymously paired with subjects in Room B; these "partners" would remain forever unknown to each other. Subjects in Room A were given \$10, and asked how much of it they would choose to send to their partners in Room B. The amount sent would be tripled by the time it reached Room B, so that each subject could potentially donate \$30 to her partner. Subjects in Room B could then decide how much of the money they received from their partner would be returned to the partner. The experiment was run with a double-blind procedure that kept even the experimenter from being able to identify the decisions of any one subject.

As Berg, *et al.* note, there is one unique subgame perfect equilibrium of this game: subjects in Room A give no money to subjects in Room B, and subjects in Room B return no money to subjects in room A. Yet, 30 of 32 subjects returned an average of \$5.16 of their \$10.00, resulting in an average payback of \$4.66 by Room B.

The authors, all economists, feel that these results raise the possibility that "trust is an economic primitive" (Berg *et al.* 1995, 123). They do not define trust, but we presume that the implicit definition is "*the belief that another will reciprocate a beneficent act not motivated by short-term self-interest (a gift).*" There is a bit of a paradox here, because once one has this belief, actions that would not be self-interested are arguably rational. For instance, if the Room

A subject believes that the Room B subject will reciprocate her gift, then it is rational to make the gift. This belief on the part of the Room A subject does not explain why the Room B subject in fact acts in a way that is consistent with the belief. On the contrary, Room A's belief in Room B's trustworthiness must be grounded in past experience of trustworthiness.

Berg, *et al.* in fact offer evidence that social history does reinforce or discourage trustworthiness. They replicated the original experiment with 28 couples who had complete reports on the decisions made by the original 32 couples – number of Room A contributors, amounts of contribution, and number and amounts of Room B responses. The median Room A contribution was identical at about \$5.00, but the level of trustworthiness increased sharply – the median payback was \$8. Furthermore, there was a stronger correlation between original contributions and paybacks. The Room A contribution of \$5 had an average payback of \$7.14 and the \$10 contributions had an average payback of \$13.17. “Taken together our two treatments provide a strong rejection of the subgame perfect prediction that Room A subjects will send no money” (137). Although each player was in a one-shot transaction, the social history reinforced trust and enhanced trustworthiness.

Reciprocity in Labor Contracts

Akerlof (1982) argues that gift exchange also occurs in labor contracts. His evidence is from firms in which employees supply unmonitored levels of effort and receive in return wages that are higher than market levels. In formalizing this idea, Rabin (1993) modifies the normal economic self-interest assumptions by assuming that people want to do well to people who help them and to punish those who harm them, and that their willingness to do so increases with a decrease in the material cost. He establishes the existence of “fairness” equilibria that depend on each player's beliefs about the other's intentions. In a Prisoner's Dilemma game, in which

the temptation to defect is not too great, mutual cooperation may be a “fairness” equilibrium because of the benefit each player gets from the mutually reinforced belief in reciprocated goodwill. One of Rabin’s applications is to Akerlof’s “gift exchange” model of worker/firm labor exchange.

Trust experiments like Berg, *et al.* (1995) have potentially great implications for principal-agency relationships. Some of these implications are clearly seen in the experiments by Fehr, *et al.* (1993; 1999a). These experiments are framed as a labor contract – an exchange of compensation for costly “effort”. Fehr’s experiments begin with an “employer” offering compensation along with a proposed level of effort. If the worker accepts the contract, he immediately and automatically gets the promised compensation, no matter what level of effort he actually provides. There is no opportunity for the labor buyer to get her money back, or punish the worker for shirking. In this situation, the subgame perfect equilibrium is for the worker to supply only the minimal level of effort, and for the employer to assume that she will receive only the minimum and pay accordingly. Any compensation beyond the minimum is a manifestation of “trust” in the same sense as the Berg experiment, and any effort level beyond the minimum is a manifestation of trustworthiness, in the sense of reciprocating the buyer’s trust.

Like Berg, Fehr, *et al.* find that employers have enough trust to offer more than the minimal, equilibrium compensation – and workers mostly honor that trust. On average, employers gave workers, *ex ante*, 42% of the surplus generated by the exchange. In response, workers chose the minimal effort level ($e = 0.1$) only 16% of the time. Furthermore, there was a positive effect of the premium offered on effort supplied. This arguably demonstrated a sense of reciprocity, not altruism. That is, employees did not unilaterally do all they could to make their employers better off. Instead, they systematically adjusted their voluntary efforts on the

employers' behalf in response to the level of the "gift" they had already received from the employer. It is a striking confirmation of Akerlof's "gift exchange" theory of employer/worker interactions (1982). As argued by Akerlof, this gift exchange is significant in scale to result in involuntary unemployment – that is, employers paying wages higher than market-clearing wages, resulting excess supply of workers.

It is also striking because, like the negotiation experiments reported in this paper, the efficiency-enhancing implicit gift-exchange occurred without repeated play; and unlike the negotiation experiments reported in this paper, the efficiency improvements were achieved *without discussion of any kind*. The result carries over to collective action settings, in which Fehr and Gächter (1999b) introduces the possibility of punishment (costly for the punisher as well as the punishee). Although normal economic assumptions would make use of the punishment a non-credible threat from the standpoint of standard game theory, the belief in reciprocity is sufficient to create nearly perfect cooperation in such settings (1999).

Security for Effort: The Canonical Gift Exchange

As Mauss emphasized in his classic anthropological study The Gift (1950), gift exchange has a paradoxical combination of voluntarism and obligation. There is no formal, enforceable requirement of a return gift. However, the recipient experiences a social obligation: one's standing in the eyes of the donor (and others in a social group) requires a return. Thus, as Mauss acknowledges, giving a gift can be a calculated, self-interested, strategic act, designed to put the recipient in the position of returning the gift at a time and in a form that is advantageous to the original donor. In the typical "big man" societies, this strategic form of gift-giving is the highest form of political self-advancement for ambitious men. Gift-giving is calculated to create a network of obligation that will guarantee social position for the giver.

Gift exchange in principal-agent relationships can certainly have the same calculated, strategic role. To use the principal-agent negotiation as an example, the strategic principal may well refuse to negotiate an expensive \$100,000 bonus for the programmer, and instead offer a smaller flat wage, which nevertheless carries with it the generous gift of financial security for the programmer – and the expectation of high effort in return.

The obligations that go with gift exchange are a potential solution to the imperfect contracting that follows from information asymmetry. That is, since the programmer's effort level is not observable, it cannot be contracted on or enforced. But, with a gift, a high effort level can be made to be an obligation of the programmer. In light of the trust experiments of Berg, *et al.*, Fehr, *et al.*, and others not reviewed here (e.g., Kollock 1994), the results of the principal-agency negotiations reported in this paper are much less surprising.

Among the MBA subjects in the principal-agent negotiation, the programmers who supplied a high effort level with a small bonus received a high flat wage. From the perspective of principal-agent theory, this flat wage should not have been effective in inducing a high effort level. After all, the flat wage was to be paid even in the event of a program failure; and the owner was never to know whether or not the programmer had supplied a high effort.

But from the perspective of the literature on reciprocity and gift exchange, however, the outcome was quite consistent with the notion that an implicit (though totally unenforceable) bargain had been struck. High effort had been traded for a high flat wage. Norms of reciprocity would require that agents who had successfully negotiated “insurance” in the form of a high flat wage would respond with high effort levels.

The behavior observed in the experiments reported in this paper, like Fehr's, can be understood as an example of Rabin's “fairness” equilibrium. The only extension is that in the

Akerlof/Rabin explication, extraordinary levels of effort are reciprocated with extraordinary shares of revenue. In our results, firms reciprocate primarily with risk-bearing rather than revenue.

CROWDING OUT: THE INCONSISTENCY OF COOPERATION AND CONTROL

A high wage will not elicit effective work from those who feel themselves outcasts and slaves, nor a low wage preclude it from those who feel themselves part of a community of free men (Robertson 1921, 244).

At one level, the theme of this paper might be thought of as using principal-agency theory as a straw man. After all, taken in its most general form, the principal-agency literature is simply suggesting that the principal can contract with the agent in such a way that she (the principal) can be confident that the agent will accomplish a task that the principal has set for him. In some cases, the principal may do this by offering financial incentives closely tied to outcome. In others, the principal may do this by close monitoring individual effort levels. In still other settings, the principal may solve her problem by working to instill a sense of trust between the agent and herself. Principal-agency may be defined broadly enough to encompass all such forms of superior/subordinate relationships.

And of course, such a perspective on principal-agency theory is unanswerable – because it is empty of content. Any hierarchical relationship – more or less formally structured, more or less successful – is a principal-agent relationship. It is, after all, not difficult to explain why an altruistic agent would take costly actions that make someone else better off. Principal-agency theory, as Moe pointed out in his early explication of the theory, was an interesting problem precisely because the agent was assumed to be *financially self-interested*. Furthermore, the

classic principal-agency theory was concerned with the efficacy of financial incentives *linked to outcomes*, not to effort levels. Once again, it is not hard to write a sufficiently motivating contract linked to effort if effort is costlessly observable. (The evocative phrase for this is a “boil in oil” contract—provide the specified effort level or suffer the dire consequences.) But as Holmstrom argued in his classic “Moral Hazard in Teams”, the effort needed to monitor any agent with an interesting task (like managing a multi-million dollar corporation, or the Department of Defense) is prohibitively expensive (1982). The purpose of principal-agent theory was to examine the possibility of replacing onerous and costly monitoring by the manipulation of self-interest through financial incentives.

Thus, non-vacuous principal-agency theory is precisely about the use of financial incentives, linked to outcomes rather than efforts, offered to self-interested agents. Such a theory leads us inevitably to conclusions such as that one that motivated the experiments in this paper: it is necessary to trade-off efficiency in risk-sharing for efficiency in incentives. And this conclusion, while it makes a non-vacuous, empirically testable statement, is found to be empirically false.

Furthermore, principal-agency, narrowly defined, would seem therefore to be an ineffective guide to managerial behavior. A manager who was sufficiently trained in principal-agency theory would be led to negotiate an outcome-linked bonus of \$100,000 or more, depending on the level of the agent’s risk aversion. However, a manager, blissfully ignorant of principal-agency theory, could negotiate a less costly flat and expect the agent to reciprocate this gift of insurance with a high (non-observable) effort level.

Put even more strongly, the principal-agency emphasis on outcome-based bonuses, forcing an unwanted risk on a risk-averse agent, may “crowd out” the kind of reciprocal gift-giving that makes the Pareto-preferred outcome possible.

The Literature on “Crowding Out”

The concept of “crowding out” can be illustrated with the example of contributing blood to a blood drive. Clearly, some people are willing to go to the pain and trouble of donating blood for no monetary gain. This is not irrational or inconsistent with the marginal analysis of standard microeconomics. These people must derive some form of intrinsic or social motivation that makes it worthwhile for them to sacrifice an hour and endure a minor amount of pain (as in Titmuss 1971).

But the problem quickly becomes more profound when one asks whether one can *increase* the amount of blood that is donated by adding a financial incentive. Economics clearly has only one answer to this question. Some people with no intrinsic or social motivation should now find it worthwhile to give blood. And those who were willing to give blood for free should certainly still find it worthwhile to give blood in exchange for their original non-pecuniary motivation *plus* the financial incentive.

However, this may not be true (Frey 1999). Paying some blood donors puts the transaction on a market basis, and thereby *decreases* the intrinsic and social motivation for donation. A potential volunteer blood donor who would give blood for free *might not give blood for \$10*. If she finds the financial incentive inadequate, she could also feel that the \$10 inducement has eliminated or diminished the social and intrinsic rewards for donation. A person who is paid for giving blood receives none of the special social status that groups typically allocate to those who make voluntary, costly contributions to group public goods. If the non-

pecuniary motivation is endogenous, and a function of the pecuniary motivation, then pecuniary rewards could *crowd out* other motivations for giving blood (Frey 1999). The net result could be that paying donors for their blood would elicit fewer blood donors than keeping it on a purely voluntary basis.

The hoodlum John Dillinger is supposed to have said, of robbing banks, “You can get more cooperation with a smile and a gun than you can with just a gun.” But of course, that statement is false, if we think of cooperation as anything more than minimal compliance. Producing a gun in a bank produces the good soldier Schweik syndrome – doing exactly what the gunman orders, and nothing more – and when one can do so safely, sabotaging the gunman’s intent by stepping on a silent alarm.

The reasons for crowding out are not too difficult to imagine, if more difficult to prove. As far as intrinsic motivation, the message that “If you do X, you will be rewarded with Y” sends the message that Y is pleasant, and X must be something unpleasant. Consequently, as John Nichols has argued, a plan to reward children who read books with pizzas will likely result in fat children who hate to read (1989). Better, he argues, to offer to reward pizza-eating with books.

Similarly, financial rewards can diminish social motivation. When a donor accepts a financial reward, she completes an exchange and ends the sense of obligation that is felt toward someone who gives a gift.

Thus, the case can be made that in the principal-agent negotiation discussed in this paper, the flat wage and the bonus have quite different motivational features. An owner who tries to negotiate a bonus is sending the message that the owner thinks the programmer requires an outcome-linked bonus to work hard – that is, the owner expects the programmer to shirk if a sufficiently high bonus is not negotiated. Whatever size bonus is negotiated then, it does not

have the nature of a gift, but of a *quid pro quo*, and it carries with it no social expectation of reciprocity. On the other hand, the advantage of the flat wage is precisely that both sides recognize that it has *no* incentive effect. A large flat wage is therefore a gift of security from the owner to the programmer, and carries with it a social obligation – one that can only be reciprocated with high effort level.

An important thing to notice about this gift exchange is that it is fragile. A negotiation that emphasizes the necessity of a high incentive bonus eliminates the sense of social obligation, and therefore crowds out the social motivation of reciprocity.

Financial incentives are thus a substitute (not a complement) for social motivation in a principal-agency relationship. The disadvantage of financial incentives is that the best that can be accomplished with financial incentives is limited by the tradeoff between efficient incentives and efficient risk-sharing. However, social motivation in the form of reciprocal gift-giving (effort for security) can achieve outcomes that are Pareto-superior to those available through financial incentives.

RECIPROCAL GIFT-GIVING IN THE WORKPLACE

Is there evidence that the kind of reciprocity evidenced in the Berg, Fehr, and negotiated principal-agent experiments can be found in the workplace? The kind of implicit contract we are thinking of is the exchange of security for risk-averse agents in exchange for unmonitorable but valuable efforts.

As the economist Stiglitz (1987) has pointed out, many firms do in fact force employees to bear an enormous risk – the risk associated with the business cycle. When demand drops in a

recession, inventories build up, and firms almost inevitably lay off employees. The employees end up bearing the risk for an outcome (a recession) for which they are in no way responsible. As a result, employees have every reason to “work to rule” – providing the minimal effort defined by rules and supervisory enforcement. They have no reason to provide non-monitorable actions or suggestions that might increase productivity, hence accelerating the day when inventories will build up resulting in their own layoffs.

An exception to the rule is Lincoln Electric Company. Lincoln Electric has had a long-held policy of no layoffs during recessions, despite the fact that its line of products (including electrical generators) is especially vulnerable to business cycle effects. Furthermore, unlike other piece-rate firms, it has a policy of never reducing the piece-rate that it offers its employees, even when the employees end up being compensated far out of line with industry averages (Miller 1992). A similar case is that of Malden Mills, where management continued to employ all workers at their Massachusetts factory during the period during which the factory was rebuilt after a devastating fire.

The net effect of these two forms of risk-bearing is a strong sense of employee loyalty and commitment that is observable in actions that go far beyond the monitorable and contractible aspects of job descriptions. Employees are noted for constant productivity improvements that are due in part to monitorable efforts but also in part to productivity tips that could not be coerced. In exchange for risk-bearing by the firm, employees enhance productivity in ways that are non-contractible.

Security/Commitment Exchanges in Public Management

It is possible to cull similar examples from the literature on public management. In general, the managers who have inspired the most legendary levels of loyalty from subordinates

are those who specifically committed to protect the employees from external risks. Those agencies that have inspired the most legendary levels of bureaucratic indolence are those where employees feel most exposed to external risks.

Robert Moses was by all accounts a demanding, even arrogant manager, driving his employees hard and expecting complete loyalty. What he offered them, in part, was his own protection:

The rewards Moses offered his men were not only power and money. If they gave him loyalty, he returned it manyfold Moses might criticize his men himself, but if an outsider tried it – even if the outsider was right, and Moses privately told his aide so – Moses would public defend him without qualification (Caro 1974, 273).

In other words, Moses offered to bear a large amount of risk himself, in exchange for a high level of effort and commitment from his subordinates.

While the exchange of security for commitment can enhance organizational performance, the absence of such an implicit contract can detract from organizational performance. Evidently the State Department, led as it is by a transitory and often distant Secretary of State, generally leaves its Foreign Service officers feeling exposed to political risks from Congress. Members of Congress, not the least of them Joseph McCarthy, have often found it politically rewarding to take the Department, and individual officials at State, to task. The result of this risk is a high level of defensive behavior, in which Foreign Service officers will do nothing that is not clearly demanded by written rules or formal hierarchical orders. Warwick (1975) provides an account of an attempt to improve State Department performance by decreasing stifling hierarchy and rules. Despite some initial indication of success, subordinates soon demanded (and got) increased cover in the form of re-instated levels of hierarchy and increasingly explicit rules. In

the absence of a managerial leader who can credibly commit to protect her subordinates, risk-averse bureaucrats will revert to unproductive but safe “working to rule”.

Scholz provides evidence of a similar kind of reciprocity in regulatory regimes, between regulatory agencies and the firms they are charged with regulating (1984, 1991). Regulatory agencies have a dominant strategy to engage in coercive, maximal enforcement, to which firms have every incentive to respond with minimal compliance (1991: 118). An exchange of information and voluntary compliance for flexible enforcement from the agency can enhance the effectiveness of regulation while reducing the costs to the firm. Scholz offers evidence that there can be a payoff in terms of reduced workplace injuries from such cooperative arrangements between state-level regulatory agencies and firms. He also shows, however, that a concern for political control by political beneficiary groups can trigger a more rigid, less cooperative regulatory style. An increased reliance on coercive regulation “crowds out” any incentive for cooperation from firms; in this case, a smile and a gun induce less cooperation from regulated firms than a smile alone.

We offer one final point on public management in the context of Brehm and Gates’ “principled bureaucrats” (1997). The evidence they offer to support their argument that selection rules choose the right type of bureaucrats – and so reduce problems of moral hazard – fits with our claim. Bureaucrats may be engaging in reciprocal gift exchanges, so that one finds a confluence of civil service protections and bureaucrats working harder (doing more of what they’re supposed to be doing) than expected. If so, the mechanism for obtaining bureaucratic compliance with political goals is neither the selection mechanism nor the selected agents, it is the gift exchange.

Further Tests

This paper has reported a set of experimental results that are not consistent with traditional principal-agency theory, but do seem consistent with recent experimental results motivated by an interest in reciprocity. Because the experiments were designed with the expectation that principal-agency theory would prove valid, they are inadequate as a rigorous test of any model of reciprocity such as Rabin's. Consequently, we offer a research agenda intended to test reciprocity against traditional principal-agency theory.

We see the important components of such an agenda to include the following. Foremost, this agenda would test comparative statics results from competing models by way of multiple treatments where only key parameters varied. At a minimum, agent efficacy, or the probability of firm success associated with high versus low effort, should vary. Traditional principal-agency theory implies that the outcome-based bonus necessary to insure high agent effort is inversely proportional to agent efficacy. However, the existence of a Rabin "fairness" equilibrium based on reciprocity is not particularly sensitive to that variable.

Second, principal-agency theory implies that the necessary bonus to insure high agent effort is positively associated with the agent's *marginal* agent effort cost, but not the absolute level of effort cost. If low effort costs \$20,000 and high effort \$40,000, that is the same marginal effort cost (and hence the same required bonus) as when low effort costs \$50,000 and high effort \$70,000. However, higher absolute agent effort cost may lead to increased agent concerns about the possibility of ending up with a negative net payoff; if risk aversion increases with agent effort cost, increased effort cost may in fact *decrease* the use of bonuses.

Finally, experimental treatments, such as different communication conditions or informational cues, may be used to manipulate subject beliefs about the other's intentions. Measurement of subject beliefs could serve as a manipulation check; if the likelihood of

negotiating relatively efficient flat wage compensation schemes is positively related to beliefs about the other's intentions, that would serve to verify Rabin's fairness equilibrium as opposed to traditional principal-agent theory.

CONCLUSION

The cooperative system must create a surplus of satisfaction to be efficient. If each man gets back only what he puts in, there is no incentive, that is, no net satisfaction for him in cooperation. What he gets back must give him advantage in terms of satisfaction; which almost always means return in a different form from that which he contributes. (Barnard 1938, 58)

This paper has proposed a somewhat different perspective on trust in principal-agent relationships than in previous writings on organizational economics. In Miller (1992), Holmstrom's impossibility result (1982) was interpreted as establishing limits on the ability of rational individuals to design incentive structures that achieve organizational efficiency, at least as one-shot games. In repeated games with long time horizons, rational individuals *may* achieve efficient equilibria, but coordinating on a particular equilibrium is the primary obstacle. This dilemma is used to justify organizational *leadership* and other coordination devices. *Trust* is interpreted as the shared belief by members of a hierarchical organization that everyone is playing the same, Pareto-preferred cooperative equilibrium. In repeated games, coordinative trust requires no special violation of individual rationality, as rational individuals will (by definition of a Nash equilibrium) play cooperatively, productively, and honestly once coordination on the appropriate Nash equilibrium is achieved. Miller (1992) is therefore a minimalist revision of rational choice organizational economics.

In this paper, however, there is no repeated game being played, so *trust* cannot be interpreted as a shared expectation regarding a coordinated equilibrium. In a one-shot

negotiation, *trust* is interpreted (in a way consistent with Rabin 1993) as the expectation that the trusted player will reciprocate a one-shot kindness with an appropriate response; and *trustworthiness* is interpreted as the obligation on the part of the trusted agent to reciprocate in this way. Note that *trust* is not a violation of rational choice, as long as the belief in reciprocity is empirically validated. Perfectly selfish individuals can trust (even manipulate!) other players by strategically giving them gifts, when they expect to achieve a net gain from the counter-gift. But why the trusted player reciprocates is a more difficult question. It requires, we believe, some (possibly modest) change in the characterization of the trusted person's utility function, which rationalizes a final-play, costly act.

This paper can also be distinguished from other works about reciprocity in employment transactions by Akerlof (1982) and Fehr, *et al.* (1993). In those cases, the mutual exchange consists of a higher level of rewards to the employee in exchange for higher levels of non-contractual effort. This paper conceives of the exchange as one that explicitly addresses the core trade-off of principal-agency theory – efficiency in risk-sharing for efficiency in incentives. The hypothetical exchange consists of a more complete bearing of the employee's risk in exchange for higher levels of non-contractual effort.

In a setting of public bureaucracy, we may suppose that this possibility reintroduces a range of managerial considerations that are secondary or omitted altogether from principal-agency theory. Foremost might be the creation and communication of reciprocity norms, combined with the selection of public employees that are willing and able to engage such norms in a productive way. For the political study of public bureaucracy, this reintroduces historical concerns about bureaucratic "types" (Downs 1967), cooperation in a hierarchy (Barnard 1938), and professionalism and its obligations (Mosher 1968).

Furthermore, the results suggest that leadership that emphasizes reciprocity with subordinates may well be more effective than a leadership style that starts from the assumption that financial incentives are the only means for motivating subordinates. The latter style of leadership will get just what it pays for, from risk-averse agents who are troubled by the burden of unwanted risk. The former style of leadership may be able to instigate a gift exchange of security for effort that leads to non-contractible levels of innovation and commitment in bureaucracy.

**TECHNICAL APPENDIX A:
A SIMPLE GAME-THEORETIC MODEL OF
THE INSURANCE/INCENTIVE TRADE-OFF**

The primary hypothesis to come out of principal-agency theory is the trade-off between efficiency in risk-sharing and efficiency in incentives. With a risk-averse agent, the best use of incentives leaves both principal and risk-averse agent worse off than trust (Kreps 1990: 585). This appendix restates the example (from Dixit and Nalebuff 1991, 302-306) with an extensive-form game.

In this simplified game, the principal moves first and may provide either a Flat Wage or a Bonus. The agent can choose either to exit, provide low effort, or high effort. At that point, Nature decides whether to provide a Success (worth \$200,000) or a Failure (worth zero). The probability of success is 80% with high effort, 60% with a low effort. These probabilities are summarized as expected payoffs at each terminus for the agent's possible actions.

The principal's best flat wage is $\mathbf{W} = \$50,000$. The reason for this is that with $\mathbf{W} < \$50,000$, the agent will choose to exit rather than provide a low effort due to effort cost; but the principal wants the agent to supply at least a low effort, because the expected payoff (60% of \$200,000 = \$120,000) is more than \$50,000. On the other hand, by backward induction, $\mathbf{W} > \$50,000$ will not result in a high effort, since (by definition) the flat wage is not outcome-contingent. With $\mathbf{W} = \$50,000$, the owner gets an expected bonus of $0.6 * \$200,000 - \$50,000 = \$70,000$.

For the principal, the best bonus is the one that will induce the agent to provide high effort. By backward induction, this means that $0.8 * \mathbf{B} - \$70,000$ must be greater than or equal to $0.6 * \mathbf{B} - \$50,000$ for a risk-neutral agent. Solving for this inequality reveals that the necessary bonus for a risk-neutral agent must be greater than \$100,000. This satisfies two other

requirements. The agent prefers a 80% chance at such a bonus to exit (individual rationality); and the principal can afford to such a bonus and still make a profit greater than that available with a flat wage, as long as:

$$0.8*(200,000-\mathbf{B}) > \$70,000$$

(or the principal's expected profit with $\mathbf{W} = \$50,000$), which implies that the principal will not pay a bonus greater than \$112,500.

However, by definition, a risk-averse agent will not be indifferent between a marginal effort cost of \$20,000 and an extra 20% chance at a bonus of \$100,000. The risk-averse agent will require a bonus \mathbf{B} that incorporates a risk premium. The more risk averse the agent, the bigger the necessary premium.

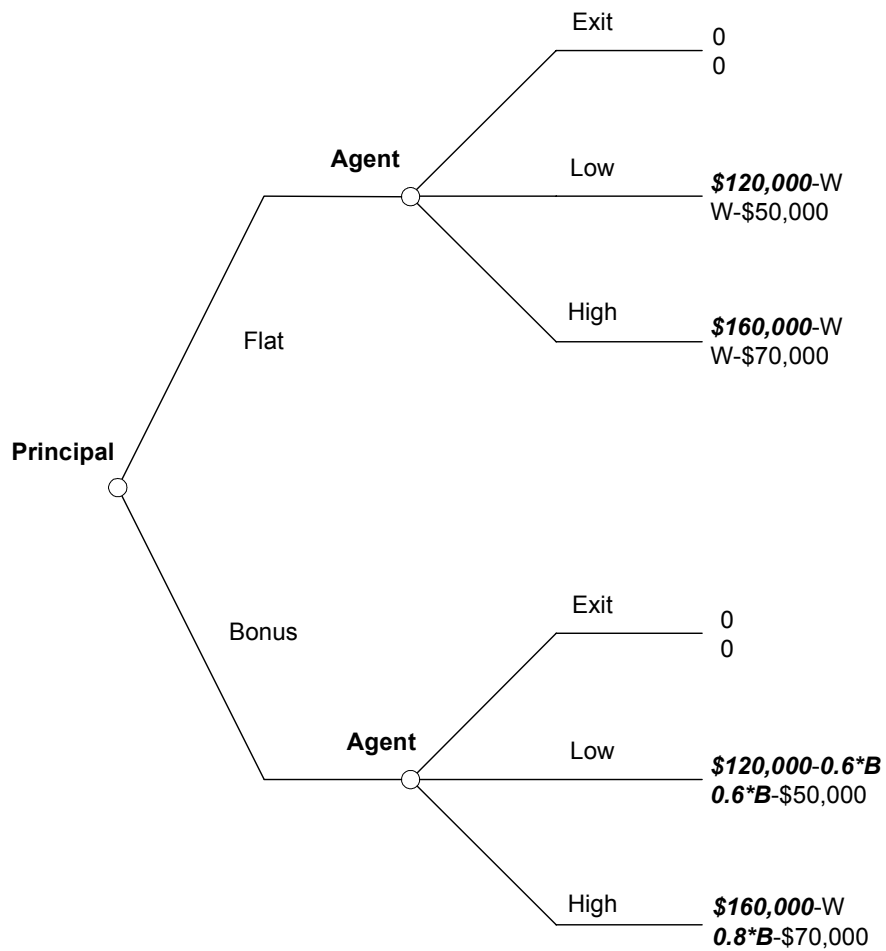
Obviously, if the agent is so risk averse that a bonus of more than \$112,500 is necessary to compensate him for the risk of outcome-based compensation, then there is *no possibility* of agreement. The principal would be better off simply paying a flat wage of \$50,000 and earning expected profits of \$70,000. But let us assume that a bonus of \$110,000 would induce a particularly risk averse agent to provide a high effort. (A standard utility function to illustrate risk aversion involves the square root of net payments; see the note at bottom of Figure A2). With any \mathbf{W} and $\mathbf{B} = \$110,000$, the following is an equilibrium of the game: the agent chooses low effort with a flat wage and high effort with the bonus, and the principal choose to use the bonus.

However, our point is that this equilibrium is Pareto suboptimal. As is generally the case, there is an outcome involving high effort and a fixed wage which would make both sides better off – the agent by reducing risk, the principal by reducing expected incentive payment. In Figure A1, the \$110,000 bonus is compared to a flat wage of \$85,000. If the agent could be trusted to

provide high effort with the flat wage, then the principal would be better off because a flat payment of \$85,000 is less expensive than the expected payout of \$88,000 with the bonus. The agent would be better off because the flat wage is more secure than the bonus, which includes a 20% chance of no payment. While the flat payment/high effort outcome is Pareto-superior to the equilibrium outcome, it is not itself an equilibrium. The agent would still have an incentive to provide low effort, and the principal would have to anticipate that and provide the more expensive bonus.

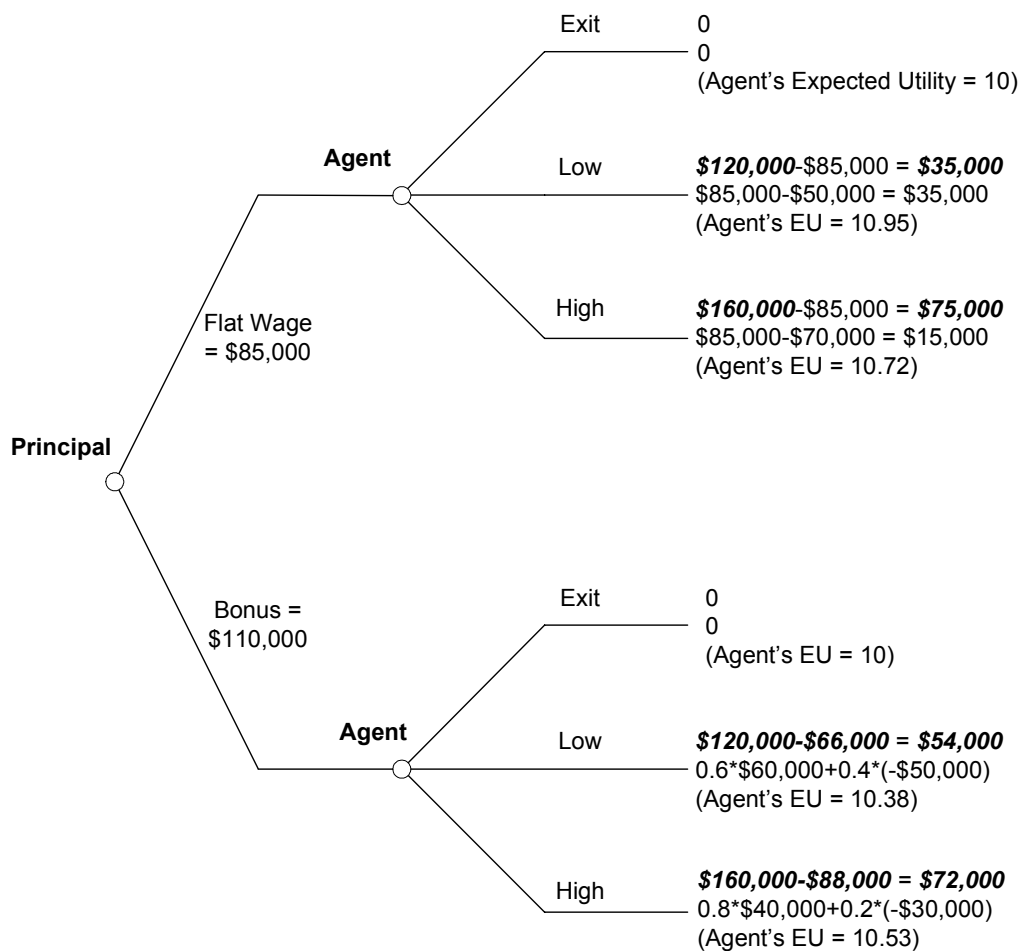
While Figure A2 assumes just two particular incentive schemes for a particular risk-averse agent, the main point is generally true: *for every incentive scheme that induces a risk-averse agent to make additional efficient effort costs, there is a flat wage scheme that both principal and agent would prefer, if the agent could be trusted to provide the same high effort.*

Figure A-1: The Principal-Agent Problem as a Generalized Extensive Form Game



*Payoffs in bold italics expected payoffs, assuming 60% or 80% chance of success, depending on agent effort; other payoffs involve no risk.

Figure A-2: An Illustrative Example with a Risk-Averse Agent*



* Assume the risk-averse agent has utility function:

$$U = (100 + P - C)^{1/2}.$$

P is the payment (W or B) in thousands, and C is effort cost for high or low effort in thousands.

TECHNICAL APPENDIX B: EXPERIMENTAL INSTRUCTIONS

Subjects were randomly paired. In each pair, one subject was randomly assigned to be “owner”, one the “programmer”. Each received the appropriate instructions, which were also read aloud. Each pair received a “contract”. The parameters are identical to those in Dixit and Nalebuff (1991). After negotiating a contract, the agent was given an opportunity to determine effort level in secret; the random element determined that each owner would never be able to deduce the agent’s effort level.

This was one in a series of classroom exercises for which subjects received participation points counting toward their grade in a required MBA strategy course. The points they earned were directly proportional to their net earnings in each of the exercises. Students took the exercises quite seriously, and entered into them with every sign of concentration and vigorous pursuit of self-interest.

We reproduce the instructions for our pilot experiment:

INSTRUCTIONS FOR OWNER **WIZARD CHESS EXERCISE**

You are the owner of a high-tech company in California trying to develop and market a new computer chess game, Wizard 1.0. If you succeed, you will make a profit of \$200,000 from the sales. If you fail, you make nothing. Success or failure hinges in part on what your expert programmer does. She can either put her heart and soul into the work, or just give it a routine shot. With high-quality effort, the chances of success are 80 percent; with routine effort, the figure drops to 60 percent.

From past experience, you have a rather good estimate of the opportunity cost of the programmer. The opportunity cost of a routine effort is \$50,000, and for a high-quality effort it is \$70,000.

Your job is to negotiate a contract with the programmer. The contract contains exactly two variables to be negotiated: the FLAT WAGE, which she will be paid no matter whether the program is a success or a failure. The BONUS will be paid only if the program is a success: that is, the BONUS will be paid if and only if you receive the \$200,000. You cannot make either her FLAT WAGE or her BONUS dependent on her effort, since effort is unobservable.

If a contract is not negotiated, both of you will earn nothing. If a contract is negotiated, then you will sign a contract. The programmer will then turn in the contract by herself, indicating whether she has decided to provide a high or routine effort.

Your points for this exercise will be determined by your profits. That is, if your company is a success,

Your points = \$200,000 – FLAT WAGE – BONUS

If your company is a failure,

Your points = - FLAT WAGE

(new page)

INSTRUCTIONS FOR PROGRAMMER

WIZARD CHESS EXERCISE

You are a programmer with the opportunity to program a new game for a high-tech company in California. If the company succeeds, the owner will make a profit of \$200,000 from the sales. Failure will generate no profits. With high-quality effort on the part of the programmer, the chances of success are 80 percent, but with a routine effort, the probability of success is only 30%. The opportunity cost of a routine effort is \$50,000, and for a high-quality effort is \$70,000.

Your job is to negotiate a contract with a FLAT WAGE and a BONUS. The FLAT WAGE will be paid no matter whether the program is a success or a failure. The BONUS will be paid if and only if the owner receives \$200,000. You cannot be paid directly for your effort, since that is unobservable.

If you do not negotiate a contract, neither of you will earn anything. If a contract is negotiated, then you will both sign a contract sheet. **WHEN YOU BRING THE CONTRACT TO THE INSTRUCTOR, THE EFFORT LEVEL MUST NOT BE FILLED OUT; EFFORT MUST BE PICKED IN THE PRESENCE OF THE INSTRUCTOR AND NOT THE OWNER.**

If your company is a success,

Your points = FLAT WAGE + BONUS – OPPORTUNITY COST

If your company is a failure,

Your points = FLAT WAGE – OPPORTUNITY COST

(new page)

CONTRACT

The PROGRAMMER agrees to develop the program for the Wizard 1.0. In return, the OWNER agrees to pay the PROGRAMMER based on the following terms:

FLAT WAGE, to be paid whether or not the program is a success: _____

BONUS, to be paid if and only if the program is a success: _____

Signed:

_____ Owner's name _____ Owner's player #

_____ Programmer's name _____ Programmer's player #

The owner will select six numbers from 1 to 10. These will be the OWNER'S numbers. The programmer will select two additional numbers from 1 to 10. These will be the PROGRAMMER'S numbers. After the programmer selects an effort level (in secret), the

programmer will select one card from a deck with 10 cards numbered one through 10. If the number picked is one of the OWNER's numbers, then the company is a success. If the number picked is one of the PROGRAMMER's numbers, then the OWNER will earn \$200,000 only if the programmer has supplied a high effort. If the number picked is neither a programmer's number or the owner's number, then the company will earn nothing, no matter what level of effort supplied by the programmer. The number selected will not be made public, so owners will never know whether the programmer supplied a high effort or a routine effort.

The owner should circle six of these numbers:

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10

Now the programmer should put a triangle around an additional two numbers from the list above.

THE FOLLOWING MUST BE FILLED OUT IN THE PRESENCE OF THE PROGRAMMER AND EXPERIMENTER ONLY:

Programmer's effort decision: _____ High Effort
_____ Low Effort

**Table 1: Effort Levels by Negotiated Bonuses in
Experimental Principal-Agent Negotiations**

	<\$49,000	\$50,000 to \$99,000	\$100,000	
HIGH EFFORT	14	22	14	50
LOW EFFORT	4	3	1	8
	18	25	15	58

Table 2: Flat Wages and Bonuses Negotiated by Shirkers and Workers

	Bonus < \$100,000	Bonus > 100,000
High Effort	Flat wage = \$56,541 Bonus = \$57,622 N=36	Flat wage = \$18,571 Bonus = \$113,357 N=14
Low Effort	Flat wage = \$58,142 Bonus = \$42,679 N=7	Flat wage = \$50,000 Bonus = \$100,000 N=1

Table 3: Rare Events Logit Estimates

Variable	Coefficient	Robust SE
Flat Wage	-0.0111	0.0280
Bonus	-0.0286	0.0250
Constant	0.6644	2.7618

REFERENCES

- Akerlof, George. 1982. "Labor Contracts as Partial Gift Exchange." *Quarterly Journal of Economics*. 97:543-69.
- Barnard, Chester I. 1938. *The Functions of the Executive*. Cambridge, MA: Harvard University Press.
- Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior*. 10:122-42.
- Berle, Adolf A. and Gardiner C. Means. 1932. *The Modern Corporation and Private Property*. New York: Macmillan.
- Blau, Peter. 1964. *The Dynamics of Bureaucracy*. Chicago: University of Chicago Press.
- Bottom, William P. 1998. "Negotiator Risk: Sources of Uncertainty and the Impact of Reference Points." *Organizational Behavior and Human Decision Processes*. 76:89-112.
- Braithwaite, Valerie and Margaret Levi. 1998. *Trust and Governance*. New York: Russell Sage.
- Brehm, John and Scott Gates. 1997. *Working, Shirking, and Sabotage: Bureaucratic Response to a Democratic Public*. Ann Arbor: Michigan University Press.
- Cook, Karen S. 2001. *Trust in Society*. New York: Russell Sage
- Dixit, Avinash and Barry Nalebuff. 1991. *Thinking Strategically*. New York: W.W. Norton.
- Downs, Anthony. 1967. *Inside Bureaucracy*. Boston: Little, Brown.
- Fama, Eugene. 1980. "Agency Problems and the Theory of the Firm." *Journal of Political Economy*. 88:288-307.
- Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl. 1993. "Does Fairness Prevent Market Clearing?" *The Quarterly Journal of Economics*. 108: 437-59.
- Fehr, Ernst, and Armin Falk. 1999. "Wage Rigidity in a Competitive Incomplete Contract Market." *Journal of Political Economy*. 107:106-134.
- Fehr, Ernst and Simon Gächter. 1999. "Cooperation and Punishment in Public Goods Experiments." University of Zurich. Working Paper No. 10.
- Fiorina, Morris. 1981. "Congressional Control of the Bureaucracy: A Mismatch of Incentives and Capabilities". In *Congress Reconsidered*. Ed. Lawrence C. Dodd and Bruce I. Oppenheimer. Washington: Congressional Quarterly.

- Frey, Bruno S. 1999. "Institutions and Morale: The Crowding Out Effect". In *Economics, Values, and Organization*. Ed. Avner Ben-Ner and Louis Putterman. New York: Cambridge University Press.
- Goodnough, Amy. 2001. "Strains of Fourth-Grade Tests Drives Off Veteran Teachers." *New York Times*. Page A1.
- Holmstrom, Bengt. 1982. "Moral Hazard in Teams." *Bell Journal of Economics*. 13:324-340.
- King, Gary and Langche Zeng. *forthcoming*. "Logistic Regression in Rare Events Data." *Political Analysis*.
- Kollock, Peter. 1994. "The Emergence of Exchange Structures: An Experimental Study of Uncertainty, Commitment, and Trust". *American Journal of Sociology*. 100:313-45.
- Kormendi, Roger and Charles R. Plott. 1982. "Committee Decisions under Alternative Procedural Rules". *Journal of Economic Behaviour and Organization*. 3:175-95.
- Lipsky, Michael. 1980. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russell Sage.
- Mauss, Marcel. 1950. *The Gift: The Form and Reason for Exchange in Archaic Societies*. New York: W.W. Norton.
- McLean Parks, Judy and Edward J. Conlon. 1995. "Compensation Contracts: Do Agency Theory Assumptions Predict Negotiated Agreements?" *Academy of Management Journal*. 38:821-38.
- Milgrom, Paul and John Roberts. 1992. *Economics, Organization, and Management*. Englewood Cliffs: Prentice-Hall.
- Miller, Gary. 1992. *Managerial Dilemmas: The Political Economy of Hierarchy*. Cambridge, New York, and Melbourne: Cambridge University Press.
- Mosher, Frederick C. 1968. *Democracy and the Public Service*. New York: Oxford University Press.
- Mirrlees, James. 1976. "The Optimal Structure of Incentives and Authority within an Organization." *The Bell Journal of Economics*. 7:105-31.
- Moe, Terry. 1984. "The New Economics of Organization". *American Journal of Political Science*. 28:739-77.
- Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics". *American Economic Review*. 83:1281-1302.

- Rich, Jue T. and John Larson. 1987. "Why Some Long-Term Incentives Fail". In *Incentives, Cooperation and Risk Sharing*. Ed. by Haig Nalbantian. Totowa: Rowan & Littlefield.
- Robertson, D.H. 1921. "Economic Incentives." *Economica*. Vol. 1 (October).
- Ross, Stephen A. 1973. "The Economic Theory of Agency: The Principal's Problem." *American Economic Review*. 63:134-139.
- Rothe, Harold F. 1970. "Output Rates Among Welders: Productivity and Consistency Following Removal of a Financial Incentive System." *Journal of Applied Psychology* 54: 549-51.
- Scholz, John T. 1984. "Cooperation, Deterrence, and the Ecology of Regulatory Enforcement." *Law and Society Review*. 18:601-46.
- Scholz, John T. 1991. "Cooperative Regulatory Enforcement and the Politics of Administrative Effectiveness." *American Political Science Review*. 85: 115-136.
- Shavell, S. 1979. "Risk Sharing and Incentives in the Principal and Agent Relationship." *Bell Journal of Economics*. 10:55-73.
- Stiglitz, Joseph E. 1987. "The Design of Labor Contracts: The Economics of Incentives and Risk Sharing". In *Incentives, Cooperation, and Risk Sharing: Economic and Psychological Perspectives on Employment Contracts*. Ed. by Haig R. Nalbantian. Totowa, NJ: Rowman and Littlefield. Pp. 47-68.
- Titmuss, Richard Morris. 1971. *The Gift Relationship: From Human Blood to Social Policy*. New York: Pantheon Books.
- Wall Street Journal*. 1988. "All Eyes on DuPont's Incentive Plan." Dec. 5.
- Weingast, Barry. 1984. "The Congressional-Bureaucratic System: A Principal-Agent Perspective with Applications to the SEC." *Public Choice*. 44:147-91.
- Whyte, William F. 1955. *Money and Motivation: An Analysis of Incentives in Industry*. New York: Free Press.
- Wilcox, Melynda Dovel. 2000. "A Crash Diet for Dot-coms." *Kiplinger's Personal Finance Magazine*. 54(8): 20-22.

ENDNOTES

¹ In a number of ways, the language of principal-agency theory has been borrowed by political science without coming to grips with its real meaning. In essence, principal-agency theory has come to be simply a metaphorical synonym for “hierarchical”. This paper achieves two proximate goals: to evaluate the principal-agency’s empirical power in its canonical form (hence the Dixit and Nalebuff example), and to extend the theoretical environment of principal-agency to reincorporate both older and more recent strains of literature with perhaps greater importance for the study of public bureaucracies.

² Actually, the theory does not require strict risk neutrality on the part of the principal—merely *less* risk aversion than for the agent. As long as that is the case, the principal is still the efficient bearer of risk.

³ An outcome “Pareto dominates” another outcome if it is preferred by both principal and agent.

⁴ This paper focuses on an interpretation of trust that is based on reciprocity. For other conceptions of trust, see Braithwaite and Levi (1998) or Cook (2001).