CARROTS THAT LOOK LIKE STICKS:
TOWARD AN UNDERSTANDING OF MULTITASKING INCENTIVE SCHEMES

Omar Al-Ubaydli
Steffen Andersen
Uri Gneezy
John A. List

Carrots that Look Like Sticks: Toward an Understanding of Multitasking Incentive Schemes
Omar Al-Ubaydli, Steffen Andersen, Uri Gneezy, and John A. List
NBER Working Paper No. 18453
October 2012
JEL No. C93,D01

ABSTRACT

Constructing compensation schemes for effort in multi-dimensional tasks is complex, particularly when some dimensions are not easily observable. When incentive schemes contractually reward workers for easily observed measures, such as quantity produced, the standard model predicts that unrewarded dimensions, such as quality, will be neglected. Yet, there remains mixed empirical evidence in favor of this standard principal-agent model prediction. This paper reconciles the literature by using both theory and empirical evidence. The theory outlines conditions under which principals can use a piece rate scheme to induce higher quantity and quality levels than analogous fixed wage schemes. Making use of a series of complementary laboratory and field experiments we show that this effect occurs because the agent is uncertain about the principal's monitoring ability and the principal's choice of a piece rate signals to the agent that she is efficient at monitoring.

Omar Al-Ubaydli
Department of Economics and Mercatus Center
George Mason University
omar@omar.ec

Steffen Andersen
University of Copenhagen
sa.eco@cbs.dk

Uri Gneezy
Rady School of Management
University of California - San Diego
Otterson Hall, Room 4S136
9500 Gilman Drive #0553
La Jolla, CA 92093-0553
ugneezy@ucsd.edu

John A. List
Department of Economics
University of Chicago
1126 East 59th
Chicago, IL 60637
and NBER
jlist@uchicago.edu

Carrots that look like sticks:
Toward an understanding of multitasking incentive schemes

Omar Al-Ubaydli, Steffen Andersen, Uri Gneezy and John A. List[1]

June, 2012

Abstract

Constructing compensation schemes for effort in multi-dimensional tasks is complex, particularly when some dimensions are not easily observable. When incentive schemes contractually reward workers for easily observed measures, such as quantity produced, the standard model predicts that unrewarded dimensions, such as quality, will be neglected. Yet, there remains mixed empirical evidence in favor of this standard principal-agent model prediction. This paper reconciles the literature by using both theory and empirical evidence. The theory outlines conditions under which principals can use a piece rate scheme to induce *higher* quantity and quality levels than analogous fixed wage schemes. Making use of a series of complementary laboratory and field experiments we show that this effect occurs because the agent is uncertain about the principal's monitoring ability and the principal's choice of a piece rate signals to the agent that she is efficient at monitoring.

JEL codes: D63, D82, J3

Keywords: gift exchange; piece rate; incentives

> "A Father, being on the point of death, wished to be sure that his sons would give the same attention to his farm as he himself had given it. He called them to his bedside and said, "My sons, there is a great treasure hid in one of my vineyards." The sons, after his death, took their spades and mattocks and carefully dug over every portion of their land. They found no treasure, but the vines repaid their labor by an extraordinary and superabundant crop." Aesop's Fables.

1. Introduction

Designing optimal incentive schemes is perhaps one of mankind's oldest activities. From the Dead Sea Scrolls to scribes on tombs of ancient kings, rudimentary and clever incentive structures to motivate a particular course of action have been extolled. For their part, economists have produced a rich assortment

of models that lend insights into the various factors that are likely to influence equilibrium market behavior. A particularly important and relatively complex problem arises when output has multiple dimensions that vary in their quantifiability. For example, school teachers are responsible for improving a wide range of students' academic and social skills, most of which are difficult to objectively measure. Factory workers often have outputs that are readily observed – quantities produced – whereas inherent product quality might be difficult or impossible to measure.

The seminal theories addressing optimal incentive schemes in multitasking were put forward by Holmstrom and Milgrom (1991) and Baker (1992). They start by explaining why simply rewarding workers on the easily measurable components may fail: workers may systematically neglect the unrewarded components of output to the detriment of the employer. This intuition is subsequently used to explain the prevalence of flat-wage structured incentives in occupations where piece rates could be easily implemented.

Although such theories are intuitively appealing, such behavior is not ubiquitously observed. For example, Lazear (2000), Shearer (2004), and Copeland and Monnet (2009) find that measured quality under a piece rate is comparable to quality under flat wage compensation. At the same time, other studies report quantity/quality tradeoffs consonant with standard neoclassical theory (Paarsch and Shearer 2000, Johnson et al. 2006; also see the discussion in Baland et al. 1999; for a comparison of piece rates and tournament schemes, see Bandiera et al. 2011).

This study reconciles the literature by first providing a theoretical exploration of the possibility of hidden benefits of using a piece rate incentive scheme. The key underlying feature of the model, which relates to the general framework of Benabou and Tirole (2003), is that both the agent *and* the principal have hidden information. Under that assumption, a piece rate compensation scheme can have a more complex effect on the quantity/quality tradeoff than currently understood.

To lend insights into our theoretical construct, we make use of an environment in which the two-sided asymmetric information has the greatest likelihood of being observed – hiring workers via a temporary employment agency and making the job last for two days. The task to be completed, preparing materials for distribution by a charitable fundraiser, permits us to measure output along two crucial dimensions (quantity and quality) with workers uncertain about our ability to measure quality. Therefore, we are able to provide a two-dimensional test of our incentive schemes, capturing the essential elements of the relevant multitasking theory, in an environment wherein there is two-sided asymmetric information: both the agent and principal have important hidden information.

We report several insights. Most importantly, consistent with the model, we find that workers under the piece rate compensation scheme deliver comparable productivity and *superior* quality to those in flat wage compensation schemes. Yet, while the empirical results are consistent with our theory, a number of caveats are necessary. One way to solve this quandary is to make use of complementary experiments. To address the underlying motivations for the main empirical insight, we begin with a laboratory experiment. By studying artificial markets that differ only in the manager's monitoring ability, we are permitted a unique glimpse into whether two-sided asymmetric information by itself can lead to such predicted consequences. In a complementary field experiment, we explore what role gift exchange might be playing in our main experiment.

Both the laboratory and the field experiments provide evidence that the data patterns observed in the main experiment are a result of two-sided asymmetric information in our naturally-occurring labor market. In this manner, two-sided asymmetric information is potentially an important feature to consider when predicting the effects of implementing various pecuniary incentive schemes. In addition, our results taken together suggest that non-pecuniary incentive schemes can work in predicted directions.

Our study also has certain methodological innovations. First, our paper showcases the benefits of combining laboratory and field data to test economic theory, provides practical advice to managers, and explores the underlying motivations of how incentives operate. Second, by overlaying randomization onto a natural labor environment we are able to explore clean tests of standard theory and generate behavioral parameters to help construct an alternative theory. Third, we use a novel subject pool: temporary workers, who represent an interesting middling ground between subjects in one-shot experiments and full-time employees in firms. Such workers represent an important component of the labor market, and to our knowledge have largely been ignored.[2]

The remainder of our paper is organized as follows. Section 2 is a simple model of piece rates. Section 3 is a piece rate field experiment. Section 4 is a piece rate laboratory experiment. Section 5 is a gift exchange field experiment. Section 6 offers concluding remarks.

2. A model of piece rates with hidden benefits

Despite humankind's constant tinkering with incentive schemes, it is not difficult to find oft-used approaches yielding seemingly perverse results. Examples within organizations are well documented (see Kerr 1975 and Baker et al. 1994), and include examples such as employees aiming to hit quantity targets regardless of quality or failing to co-operate with other employees. Unwanted incentive effects generally arise when a job requires workers to perform several tasks but only some are measured and rewarded. Unsurprisingly, in such circumstances, the worker will often concentrate her efforts on the rewarded tasks to the detriment of the overall organizational objectives.

This insight is formally shown in the seminal work of Holmstrom and Milgrom (1991) and Baker (1992). Holmstrom and Milgrom (1991; henceforth HM) introduce a multitasking model which explains why incentive pay is often not appropriate even when accurate performance measures are available. HM further describe how the subset of tasks performed within an individual job and the method of pay for that job, are jointly determined. It is predicted that subsets of tasks will be grouped around the costs of measuring and rewarding performance. Some workers will perform a set of easy to measure tasks and will be paid based on measured performance. Others will perform a set of difficult to measure tasks and will receive a fixed wage.

In this paper we take an alternative approach, outlining the conditions under which there are actually hidden benefits in using a piece rate scheme. Our starting point is the Benabou and Tirole (2003; henceforth BT) model in which agents hold hidden information, and update their beliefs regarding the

---

[2] Professional temporary workers are widespread. In 2005 alone, US staffing companies employed an average of 2.9 million temporary and contract workers. Further, on any given day the staffing industry employs more than 2% of the US work force (American Staffing, 2006).

interaction based on the incentives offered by the principal. The principal's choice of contract signals something about her private information. This kind of decision calculus can result in the realization of hidden benefits, in that the standard neoclassical model would not capture them. BT deploy these models of two-sided asymmetric information to understand how incentives can signal information. The following model is stylized for expositional purposes. Specific functional forms are used to illustrate the possibility of certain results.

<center>A. Setup</center>

Consider a principal-agent problem with two-dimensional effort, $e = (e_1, e_2)$, where e_1 denotes quantity and e_2 denotes quality. Initially, quantity (e_1) is unobservable. To implement any contract that is based on e_1, the principal must incur a fixed cost R; it represents the resources needed to, e.g., count and track an individual worker's output. Paying a flat wage allows the principal to avoid this cost.

Quality (e_2) lies in the range $[0, \bar{e}]$; any quality below $\bar{e}$ is described as substandard.[3] It is not feasible to monitor quality (e_2) perfectly, or, more specifically, in a manner in which permits fully-enforceable quality-contingent contracts. This could be because quality is so difficult to define objectively. However the principal can detect substandard quality with probability $t(e_2)$.

$$t(e_2) = \theta(\bar{e} - e_2), \qquad \theta \in [0, 1/\bar{e}]$$

The lower the quality, the more likely it is to be detected. If the principal detects substandard quality $(e_2 < \bar{e})$, then she imposes a fine f (this can represent being fired).[4] The fine is lost and does not enter the principal's objective function directly.

The cost of exerting effort to the agent is:

$$C(e) = \frac{1}{3}[(e_1 - 1)^2 + (e_2 - 1)^2 + (e_1 - 1)(e_2 - 1)]$$

Note that the cost of effort function $C(e)$ is minimized at $e_1 = e_2 = 1 < \bar{e}$. This reflects the intrinsic satisfaction from producing something (HM).[5] The function is convex, implying that the marginal cost of e_i is increasing in the both e_1 and e_2.

The output resulting from the agent's effort is:

$$Y(e) = 2(e_1 + e_2)$$

This accrues to the principal. Finally, the compensation scheme is linear:

$$w = \alpha + \beta e_1$$

---

[3] Our focus is the effect of financial incentives on error rates. For those interested in a more cognitive and neuropsychological analysis of the causes and treatments of errors, see for example Wickens (1980), Reason (2000) and Kane and Engle (2003).

[4] The fine is fixed in size; principals would ideally like to make it larger but legal or union-based restrictions prevent this.

[5] Alternatively, we can imagine a model where managers can detect and punish low levels of effort such that the net marginal disutility of effort is negative at low levels of effort, but that after a certain point, detection probability is sufficiently low as to be dominated by the disutility of effort.

Minimum wage laws stipulate that $\alpha \geq \bar{\alpha}$. For the remainder of this note, we fix $\alpha = \bar{\alpha}$ since it will follow trivially from the solution to all the principal's problems.[6]

Let D be a dummy variable that takes the value 1 when $\beta > 0$ and 0 when $\beta = 0$. The principal's payoff is output minus compensation and any costs of implementing a piece rate:

$$u_P = Y(e) - w - DR$$

Let $f\theta = F$ and assume that $F < 1$. The agent's payoff is compensation minus the cost of effort and the expected cost of being caught delivering substandard quality:

$$u_A = w - C(e) - F(\bar{e} - e_2)$$

Two features of the environment are the private information of the principal: how much she has to pay to be able to implement a piece rate (R) and the parameter governing how likely she is to detect substandard quality (θ), and hence F. The order of play is as follows.

1. Principal privately draws (R, F)
2. Principal selects β
3. Agent observes β and selects (e_1, e_2)

All features of the game are common knowledge.

B. One-sided asymmetric information

As is conventional when analyzing piece rates (HM), we eliminate the asymmetric information about the principal's parameters, implying that (R, F) is exogenously fixed. Thus asymmetric information is limited to the agent's choice of quality (e_2). The agent's problem is therefore:

$$\max_{e_1, e_2} \left\{ \bar{\alpha} + \beta e_1 - F(\bar{e} - e_2) - \frac{1}{3}[(e_1 - 1)^2 + (e_2 - 1)^2 + (e_1 - 1)(e_2 - 1)] \right\}$$

Proposition 1 (Holmstrom and Milgrom 1991): Under one-sided asymmetric information, increasing the piece rate factor β leads to higher quantity (e_1) and lower quality (e_2).

Proof: See appendix. ∎

This is the classic multitasking result: piece rates generate a quantity/quality tradeoff. The result is driven by the complementarity in the cost of both efforts. In equilibrium, the principal's choice of compensation (β) will depend upon the model's main parameters (R, F).

Despite the theoretical and intuitive appeal of this main HM prediction, finding evidence of the mechanism, especially in field settings, can be tricky. For the econometrician to have a high powered test, she needs access to accurate quality data. If such data exist, then the manager's access to such data is likely to be at least as good as the econometrician's, in which case the manager can implement quality-contingent contracts, thereby circumventing the quantity/quality tradeoff. This might be why studies such

---

[6] The agent's reservation utility is assumed to be sufficiently low so as to never bind.

as Lazear (2000), Shearer (2004), and Copeland and Monnet (2009) find that quality under a piece rate is comparable to quality under flat compensation.

Yet, there is evidence at odds with these findings. Several field studies have found quantity/quality tradeoffs (Paarsch and Shearer 2000, Johnson et al. 2006; also see the discussion in Baland et al. 1999) that exist when moving from a fixed wage scheme to a piece rate contract. Particularly in the case of Johnson et al. (2006), a fruitful strategy is for the econometrician to collect quality data with substantially more vigor and expense than the manager would realistically deploy. Without further theory and empirical evidence, it is difficult to reconcile the disparate observations in the literature.

C. Two-sided asymmetric information

Following BT, now assume that there are two types of principals. High types (henceforth highs) have $(R, F) = (2,0)$ and low types (henceforth lows) have $(R, F) = (0, 1/2)$. Lows are good at monitoring (or find it cheap to do so) while highs are bad at monitoring (or find it expensive to do so). We assume complementarity between monitoring quantity and quality, which is intuitive.

Proposition 2 (Benabou and Tirole 2003): Under two-sided asymmetric information, there exists a separating sequential equilibrium where:

- Lows offer a piece rate ($\beta > 0$), highs offer flat compensation ($\beta = 0$)
- Both types of effort are higher under a piece rate than under flat compensation

Proof: See appendix. ∎

The main result of our theory is that effort of both types is higher under the piece rate. One might then ask, why do high types actually use a flat compensation scheme? This happens because their monitoring costs on quantity (R) are so high that the cost of implementing the higher effort outweighs the benefits of both efforts being higher.

In contrast to one-sided asymmetric information, two-sided asymmetric information (under the above parameter values) implies the following. Given the principal's values of (R, F), she <u>does not</u> face a quantity/quality tradeoff to increasing β; both improve when switching from flat compensation to a piece rate.

The result is driven by the complementarity in monitoring quantity and quality. The compensation scheme selected signals monitoring costs. If the principal selects the piece rate, the agent thinks to herself: "The principal would only do that if she is good at monitoring quantity (e_1), which also means that she's good at monitoring quality (e_2), so I'd better shape up on both fronts."

Clearly, different functional forms or a different relationship between the ability to monitor quantity and quality could lead to different results. Moreover the above equilibrium is not necessarily unique. As is discussed extensively in BT, two-sided asymmetric information leads to an "anything goes" environment. The above example does highlight, however, that reasonable assumptions can generate hidden benefits of moving from a fixed wage scheme to a piece rate. The question naturally becomes an empirical one.

3. Experiment I: Piece rate field experiment

Finding a field environment conducive to testing our theory is difficult because one needs to simultaneously find a task that includes the dimensions necessary to monitor quality and quantity as well as the necessary two-sided asymmetries in information. The prospect to design and test the efficacy of various incentive schemes in an actual work environment presented itself when we agreed to help a non-profit organization with a capital campaign in which several thousand potential donors were to receive direct mail solicitations. Our duties included preparing and distributing the solicitations.

According to our model, in an environment where quality is imperfectly monitored, a necessary condition for piece rates to improve quality is for the agent to have imperfect information about the principal.[7] We chose to conduct the experiment using a short-term job and to hire workers via a temporary employment agency, both of which are natural features of solicitation preparation. Such an environment and hiring procedure serve to create a large degree of uncertainty from the agent's side about the principal's monitoring capabilities.

Beyond learning about incentive schemes in such a setting, we view this approach as providing insights into the underpinnings of a worker pool that has not been studied in a controlled experiment. According to the American Staffing Association's quarterly employment and sales survey, temporary workers are quite broad-based (American Staffing 2006): at least 10% of the US labor force has been employed as a temporary worker, and at any given time more than 2% of staff workers, or nearly 3 million workers, are temporaries.

A. Treatments

In the control group, workers were paid a flat hourly wage of $9/hr, which corresponded to the market wage for a task such as envelope packing. In the treatment, workers were paid minimum wage ($8/hr) plus $0.20 per envelope packed.[8] The nature of the task (see below) implied that envelopes were packed at a rate of approximately 13 envelopes per hour, resulting in an effective hourly wage of approximately $10.60/hr.[9]

Workers were hired to work for two days and they were explicitly told that this was a one-off job. The first day of work was six hours, which included 15 minutes of orientation and a 30 minute paid lunch break. The second day of work was a four hour period. Workers were informed that they would be invited back on the second day should their performance on the first day be deemed satisfactory.[10]

---

[7] This is also assuming complementarity in the disutility of effort.

[8] We conducted sessions in summer 2009 and summer 2010. By summer 2010, the minimum- and market wages had risen by $0.25, and so we increased all our wages accordingly. All of our regressions contain year effects to control for this potential level effect.

[9] The fact that the effective wage rate in the piece rate is higher than that in the flat hourly scheme raises the possibility of confounding with a gift effect. We explore this in the complementary experiment in section 5.

[10] No worker was terminated after the first day. A few of the workers left early or simply did not show up on the second day (some of whom had an excuse). For our empirical analysis, we include data for these workers. However dropping these workers from our dataset does not alter any of our results.

We ran four control sessions and four treatment sessions across four (pairs of) days. On each of those days, we ran a control session *and* a treatment session concurrently in separate, identical rooms. This ensures that day effects cannot confound our empirical analysis.

B. Recruitment

We made inquiries to numerous employment agencies to hire workers to pack envelopes for our non-profit organization. We personally interviewed several agencies and selected the firm that we regarded as the best balance between the price charged and the flexibility required to run the natural field experiment.[11] In the end, we contracted with a local temporary-employment agency that was broad based and had years of experience in this area of work – FurstStaffing Services.

We worked closely with the temporary work agency to recruit workers for our task in a natural manner. This was done by placing advertisements on various employment websites. The advertisements simply requested envelope packers for a non-profit organization at the University of Chicago, and requested that potential workers contact the employment agency if they had interest. Importantly, we followed standard procedure by not mentioning compensation or the exact date(s) of employment on the advertisements at this stage.

Recruitment generally followed four steps. First, when potential workers contacted the employment agency expressing interest in the job, the agency followed its usual protocol of inviting applicants for a brief interview, which included completion of paperwork that provided a demographic profile. Second, workers were informed that the employment agency would contact them soon with further details about the job, including compensation and the date(s) at which employment was available.

These first two steps provided us with a large pool of potential workers. In step three, potential workers were randomly allocated to one of the treatments. In the fourth step, staff of the employment agency personally called each potential worker and inquired about their availability on their assigned day (corresponding to their allocated treatment). If the prospective employee replied in the affirmative, they were informed of the compensation package, and asked to confirm working for the specified compensation on the chosen date. If the prospective worker declined after learning about the wage, they were eliminated from the sample.[12]

If the prospective employee was not available to work on a specific date, while still being unaware of the compensation scheme, they were offered another randomly selected date. Importantly, no individual was offered more than one wage or made aware of the existence of a wage that differed from the one that they accepted or rejected. The target recruitment for each session was 10 workers; this necessitated hiring and confirming 12 workers given an expected show-up rate of 80%.

---

[11] Without letting the agencies know about our plans, it was clear through these initial interviews that many of the agencies were not inclined to provide the necessary control to complete a viable experiment.

[12] While some potential workers rejected a session due to conflicting commitments, nobody ever declined upon hearing the compensation scheme conditional on expressing willingness to work at that time/date. There were no significant differences in show-up rates across treatments. One demographic variable (having some college education) was marginally statistically significant in predicting show-up rates; this was of no concern because it was never significant in any of the outcome regressions, and because it was balanced across treatments as indicated by our ANOVA. No other demographic variables predicted show-up rates. Informal discussions with the employment agency suggested that random personal emergencies were the modal reason for no-shows.

C. The task

Upon arriving on site, each worker was greeted, signed-in, and seated at an individual desk. The monitor's desk was placed in the center of the room. Upon arrival of all workers, the monitor gathered the workers to receive instructions. The monitor proceeded with a 15 minute orientation.[13]

Workers were informed that the task was part of a charitable fundraising drive. In this drive, several thousand letters needed to be prepared and mailed. The monitor informed workers that there was a stack of letters placed on their desk, as well as other support materials. The worker was to: (1) match the top letter from the stack with the matching address label, (2) read the letter to identify the correct accompanying materials, and (3) insert the necessary materials into the envelope. It was stressed that great importance lie in correctly producing and packing the envelopes. Workers were informed that they should then place completed envelopes to the side on their desk. We further informed workers that monitors would collect the completed envelopes during the day (in practice we collected them on the top of each hour).

In the final preparatory step, workers completed important paperwork that accompanied each letter. This time-consuming task had the worker verify each letter that was sent to ensure that we did not send multiple letters to each household. Importantly, this task was simple to complete accurately, only requiring that the worker could read and write.[14] We did not collect individual paperwork until the close of the work day, sending a signal to the worker that this was a component of effort that would not be monitored as closely as the count of letters. The monitor closed the introduction by asking for questions. After all questions were answered, work commenced.

In sum, workers were recruited to a natural work environment to complete a task for compensation. The task and wage profiles were similar to many other jobs they had accepted previously from employment agencies. The task involved a range of activities, each of which might include defects, or errors.[15] This permits us to observe both the quantity and quality of the workers' output. In total, there are 12 potential errors, but the effect of each on output (the final output of an envelope is a successful contribution by the recipient of the specific letter) is uncertain. We classified the errors into three broad categories reflecting the error's importance based on introspection.[16]

- **Critical errors**: all but guaranteed failure of the letter to generate contributions. These included incorrectly matching the address label and letter (e.g., the household of Mr. J. Smith receives a letter addressed to the household of Mr. R. Jones) or forgetting to include either the address label or the letter.
- **Non-critical errors**: represent a portmanteau for remaining errors in matching the materials, e.g., failing to include a complimentary bookmark when the letter states that the recipient should get one, or omitting a copy of the newsletter.

---

[13] There was no evidence of any seating effects on productivity (e.g., neighbor effects or position effects). See Falk and Ichino (2006) for an example.

[14] Literacy was a requirement at the recruitment process that was covered by the employment agency.

[15] To prevent unconscious bias in data gathering, the research assistants who checked the envelopes for errors were blind to treatment.

[16] The solicitation failed to generate a response rate high enough to allow an econometric estimation of the effects of each error type, but the effect sizes were ordered consistently with the 3 broad categories.

- **Recording errors**: reflect the administrative task that was completed directly after the envelope-packing task. This was a time-consuming activity that required the workers to look up addresses in a long list. Errors of this class involve failure to complete it correctly or failure to attempt to complete the task. We viewed this error type as one that is not necessarily cognitive, rather it represented a simple task that would be time-consuming, and therefore yield fewer letters per hour.

At no point were workers made aware of that fact that they were taking part in an experiment, thus we are executing a natural field experiment in the parlance of Harrison and List (2004).

D. Prediction

Unexpected effects of a piece rate would be an increase in quantity *and* quality from the fixed wage baseline. A necessary condition for this to occur is two-sided asymmetric information.

Prediction: As a consequence of the two-sided asymmetric information, in the piece rate session (treatment), quantity and quality will be higher than in the flat hourly session (control).

E. Results

Table 1 contains details about when the sessions were executed. The table includes information about sessions for Experiment III (denoted Gift in the Compensation column), which will be discussed later. For the empirical analysis in Experiment I, due to the possibility of day effects, we omit data from group 1.

Group	Compensation	Start date	Workers
1	Flat	17 Aug 2009	12
1	Gift	17 Aug 2009	11
2	Flat	27 Aug 2009	10
2	Gift	27 Aug 2009	6
2	Piece rate	27 Aug 2009	9
3	Flat	1 Sep 2009	6
3	Gift	1 Sep 2009	10
3	Piece rate	1 Sep 2009	8
4	Flat	9 Aug 2010	10
4	Piece rate	9 Aug 2010	9
5	Flat	12 Aug 2010	9
5	Piece rate	12 Aug 2010	5

Table 1: Session dates

Table 2 contains the sample statistics for the main dependent variables and the demographic controls. A one-way multivariate ANOVA fails to reject equality of the demographics across the two treatments (p = 0.42), suggesting that our randomization was effective in balancing the important variables across the treatment cells.[17]

	Control: Flat wage	Treatment: Piece rate
Envelopes per hour	14 (4.3)	13 (2.7)
Critical errors per envelope	0.034 (0.17)	0.0032 (0.0096)
Non-critical errors per envelope	0.47 (0.34)	0.35 (0.33)
Recording errors per envelope	0.49 (0.43)	0.20 (0.25)
Female (dummy)	0.63 (0.49)	0.55 (0.51)
Age (years)	36 (11)	35 (11)
Black (dummy)	0.83 (0.38)	0.58 (0.50)
Some college education (dummy)	0.43 (0.50)	0.39 (0.50)
Bachelor's degree or more (dummy)	0.29 (0.46)	0.32 (0.48)
Number of workers	35	31

Table 2: Sample means and standard deviations in parentheses for Experiment I
All figures are to two significant figures.

Result 1: Output per hour under the piece rate is slightly lower than output per hour under the flat hourly compensation scheme (a difference of 0.84 envelopes per hour = 23% of a standard deviation), but the difference is statistically insignificant at conventional levels.

A Mann-Whitney test on output in control (flat) vs. treatment (piece rate) yields a p-value of 0.70. In Model 1 in Table 3, a regression with demographic controls and day effects confirms the result.[18]

---

[17] If we conduct a series of 2-sample t-tests rather than one multivariate ANOVA, then we find that there are significantly less black people in the treatment. However since in our sample, black people commit less errors than non-blacks, and our main results revolve around the treatment leading to less errors, this minor imbalance only serves to reinforce our results. Regardless our results are never affected by the inclusion of demographic controls.

[18] Running a regression where we have hourly data with time effects, demographic controls, and random effects to control for individual effects yields analogous results. This is true for all the models in Table 3 (and Table 6 further below).

Model	1	2	3	4
Dependent variable	Envelopes / hr	Critical errors / env	Non-critical errors / env	Recording errors / env
Piece rate (dummy)	-0.80	-0.055*	-0.066	-0.33***
	(0.95)	(0.031)	(0.083)	(0.089)
Female (dummy)	1.9*	0.040	-0.11	0.021
	(0.98)	(0.032)	(0.086)	(0.092)
Age (years)	-0.12	0.013	-0.022	-0.030
	(0.40)	(0.013)	(0.035)	(0.037)
Age (years) squared	0.0013	-0.00019	0.00033	0.00046
	(0.0052)	(0.00017)	(0.00046)	(0.00049)
Black (dummy)	-0.80	-0.10***	0.25**	-0.073
	(1.1)	(0.037)	(0.10)	(0.11)
Some college (dummy)	0.66	0.045	0.019	0.0023
	(1.2)	(0.038)	(0.10)	(0.11)
Bachelor's degree (dummy)	-0.31	-0.020	-0.13	0.041
	(1.24)	(0.040)	(0.11)	(0.12)
Observations	66	66	66	66
R-squared	0.15	0.20	0.27	0.34

Table 3: Regression results for Experiment I

Figures in parentheses are standard errors. All regressions control for day effects. Asterices denote statistical significance: * = 10% level, ** = 5% level, *** = 1% level. All figures are to two significant figures.

Result 2a: Critical errors per envelope under the piece rate are slightly less frequent than critical errors per envelope under the flat hourly compensation scheme (a difference of 0.031 critical errors per envelope = 25% of a standard deviation), and the difference is statistically significant.

A Mann-Whitney test of critical errors per envelope yields a p-value of 0.022. In Model 2 in Table 3, controlling for demographics and day effects leads to a larger treatment effect (0.055 critical errors per envelope = 45% of a standard deviation), though the p-value drops to marginal significance (0.083).

Result 2b: Non-critical errors per envelope under the piece rate are less frequent than non-critical errors per envelope under the flat hourly compensation scheme (a difference of 0.12 non-critical errors per envelope = 36% of a standard deviation), but the difference is statistically insignificant.

A Mann-Whitney test of non-critical errors per envelope yields a p-value of 0.14. In Model 3 in Table 3, controlling for demographics and day effects leads to a weakening of the treatment effect (0.066 non-critical errors per envelope = 19% of a standard deviation) and a diminution of the p-value to 0.43.

Result 2c: Recording errors per envelope under the piece rate are much less frequent than recording errors per envelope under the flat hourly compensation scheme (a difference of 0.29 recording errors per envelope = 75% of a standard deviation), and the difference is statistically significant.

A Mann-Whitney test of critical errors per envelope yields a p-value of 0.0065. In Model 4 in Table 3, controlling for demographics and day effects leads to a larger treatment effect (0.33 recording errors per envelope = 87% of a standard deviation).

<div align="center">F. Discussion</div>

One of the virtues of Experiment I is that we have precise quality data in an environment where the workers cannot be *certain* of such accuracy.[19] Under these conditions, one would initially expect the HM quantity/quality tradeoff (see the discussion at the end of section 2.B).

Though output was approximately equal under piece rate and flat compensation, there is substantive evidence that quality was superior under the piece rate. This constitutes an unexpected result in light of the HM model. The results correspond to an environment better described by our model. Critical evaluation of the results, however, reveals that sharp tests of the exact underlying mechanisms at work are not possible because of two distinct drawbacks.

First, to provide more definitive support for our model, we would ideally repeat our field experiment with only one difference across the treatment cells: workers would be certain of the manager's monitoring ability in one case and uncertain in the other. In this case, observing an HM quantity/quality tradeoff would cement our conclusion that the results of Experiment I reflected the mechanism in our model. Yet, this is not feasible in our natural field experimental environment without introducing additional confounds. The only way to eliminate (credibly) workers' uncertainty about our monitoring capability is to convince them that it is very high, which precludes a treatment effect since workers will be forced to deliver high quality in all treatments (this happened in Lazear 2000, where workers were explicitly aware of fines for poor workmanship).

Accordingly, for workers to be *certain* that our ability to monitor quality was moderate or low, they would have to work with us for a long time so that they could acquire firsthand evidence (this is what occurred in Johnson et al. 2006). It is not possible with temporary workers hired for a two-day job. We therefore use a laboratory experiment to examine if the two-sided asymmetric information by itself can have the predicted theoretical effect. We discuss this experiment in the next section.

Second, an alternative, and arguably simpler, explanation for our results is gift exchange (Akerlof 1982). Recall that at average productivity levels, piece rate workers earned $1.60 more per hour than flat compensation workers, the latter being paid the market wage of $9 per hour.[20] Perhaps the superior quality is an expression of reciprocity for our 'gift' of above-market wages. To explore this possibility, we conducted a parallel gift exchange natural field experiment, which we discuss in Section 5.

---

[19] This is one of the reasons it is important that the workers were unaware of their participation in an experiment. We hired Staff at substantial expense to check every envelope unaware of the experimental treatments.

[20] We would have ideally eliminated this gap, but a minimum wage laws prevented us from forcing the fixed component any lower, and we wanted to avoid having a trivial per-envelope fee of $0.07.

4. Experiment II: Lab experiment to differentiate theories

The goal of our laboratory experiment is to test cleanly the mechanisms behind the HM model and our model. While our lab experiment is admittedly artificial, in that it explicitly suppresses many of the rich characteristics of well-functioning labor markets, it affords us an opportunity to precisely manipulate the information and expectations of managers and workers. In this manner, we are able to test the direct effect of two-sided asymmetric information while maintaining a strong theoretical link.

A. Environment

The environment was a simplified version of the model in Section 2. Participants were undergraduate students from the University of California at San Diego. After arrival and check-in, participants were randomly divided into managers and workers. After 2 practice rounds, they interacted for 40 rounds with random rematching after each round (random stranger). The Appendix contains the experimental instructions.

Every round started with a manager being paired with a worker. Each manager learned her type (low vs. high), she selected the compensation scheme (flat vs. piece rate), the worker saw the manager's choice (but not the manager's type) and then selected both dimensions of effort (e_1 and e_2).[21]

Managers were high types with probability 0.75 and low types with probability 0.25. High managers had to pay 40 to implement a piece rate while low managers did not ($R = 40$). Both efforts were constrained to the values $\{1,2\}$; $\bar{e} = 2$. High managers could not enforce any quality-contingent contract ($\theta = 0 \Rightarrow F = 0$). Low managers would always detect low quality ($\theta = 1$), and this would result in an automatic 30 point fine for the worker ($f = 30 \Rightarrow F = 30$).

The cost of effort is convex: $C(1,1) = 0, C(1,2) = C(2,1) = 10, C(2,2) = 30$, and output is given by $Y(e) = 25(e_1 + e_2)$. Flat compensation corresponded to $\alpha = 30$ and $\beta = 0$, and piece rate compensation corresponded to $\alpha = 20$ and $\beta = 30$. Figure 1 provides the payoff table for both managers and workers.

	$e_1 = 1, e_2 = 1$	$e_1 = 2, e_2 = 1$	$e_1 = 1, e_2 = 2$	$e_1 = 2, e_2 = 2$
Low manager: Flat	20, 0	45, -10	45, 20	70, 0
Low manager: Piece rate	30, -10	25, 10	55, 10	**50, 20**
High manager: Flat	**20, 30**	45, 20	45, 20	70, 0
High manager: Piece rate	-10, 20	-15, 40	15, 10	10, 20

Figure 1: Manager (first figure) and worker (second figure) payoffs

The game's sequential equilibrium (recall that the manager's type is her private information) is analogous to Proposition 2 above: low managers always select a piece rate, high managers always select flat

---

[21] See Fehr and Schmidt (2004) for a laboratory comparison of piece rates and bonus contracts.

compensation; workers respond to a piece rate with $e_1 = e_2 = 2$ and to flat compensation with $e_1 = e_2 = 1$. The equilibrium is bolded in Figure 1.

B. Treatments and predictions

In the control (*exogenous*), it was common knowledge that a computer randomly selected the compensation scheme on the manager's behalf, thereby destroying any signaling value of the compensation scheme. The exact probability of a piece rate was never reported to the participants; we set it to 0.25 to mimic the unconditional moments of the sequential equilibrium.

In the control version, the risk-neutral equilibrium is for workers to respond to a piece rate with $e_1 = 2, e_2 = 1$ and to flat compensation with $e_1 = e_2 = 1$. Under mild risk-aversion, workers will respond to flat compensation with $e_1 = 1, e_2 = 2$ while still responding to a piece rate with $e_1 = 2, e_2 = 1$. Under extreme risk-aversion, workers will respond to a piece rate with $e_1 = 2, e_2 = 2$ and to flat compensation with $e_1 = 1, e_2 = 2$. The convexity of the cost of effort guarantees that quality (e_2) is never higher under the piece rate and that it is sometimes lower, as in HM.

Prediction: In the exogenous treatment, workers have higher quantity (e_1) and lower quality (e_2) in a piece rate than under flat compensation.

In the treatment (*endogenous*), it was common knowledge that the manager selected the compensation scheme after privately learning her type. This corresponds to the sequential equilibrium in our theory.

Prediction: In the endogenous treatment, workers have both higher quantity (e_1) and quality (e_2) in a piece rate than under flat compensation.

For each of the two treatments, we ran two versions. In the fixed version, participants were assigned the role of either manager or worker at the start of the experiment and they remained in that role for the entire experiment (40 rounds). In the rotating version, participants had an equal chance of being a manager or a worker at the start of each round. We ran the rotating version because we anticipated faster convergence to equilibrium behavior; as it happens, both versions exhibited rapid convergence to the equilibrium.

C. Results

We ran 18 sessions in the spring of 2011: 9 exogenous (control) and 9 endogenous (treatment). For each treatment, 6 of the 9 were fixed and 3 of the 9 were rotating. Each session had 10 participants and 40 rounds, yielding a total of 400 manager-worker observations per session. The summary statistics are in Table 4.

Treatment	Exogenous: Fixed	Exogenous: Rotating	Endogenous: Fixed	Endogenous: Rotating
% of low managers choosing piece rate	24%	25%	72%	88%
% of high managers choosing piece rate	27%	27%	1%	1%
e_1 under flat compensation	1.0 (0.10)	1.0 (0.18)	1.0 (0.15)	1.0 (0.14)
e_2 under flat compensation	1.5 (0.50)	1.7 (0.47)	1.3 (0.47)	1.2 (0.42)
e_1 under piece rate	2.0 (0.18)	2.0 (0.20)	1.9 (0.23)	1.9 (0.29)
e_2 under piece rate	1.2 (0.36)	1.3 (0.44)	1.8 (0.38)	1.8 (0.36)

Table 4: Sample means and standard deviations in parentheses for Experiment II
All figures are to two significant figures.

In the fixed versions, in any given round, participants assigned the role of workers found themselves at one of two nodes: flat compensation vs. piece rate; participants assigned the role of managers also found themselves at one of two nodes: low type vs. high type. In the rotating versions, in any given round, participants found themselves at one of the four above nodes since they were randomly assigned the role at the beginning of each round.

In our empirical analysis, for each participant we generate one observation per node by averaging across rounds her decisions at *that* node. For example, if in the fixed exogenous treatment, participant 3 played $e_1 = 1, e_2 = 1$ on 10 of the 20 times she faced flat compensation and she played $e_1 = 1, e_2 = 2$ on 10 of the 20 times she faced flat compensation, then her observation for the flat compensation node is $\bar{e}_1 = 1, \bar{e}_2 = 1.5$. This is a conservative and simple way of dealing with within-participant dependence in observations.

Let $\bar{D}_i^L$ denote manager i's relative frequency of choosing a piece rate when low and $\bar{D}_i^H$ when high. Let $\left(\bar{e}_{1i}^p, \bar{e}_{2i}^p\right)$ denote worker i's average effort vector in response to a piece rate and $\left(\bar{e}_{1i}^f, \bar{e}_{2i}^f\right)$ in response to flat compensation.

Result 3: When the choice of compensation scheme has no signaling value about the manager's monitoring ability (exogenous treatment), workers deliver higher quantity (e_1) and lower quality (e_2) in a piece rate than under flat compensation.

In the fixed version, the average of $\left(\bar{e}_{1i}^p, \bar{e}_{2i}^p\right)$ for the 30 workers is (2.0, 1.2) and the average of $\left(\bar{e}_{1i}^f, \bar{e}_{2i}^f\right)$ for the 30 workers is (1.0, 1.4): quantity is higher and quality is lower under the piece rate.[22] Using a paired-values Wilcoxon test, both differences are significant at the 0.1% level (for the remainder of this section, all tests are paired-values Wilcoxon tests). In the rotating version, we have $\left(\bar{\bar{e}}_1^p, \bar{\bar{e}}_2^p\right) = (2.0, 1.3)$ and $\left(\bar{\bar{e}}_1^f, \bar{\bar{e}}_2^f\right) = (1.0, 1.7)$. The differences are also significant at the 0.1% level. These data are consistent with the HM model.

Result 4: When managers can choose the compensation scheme (endogenous treatment), high managers virtually always select flat compensation while low managers very frequently select piece rate compensation.

In the fixed version, the average of $\bar{D}_i^L$ for the 30 managers is 0.74 and the average of $\bar{D}_i^H$ for the 30 managers is 0.015. This difference is significant at the 0.1% level. In the rotating version, we have $\left(\bar{\bar{D}}^L, \bar{\bar{D}}^H\right) = (0.90, 0.014)$. The difference is also significant at the 0.1% level.

Result 5: When managers can choose the compensation scheme (endogenous treatment), workers deliver higher quantity (e_1) and quality (e_2) in a piece rate than under flat compensation.

For the 30 workers in the fixed version, $\left(\bar{\bar{e}}_1^p, \bar{\bar{e}}_2^p\right) = (1.9, 1.8)$ and $\left(\bar{\bar{e}}_1^f, \bar{\bar{e}}_2^f\right) = (1.0, 1.3)$. These difference are significant at the 0.1% level. For the 30 workers in the rotating version, $\left(\bar{\bar{e}}_1^p, \bar{\bar{e}}_2^p\right) = (1.9, 1.8)$ and $\left(\bar{\bar{e}}_1^f, \bar{\bar{e}}_2^f\right) = (1.0, 1.3)$, and again these differences are significant at the 0.1% level. Results 4 and 5 are consonant with the major features of our model.

At the end of all sessions, while we calculated earnings and prepared the paperwork for payments, we asked participants to describe and explain their strategies in an open-response survey. The modal responses matched the mechanisms and inferences described in the HM and our model.

<div align="center">D. Discussion</div>

Our theory provides a clear mechanism for the relation between quality and quantity. The key ingredient is substantial worker uncertainty about the manager's monitoring ability. In principle, if this uncertainty is eliminated, then the HM quantity/quality tradeoff re-emerges. This prediction is supported by the results of Experiment II.

5. Experiment III: Gift exchange field experiment

Akerlof (1982) proposed gift exchange as an alternative solution to the multitasking problem. The idea is that even when a worker is paid a flat wage, the level of pay influences effort, even in situations void of reputational considerations. In contrast to the traditional assumptions in economic models, workers are not mute to reciprocal motivations in the gift exchange model. Once receiving a gift in the form of higher

---

[22] These sample means differ slightly from those reported in Table 4 since we are averaging over averages, where as the sample means in Table 4 come from one gross average.

than market clearing wages, workers reciprocate with higher effort levels (positive reciprocity); when receiving less than they believe they deserve, workers punish the firm via sabotage or other more subtle measures (negative reciprocity; see Andreoni et al. 2003 for a comparison of positive and negative reciprocity; also see Bandiera et al. 2005 for a discussion of the effect of social preference between workers). The reliance of such schemes on non-egotistical preferences suggests that they can be thought of as non-pecuniary, despite having a pecuniary component in the literal sense.[23]

In an effort to compare pecuniary and non-pecuniary schemes as solutions to the multitasking problem using a natural field experiment, while eliminating the possibility of the piece rate results being driven by a gift effect, we constructed a gift treatment with a very generous gift and without a piece rate component.

A. Treatments, recruitment and the task

The control is identical to the flat compensation scheme in Experiment I. Similar to Gneezy and List (2006), the treatment, which we denote the 'gift' scheme, involves hiring workers at a wage rate of "between $9/hr and $18/hr" (see the appendix for the precise wording; see Charness 2000 for a discussion). When workers arrive at the workplace, immediately following the orientation, they are informed that their wage will be $18/hr. This is an exceptionally large gift for workers performing a task as straightforward as packing envelopes. All other features are identical, including the duration (two days) and the task.

We ran three control sessions and three treatment sessions across three pairs of days. On each of those days, we ran a control session and a treatment session concurrently in separate, identical rooms to ensure that day of the week effects would not confound interpretation. Two of the three control sessions corresponded to two of the four control sessions used in Experiment I.

B. Prediction

Numerous field and laboratory studies have found evidence of gift exchange in labor markets (see Fehr et al. 2009 for a comprehensive survey). While the evidence on gift exchange in multitasking environments is more limited, the theory (Akerlof 1982) suggests a straightforward generalization of the results from unidimensional output.

Prediction: As a consequence of positive reciprocity, in the gift session (treatment), quantity and quality will be higher than in the flat hourly session (control).

C. Results

Table 2 contains the sample statistics for the main dependent variables and the demographic controls. A one-way multivariate ANOVA fails to reject equality of the demographics across the two treatments (p = 0.93), suggesting that our randomization effectively allocated workers into treatment.

---

[23] This conjecture is typically termed the "fair wage-effort" hypothesis. An alternative underlying mechanism that can create similar data patterns is denoted the "efficiency wage theory," which surmises that wages above market-clearing levels occur because these wage profiles induce workers to be motivated in an effort to avoid being fired, which economizes on firm-level monitoring (see, e.g., Katz, 1986). For studies of the interaction between social preferences and financial incentives, see Meier (2007), Goette and Stutzer (2010) and Ariely et al. (2009).

	Control: Flat wage	**Treatment: Gift**
Envelopes per hour	11 (3.8)	12 (6.2)
Critical errors per envelope	0.0051 (0.011)	0.0053 (0.013)
Non-critical errors per envelope	0.51 (0.32)	0.49 (0.29)
Recording errors per envelope	0.52 (0.46)	0.33 (0.38)
Female (dummy)	0.54 (0.51)	0.67 (0.48)
Age (years)	38 (13)	36 (13)
Black (dummy)	0.75 (0.44)	0.74 (0.45)
Some college education (dummy)	0.36 (0.49)	0.37 (0.49)
Bachelor's degree or more (dummy)	0.32 (0.48)	0.37 (0.49)
Number of workers	28	27

Table 5: Sample means and standard deviations in parentheses for Experiment III
All figures are to two significant figures.

Result 6: Output per hour under the gift scheme is slightly higher than output per hour under the flat hourly compensation scheme (a difference of 0.70 envelopes per hour = 14% of a standard deviation), but the difference is statistically insignificant.

A Mann-Whitney test on output in control (flat) vs. treatment (gift) yields a p-value of 0.74. In Model 1 in Table 6, a regression with demographic controls and day effects confirms the result.[24]

---

[24] For Result 3 and the remaining results, there is no evidence of a decaying treatment effect similar to that reported by Gneezy and List's (2006) data entry experiment.

Model	1	2	3	4
Dependent variable	Envelopes / hr	Critical errors / env	Non-critical errors / env	Recording errors / env
Gift (dummy)	0.40	-0.00017	-0.012	-0.16
	(1.2)	(0.0035)	(0.086)	(0.11)
Female (dummy)	2.6*	0.0032	-0.079	0.070
	(1.3)	(0.0038)	(0.094)	(0.12)
Age (years)	-0.27	-0.00092	-0.022	-0.0098
	(0.36)	(0.0011)	(0.026)	(0.033)
Age (years) squared	0.0015	0.000013	0.00032	0.000078
	(0.0044)	(0.000013)	(0.0033)	(0.00042)
Black (dummy)	-0.011	-0.0021	0.14	0.10
	(1.3)	(0.0040)	(0.099)	(0.13)
Some college (dummy)	-0.013	0.0060	-0.053	-0.15
	(1.5)	(0.0043)	(0.11)	(0.14)
Bachelor's degree (dummy)	0.84	0.00059	-0.099	-0.14
	(1.5)	(0.0044)	(0.11)	(0.14)
Observations	55	55	55	55
R-squared	0.46	0.15	0.17	0.19

Table 4: Regression results for Experiment III

Figures in parentheses are standard errors. All regressions control for day effects. Asterices denote statistical significance: * = 10% level, ** = 5% level, *** = 1% level. All figures are to two significant figures.

Result 7a: Critical errors per envelope under the gift scheme are negligibly more frequent than critical errors per envelope under the flat hourly compensation scheme (a difference of 0.00014 critical errors per envelope = 1% of a standard deviation), and the difference is statistically insignificant.

A Mann-Whitney test of critical errors per envelope yields a p-value of 0.57. In Model 2 in Table 6, controlling for demographics and day effects confirms this result.

Result 7b: Non-critical errors per envelope under the gift scheme are negligibly less frequent than non-critical errors per envelope under the flat hourly compensation scheme (a difference of 0.017 non-critical errors per envelope = 5% of a standard deviation) and the difference is statistically insignificant.

A Mann-Whitney test of non-critical errors per envelope yields a p-value of 0.96. In Model 3 in Table 6, controlling for demographics and day effects confirms this result.

Result 7c: Recording errors per envelope under the gift scheme are less frequent than recording errors per envelope under the flat hourly compensation scheme (a difference of 0.19 recording errors per envelope = 44% of a standard deviation), but the difference is marginally significant at best.

A Mann-Whitney test of critical errors per envelope yields a p-value of 0.063. In Model 4 in Table 6, we see that controlling for demographics and day effects leads to a slight diminution of the treatment effect

(0.16 recording errors per envelope = 37% of a standard deviation) and it loses its marginal significance (p = 0.16).

<div align="center">D. Discussion</div>

Given the strong evidence of gift exchange reported in the literature survey by Fehr et al. (2009), we were initially surprised by the small and statistically weak gift effects detected. This was true despite of the fact that gifts used in our study were large– paying unskilled workers $18/hr to pack envelopes.

However, an important difference between our environment and many of the gift exchange environments considered in the literature is the presence of explicit and implicit reputational concerns. First, we told workers that they would only be invited back on the second day if their performance on the first day was deemed satisfactory (which is not uncommon for positions secured via a temporary employment agency).

Second, the employment agency itself generates dynamic incentives for the workers: it is common knowledge that workers who perform well (or who avoid performing particularly badly) in a temporary job are more likely to be given future assignments by the employment agency as the agency cultivates its own reputation for delivering good workers. Consequently, it is likely that these reputational concerns swamp any gift effect.[25]

Regardless of their explanation, the results of Experiment III reinforce our initial interpretation of the results from Experiment I. We have shown that in response to an incredibly generous gift, workers (in this environment) barely change their behavior. It follows that the much smaller gift embodied in the piece rate (at average productivity, 18% of the gift in the gift scheme) is not driving the improvement in quality observed in the piece rate.[26] This experiment, therefore, provides further support that we have pinpointed the underlying mechanism at work in our theory to explain the hidden benefits of a piece rate scheme in our multitasking environment.

6. Summary and conclusion

In one-shot environments, both pecuniary and non-pecuniary incentive schemes can potentially solve multitasking problems. An inherent problem with pecuniary incentive schemes, however, is that workers might substantially reduce their effort on tasks that produce unobservable outputs as they seek the salient reward to observable effort. The problem with non-pecuniary schemes is that if social preferences are weak, then they are unlikely to be cost effective.

---

[25] If reputational concerns are swamping incentive effects, then in principle, incentive effects should be more visible on the second day of work since only one of the two reputational avenues is in effect on the second day. Follow-up regressions (omitted for parsimony) lend some support to this prediction; treatment effects are almost always directionally closer to the incentive-models' predictions in the second day than in the first; however the changes in the treatment effects are typically not statistically significant, suggesting that the majority of swamping is due to the reputational effect of the employment agency.

[26] Even if we were to ignore the insignificance of the effect of the gift on recording errors in Model 4 in Table 2 and ignore the fact that the gift was incredibly generous, the treatment effect remains 50% of the treatment effect detected in the piece rate treatment.

This study provides empirical insights into the multitasking problem using a unique sample of workers. We were able to acquire distinctive quality data by expending an unprofitable amount on monitoring quality. Employees cannot be certain that quality is being monitored this precisely, and so the classic multitasking theory applies.

Our results paint an interesting, and unexpected, picture: when moving from a fixed wage scheme to a piece rate scheme, it is possible to enhance both the observed (quantity) and unobserved (quality) outputs. This follows because of two-sided asymmetric information: both the agent and principal have important hidden information. When both the agent and the principal have hidden information, a piece rate compensation scheme can have hidden benefits; that is, the quantity/quality tradeoff is not expected under certain parameter values.

We also find that workers respond to non-pecuniary incentives, but only marginally. One overarching lesson learned from this exercise is that lab and field data can importantly complement one another to further our understanding of the economics of incentive schemes. Interestingly, as a whole, behavioral differences across the incentive schemes are much less than what one would have expected from parallel laboratory experimental results. We believe that this is chiefly a result of the intermediate reputational concerns in contrast to the one-shot frame of laboratory studies. Yet, this area is certainly ripe for future research.

References

Akerlof, G. (1982). "Labor Contracts as Partial Gift Exchange," Quarterly Journal of Economics, 97, p543-569.

American Staffing Association (2006). American Staffing 2006: Annual Economic Analysis. http://www.americanstaffing.net/statistics/economic.cfm.

Andreoni, J., W. Harbaugh and L. Vesterlund (2003). "The carrot or the stick: rewards, punishments and cooperation," American Economic Review, 93, p893-902.

Ariely, D., A. Bracha and S. Meier (2009). "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially," American Economic Review, 99, p544-555.

Baker, G. (1992). "Incentive contracts and performance measurement," Journal of Political Economy, 100, p598-614.

Baker, G., R. Gibbons and K. Murphy (1994). "Subjective performance measures in optimal incentive contracts," Quarterly Journal of Economics, 109, p1125-1156.

Baland, J., J. Dreze and L. Leruth (1999). "Daily wages and piece rates in agrarian economics," Journal of Development Economics, 59, p445-461.

Bandiera, I. Barankay and I. Rasul (2005). "Social preferences and the response to incentives: evidence from personnel data," Quarterly Journal of Economics, 120, p917-962.

Bandiera, I. Barankay and I. Rasul (2011). "Team incentives: evidence from a firm-level experiment," Journal of the European Economic Association, forthcoming.

Benabou, R. and J. Tirole (2003). "Intrinsic and extrinsic motivation," Review of Economic Studies, 70, p489-520.

Brandts, J. and G. Charness (2004). "Do labor market conditions affect gift exchange? Some experimental evidence," Economic Journal, 114, p684-708.

Charness, G. (2000). "Responsibility and effort in an experimental labor market," Journal of Economic Behavior and Organization, 42, p375-384.

Copeland, A. and C. Monnet (2009). "The welfare effects of incentive schemes," Review of Economic Studies, 76, p93-113.

Falk, A., and A. Ichino (2006). "Clean evidence on peer effects," Journal of Labor Economics, 24, p39-57.

Fehr, E. and K. Schmidt (2004). "Fairness and incentives in a multi-task principal-agent model," Scandinavian Journal of Economics, 106, p453-474.

Fehr, E., C. Zehnder and L. Goette (2009). "A behavioral account of the labor market: The role of fairness concerns," Annual Review of Economics, 1, p355-384.

Goette, L. and A. Stutzer (2010). "Blood donations and incentives: evidence from a field experiment," Unpublished paper.

Gneezy, U. and J. List (2006). "Putting Behavioral Economics to Work: Field Evidence of Gift Exchange," Econometrica, 74, p1365-1384.

Harrison, G. and J. List (2004). "Field experiment," Journal of Economic Literature, 42, p1009-1055.

Holmstrom, B. and P. Milgrom (1991). "Multitask principal-agent analyses: incentive contracts, asset ownership and job design," Journal of Law, Economics and Organization, 7, p24-52.

Johnson, R., D. Reiley and J. Munoz (2006). "The war for the fare: how driver compensation affects bus system performance," Working paper, University of Arizona.

Kane, M. and R. Engle (2003). "Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to stroop interference," Journal of Experimental Psychology, 132, p47-70.

Katz, L. (1986). "Efficiency wage theories: a partial evaluation," in NBER Marcoeconomics Annual, ed. S. Fischer, Cambridge, MA: MIT Press.

Kerr, S. (1975). "On the folly of rewarding A while hoping for B," Academy of Management Journal, 18, p769-783.

Lazear, E. (2000). "Performance pay and productivity," American Economic Review, 90, p1346-1361.

Meier, S. (2007). "Do subsidies increase charitable giving in the long run? Matching donations in a field experiment," Journal of the European Economic Association, 5, p1203-1222.

Paarsch, H. and B. Shearer (2000). "Piece rates, fixed wages and incentives effects: statistical evidence from payroll records," International Economic Review, 41, p59-92.

Reason, J. (2000). "Human errors: models and management," British Medical Journal, 320, p768-770.

Shearer, B. (2004). "Piece rates, fixed wages and incentives: evidence from a field experiment," Review of Economic Studies, 71, p513-534.

Wickens, C. (1980). "The structure of attentional resources," in R. Nickerson (ed), *Attention and Performance*; Princeton University Press: Princeton, NJ.

Appendix: Proofs

Proof of Proposition 1

The agent's problem is:

$$\max_{e_1,e_2}\left\{\bar{\alpha} + \beta e_1 - F(\bar{e} - e_2) - \frac{1}{3}[(e_1 - 1)^2 + (e_2 - 1)^2 + (e_1 - 1)(e_2 - 1)]\right\}$$

It is globally concave. The first-order conditions are:

$$e_1: \beta - \frac{2}{3}(e_1 - 1) - \frac{1}{3}(e_2 - 1) = 0$$

$$e_2: F - \frac{2}{3}(e_2 - 1) - \frac{1}{3}(e_1 - 1) = 0$$

Rearranging and solving yields:

$$e_1^*(\beta, F) = 1 + 2\beta - F, \qquad e_2^*(\beta, F) = 1 + 2F - \beta$$

The assumption that $F < 1$ guarantees that both efforts are positive. Finally differentiating with respect to β yields:

$$\frac{\partial e_1^*(\beta, F)}{\partial \beta} = 2 > 0, \frac{\partial e_2^*(\beta, F)}{\partial \beta} = -1 < 0 \ \blacksquare$$

Proof of Proposition 2

Highs have $(R, F) = (2, 0)$, lows have $(R, F) = (0, 1/2)$. Let (μ_R, μ_F) denote the agent's beliefs about the value of (R, F).

Notice that μ_F enters the agent's problem directly but that μ_R does not. On the other hand, R directly enters the principal's problem but F only enters the principal's problem via the agent's behavior, which is mediated by μ_F.

The first step is to demonstrate that if $1/4 < \beta < 1$ and given the proposed equilibrium beliefs, then it is optimal for the agent to behave in the manner described by the proposed equilibrium.

In the proposed separating equilibrium, we have that:

$$\mu_F = \begin{cases} \dfrac{1}{2} \text{ if } \beta > 0 \\ 0 \text{ if } \beta = 0 \end{cases}$$

Recall that under certainty:

$$e_1^*(\beta, F) = 1 + 2\beta - F, \qquad e_2^*(\beta, F) = 1 + 2F - \beta$$

Given imperfect information and equilibrium beliefs, we have that:

$$e_1^*(\beta, \mu_F(\beta)) = \begin{cases} \dfrac{1}{2} + 2\beta \text{ if } \beta > 0 \\ 1 \text{ if } \beta = 0 \end{cases} \quad e_2^*(\beta, \mu_F(\beta)) = \begin{cases} 2 - \beta \text{ if } \beta > 0 \\ 1 \text{ if } \beta = 0 \end{cases}$$

Thus if $1/4 < \beta < 1$ then both efforts are higher under the piece rate than under the flat compensation.

The next step is to show that given the proposed equilibrium beliefs, it is optimal for the principal to behave in the manner described by the proposed equilibrium, thereby rationalizing the agent's beliefs.

Recall that F only affect's the principal's decision via the agent's beliefs μ_F. Thus in equilibrium, the difference between the decisions of lows and highs is based on the difference in R.

First solve the high principal's problem $(R = 2)$:

$$\max_{\beta \geq 0} \{ 2e_1^*(\beta, \mu_F) + 2e_2^*(\beta, \mu_F) - \beta e_1^*(\beta, \mu_F) - DR \}$$

$$\beta > 0 \Rightarrow 2e_1^*(\beta, \mu_F) + 2e_2^*(\beta, \mu_F) - \beta e_1^*(\beta, \mu_F) - DR = (2 - \beta)\left(\frac{1}{2} + 2\beta\right) + 2(2 - \beta) - 2$$

$$= \frac{1}{2}(6 + 3\beta - 4\beta^2)$$

Maximizing $(6 + 3\beta - 4\beta^2)$ with respect to β yields $\beta^* = 3/8$; substituting this back in:

$$\beta = \frac{3}{8} \Rightarrow \frac{1}{2}(6 + 3\beta - 4\beta^2) = \frac{105}{32}$$

In contrast if a high principal offers $\beta = 0$:

$$\beta = 0 \Rightarrow 2e_1^*(\beta, \mu_F) + 2e_2^*(\beta, \mu_F) - \beta e_1^*(\beta, \mu_F) - DR = 4 > \frac{105}{32}$$

Consequently given the beliefs, highs prefer to offer flat compensation.

Now solve the low principal's problem ($R = 0$):

$$\max_{\beta \geq 0}\{2e_1^*(\beta, \mu_F) + 2e_2^*(\beta, \mu_F) - \beta e_1^*(\beta, \mu_F) - DR\}$$

$$\beta > 0 \Rightarrow \beta^* = \frac{3}{8} \Rightarrow u_P = \frac{169}{32}$$

In contrast if a low principal offers $\beta = 0$:

$$\beta = 0 \Rightarrow 2e_1^*(\beta, \mu_F) + 2e_2^*(\beta, \mu_F) - (\beta + Ds)e_1^*(\beta, \mu_F) = 4 < \frac{169}{32}$$

Consequently given the beliefs, lows prefer to offer a piece rate where $\beta = 3/8$. Taken together, these render the agent's beliefs consistent for the cases where $\beta = 0$ and $\beta = 3/8$. Beliefs for all other values of β are off equilibrium implying no consistency constraints from the agent's side, but we have shown that setting $\mu_F = 1/2 \; \forall \; \beta > 0$ rationalizes the principal's behavior.

The final step is to note that since highs set $\beta = 3/8$, the condition $1/4 < \beta < 1$ is satisfied, which was necessary for agents to behave consistently with the proposed equilibrium. ∎