



Performance pay and worker cooperation: Evidence from an artefactual field experiment[☆]

Stephen Burks^{a,b}, Jeffrey Carpenter^{c,b}, Lorenz Goette^{d,b,*}

^a Division of Social Science, University of Minnesota-Morris, 600 East 4th St., Morris, MN 56267, United States

^b Institute for the Study of Labor (IZA), Schaumburg-Lippe-Str. 5-9, D-53113 Bonn, Germany

^c Department of Economics, Middlebury College, Middlebury, VT 05753, United States

^d Research Department, Federal Reserve Bank of Boston, 600 Atlantic Ave., Boston, MA 02210, United States

ARTICLE INFO

Article history:

Received 15 April 2007

Received in revised form 24 December 2007

Accepted 16 February 2008

Available online 14 February 2009

JEL classification:

C72

C78

C93

D23

J33

J54

Z13

Keywords:

Field experiment

Social preferences

Compensation schemes

Sorting

ABSTRACT

We report the results of an artefactual field experiment with bicycle messengers in Switzerland and the United States. Messenger work is individualized enough that firms can choose to condition pay on it, but significant externalities in messenger behavior nonetheless give their on-the-job interactions the character of a social dilemma. Second-mover behavior in our sequential prisoner's dilemma allows us to characterize the cooperativeness of our participants. Among messengers, we find that employees at firms that pay for performance are significantly less cooperative than those at firms that pay hourly wages or who are members of cooperatives. To examine whether the difference is the result of treatment or selection we exploit the fact that firm type is location-specific in Switzerland and that entering messengers must work in performance pay firms in the U.S.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Economists have understood that the organization of work affects the monetary incentives of the workforce for more than three decades (Alchian and Demsetz, 1972), and for most of that time the standard assumption has been that money maximization is a sufficiently good description of the goals of agents to be used as a stand-alone proxy for all significant economic motivations. However, recent laboratory experiments have caused economists to rethink the link between monetary

[☆] We thank Colin Camerer, John List, Peter Matthews, and two thoughtful referees for comments. We gratefully acknowledge financial support from the MacArthur Foundation's Research Network on the Nature and Origin of preferences and grants to Burks by the University of Minnesota (Grant-in-Aid Program, Office of the Dean of the Graduate School) and to Carpenter by the National Science Foundation (CAREER 0092953). Andrew Anderson, Alex Bertolas, Kyle Carlson, and Tyler Williams provided excellent research assistance. The views expressed herein do not necessarily reflect the views of the Federal Reserve Bank of Boston or the Federal Reserve System.

* Corresponding author. Tel.: +1 617 973 2189; fax: +1 617 973 3957.

E-mail addresses: svburks@morris.umn.edu (S. Burks), jpc@middlebury.edu (J. Carpenter), lorenz.goette@bos.frb.org (L. Goette).

incentives and motivation.¹ Not only is there substantial evidence showing that subjects often discount material incentives and act according to social norms or preferences such as reciprocity (Fehr and Fischbacher, 2002), but also they sometimes ignore material incentives altogether and act based on intrinsic motivation (Falk and Fehr, 2002).

Teamwork in particular characterizes a work process in which the material incentives faced by individual employees can be confounded not only because individual and group monetary incentives draw individuals towards different actions, but also because *social preferences* can interact with material ones in unexpected ways. Social preferences are preferences over monetary outcomes informed not only by the desire for one's own well-being, but also by the beliefs, intentions, and well-being of others. They can reinforce material incentives, but they can also render them ineffective or even make material rewards counterproductive. Furthermore, there is potential for dynamic effects if preferences, whether material or social, are endogenous with respect to the organization of work (Bowles, 1998).

A standard recommendation of labor economists in response to the team production problem is that employers should pay for individual performance, when individual performance is sufficiently measurable, to align the incentives of individual employees better with those of the employer (e.g., Lazear, 2000). However, there are a number of possible problems with this prescription. First, even when production is individualized enough that performance pay can be used, the remaining teamwork features of the work process can cause interactions between employees to have a social dilemma character. Second, such a prescription does not account for the response of socially motivated employees to the implementation of performance pay. Economists have sound predictions for the response of egoists to performance pay, but laboratory evidence suggests that reciprocally, or otherwise socially motivated agents, will not respond in the same way (e.g., Fehr and Gächter, 2002). On top of these important issues, and the point of departure for the current study, performance pay itself may partially determine worker social preferences in ways that, at a minimum, have thus far not been modeled and may be deleterious to production.

We provide empirical evidence addressing whether pay schemes affect the distribution of social preferences in a team production setting in which individual production is highly measurable and some employers have chosen to implement performance pay. Specifically, we measure the degree to which performance pay increases the number of egoistically categorized bicycle messengers in a social dilemma experiment where the social framing is provided by their workplace relationships. To connect our evidence with that in the conventional experimental economics literature, we also compare the behavior of our field population to that of a student control group under the same protocol.²

Bicycle messengers are especially suitable subjects for this project because their work, delivering message parcels within a short time period, is essentially individualized production. Even though there are many distinct sub-tasks involved in the production of one delivery, the resulting individual performance is measurable enough that firms can choose to condition pay on it. However, even the work of messengers paid by performance has a significant teamwork component because some deliveries are better than others (i.e., pay at a higher rate, require less effort, or position the messenger more advantageously for later runs), and messengers can take actions to influence the deliveries they are assigned. Mis-assigned deliveries (e.g., when the closest free messenger does not make the delivery) create externalities for other messengers, and to the extent that such mis-assignments spatially or temporally misallocate productive resources, potentially also affect the profitability of the firm.

Our research has uncovered three major findings. First, we find that messengers are considerably more cooperative than students. Like Carpenter et al. (2005a), this suggests two things: that the workplace frame matters and that, taken with a number of other recent studies (e.g., List, 2004 and Carpenter et al., 2008), students in the laboratory tend to produce the lower bound of social preference measures.

Second, and perhaps most importantly, we find that the compensation scheme correlates with cooperative behavior. Messengers at firms that have adopted performance pay are more likely to behave completely selfishly in our experiment (a sequential prisoner's dilemma game) and also (correctly) anticipate this from other messengers at their firm. This result is not due to demographic differences between messengers at the different types of firms.

Third, we try to distinguish between different possible explanations for this correlation. Our empirical setup allows us to rule out self-selection of messengers based on their cooperative preferences into different types of firms. We also provide additional evidence to show that there are strong high-effort norms at non-commission rate firms and that informal norm enforcement is more prevalent at these firms. We use this evidence to argue that the observed difference in social preferences is the result of one of the following two channels: Either the compensation scheme affects the social preferences of the messengers directly (as in Bowles, 1998 or Levitt and List, 2007), or non-commission firms screen their applicants to select individuals who behave in a non-selfish way.

2. The structure and incentives of the courier industry

Bicycle messengers deliver parcels as small as letters and as big as boxes that require the rider to attach a trailer to his or her bike. Among the workers, a parcel assignment is known as a "tag." Many of the tags (among our subjects, on average 37%) involve ferrying legal documents around the financial centers of the cities in which we conducted the experiment.

¹ As one referee pointed out, these experiments corroborate a long tradition of industrial research on workplace relations (Marshall, 1890; Mayo, 1933; Barnard, 1938; Roethlisberger and Dickson, 1939; Roy, 1952; Bewley, 1999) and are the inspiration for a number of new studies (e.g., Bandiera et al., 2005).

² Hence, our study may be characterized as an *artefactual* field experiment according to the Harrison and List (2004) and Carpenter et al. (2005b) criterion.

However, the pay per tag and the effort required to deliver a parcel both vary. In general, the price charged to a sender depends on the number of “zones” the courier needs to cross to make the delivery, but even for commission paid couriers, the relationship between effort and remuneration is not perfect. For instance, legal tags tend to pay more for the time required, some buildings are easier to access for pickups or deliveries with a bicycle than others, and, in cities like Zurich and San Francisco, messengers may encounter non-trivial hills. In addition, on commission pay, empty miles carry a zero pay rate, and some tags go to destinations that impose a higher chance of a long empty run to the next pickup than do others.

Most couriers work on commission. The average commission paid messenger in our sample is 45% of the revenue the tag generates for the firm (40% is a common norm). Among the non-commission riders, some are paid an hourly wage (the average is \$17.71/hour among our subjects), and others are members of coops that share revenues based on the proportion of total work hours contributed. On average, the revenue sharing coops in our sample pay out 2.4 times a month and share 59% of total revenue.

In addition to variation in pay schemes, there is some institutional variation in the way that tags are allocated to couriers. Most couriers (94% among our subjects) receive assignments directly from a central dispatcher over handheld radios or phones. This procedure is known as *allocated dispatch*. Despite the primary control held by the dispatcher under this system, messengers can significantly affect their assignments at most firms by how they report their work progress and whether they respond to a particular call. The spatial distribution of pickup and delivery locations means that by reporting a tag done either early or late, the messenger can affect the dispatcher’s information about whether their location is optimal for taking on a particular new tag. Since it is usually not smart to try to respond to radios and phones while riding in traffic, a messenger can choose to ignore some calls without penalty.

Some firms have their radios set up so that each messenger can hear the assignments made to some or all of the others, which provides information that individual messengers can use either cooperatively or strategically. The remaining few couriers claim tags in the *free call* system, wherein a central dispatcher broadcasts each tag to all of the riders, and the rider who responds first makes the pickup.

Under both dispatch systems the individual material incentives at firms that pay commission are those of a common pool resource problem, although the degree of the problem is most severe under free call. Couriers can choose to “cherry pick” (free call) the best tags directly, or manipulate the dispatcher’s information to make getting good tags more likely (allocated dispatch), regardless of whether or not they are the best suited from the firm’s perspective to make the delivery. At firms that pay hourly wages or a share of total firm revenues, the incentives are those of a public goods game. Couriers can choose to lay back and avoid high effort tags, free-riding (literally) on the efforts of their coworkers. As a result the degree of egoistic behavior of its messengers can affect the profitability of the firm because uncooperative behavior can lead to the misallocation of messenger resources both in time and over the geographic area served by the firm.

3. Our field experiment

We used a strategic form of the sequential prisoner’s dilemma game (PD) to measure the cooperative predispositions of the bicycle messengers who took part in our experiment. Each messenger filled out a strategy sheet for their choice as a first-mover, and then for their choices, conditioned on first-mover actions, as a second-mover. They were informed that a coin toss would decide which of their roles would be activated when we made final payouts.

Playing the PD sequentially provides us with a well-defined typology of player strategies.³ In the simultaneous version of the PD one can not distinguish, for example, between egoists who simply play the dominant strategy and conditional cooperators who defect, but would have cooperated if their partner had done so. By looking at the second-mover strategy chosen in the sequential PD, one can identify four possible “types”: *Egoists*, who defect no matter what; *Altruists*, who cooperate regardless of what the first-mover has done; *Conditional Cooperators*, who cooperate only if the first-mover cooperates; and what we have termed *Wingnuts*, who defect if the first-mover cooperates, but cooperate if the first-mover defects. We interpreted “wingnuts” as participants who did not understand the experiment.

In theory, using the sequential form of the PD means that any expectations second-movers may have formed about the distribution of cooperative types in the population are irrelevant when implementing a conditional strategy. Despite this, and following Manski (2002), we chose to elicit expectations about the cooperativeness of the pool of players from each of our participants. To give our participants the incentive to think about their estimates, they were paid an additional amount of money for being close to the true distribution of choices. These expectations data provide us with another dependent variable that might be explained by differences in the employment contract. Not only might the distribution of social preferences be endogenous to the use of performance pay, but these differences might also correspond to differences in what our messengers expect from each other. For example, if more egoism occurs under piece rates, it would not be surprising if messengers were to show awareness of this in their expectations of how other messengers in their firm will react to social dilemma incentives.

We chose to maximize the amount of information we gathered from each of our participants by using the strategic form and asking participants to make choices in both first- and second-mover roles both because of the type identification issues discussed above and because the high effort and expense of field work with a small and widely dispersed subject pool make

³ This could, however, come at the cost of biasing our measure more towards cognition than affect, but the evidence does not suggest that this is a problem in the PD (Brandts and Charness, 2000).

Table 1
Participant (messenger) demographics.

	Obs.	Mean	Std. Dev.	Min	Max
Beliefs: fraction of Bs cooperating if A cooperates	252	0.67	0.26	0	1
Paid on commission	252	0.82	0.39	0	1
Dispatched by free call	249	0.06	0.24	0	1
Tenure: months at current job	252	36.48	34.94	0	162
Hours per week worked	250	27.48	14.39	1	80
Household income (in dollars)	251	22,411.35	11,310.28	5600	45,000
Swiss	252	0.55	0.50	0	1
Male	252	0.85	0.35	0	1
Age	252	29.60	6.47	19	52.50
Married	252	0.15	0.35	0	1
Student	252	0.34	0.47	0	1
Years of schooling	247	13.85	20.40	9	18

every datum valuable. This raises the question of how comparable our results are to those from simultaneous single-role experiments. To provide a link to the standard lab literature we ran control sessions with University of Zurich students.

We deliberately put a mild frame on our instructions with messenger subjects because we were specifically interested in our participants' cooperative predispositions at work.⁴ The extent of the frame was to label the two roles as “messenger A” and “messenger B,” and to refer to the choices as “cooperate” and “not cooperate.” The balance of the instructions was neutrally worded. Our goal in this approach was to let the subjects bring with them whatever frame was cued by the fact that they are playing with fellow messengers, that the physical location was associated with their work, and that the study was of messengers.⁵

The details of our protocol are as follows. We conducted the experiment at relatively high stakes: Mutual cooperation yielded \$30 each on top of \$15 for filling out our survey. The maximum a subject could earn was \$64 from not cooperating with a cooperator and accurately estimating the distribution of types in the population; the average payout was \$44 in San Francisco and 59CHF in Switzerland.⁶ We first handed out surveys to as many messengers as we could find in the places in which the messengers congregate. In San Francisco we dropped surveys at messenger service offices and then spent ten days hanging around at “the wall” in the financial district, a small public gathering area where the messengers gather to eat lunch. In Switzerland, we went directly to the breakrooms at the offices of the messenger services or to public areas immediately outside to distribute surveys and collect participant decisions. When returning the survey, the messengers were asked to stay for an additional 15 min to read and fill out the decision sheets for our experiment.

Because it was impossible to gather all our participants at once, we created a protocol that allowed us to collect one observation at a time. Participants read approximately one page of instructions and filled out six control questions. Once a participant had answered the questions correctly, (s)he was allowed to proceed. We first asked participants to give their expectations about how many of the other participants would cooperate if the respondent cooperated as the first-mover, and then repeated the question for the case in which the respondent did not cooperate as the first-mover. Participants then chose whether to cooperate or not as the first-mover.

The second task of each respondent was to submit a strategy as the second-mover. Participants were first asked to give their expectations about how many first-movers would cooperate. They then submitted a strategy: what they wanted to do if the first-mover cooperated, and what they wanted to do if the first-mover did not cooperate.

As we explained to the participants, to generate final payoffs all the responses we collected were pooled and matched at random, with first- versus second-mover roles assigned by the flip of a coin. Payoffs were determined by the intersection of the matched responses and the degree to which the expectations of the subjects were correct. We asked the participants to indicate whether they preferred to pick up their payoffs (in private) at a later pre-specified date and place, or to have their payoffs mailed.

4. Our participants

We gathered data from 252 messengers; 139 worked in Zurich, Basel, Bern, or Luzern, and the remaining 113 worked in San Francisco.⁷ Table 1 provides summary statistics from our participant demographics. Our messengers were rather

⁴ The exact wording of our instructions is available upon request.

⁵ As it turns out, this choice seems to have been sensible because our type data shows some “construct validity.” Specifically, egoists respond significantly differently than conditional cooperators and altruists to questions about their work and life experience such as I try to help out other messengers at work; How often do you lend small amounts to friends; How often do you give to panhandlers; I let messengers who work hard know that I appreciate them; If the weather is bad, I sometimes take it easy.

⁶ All experiments were conducted during the period from May through September, 2003.

⁷ In San Francisco, the anonymous counterparts of each subject in the experiment were drawn from a group that matched the firm size of the subject, messengers either employed at small firms, at medium-sized firms, or at large firms. In Switzerland, where all the firms were medium or large, and two were unique to their home city, we grouped messengers within each firm.

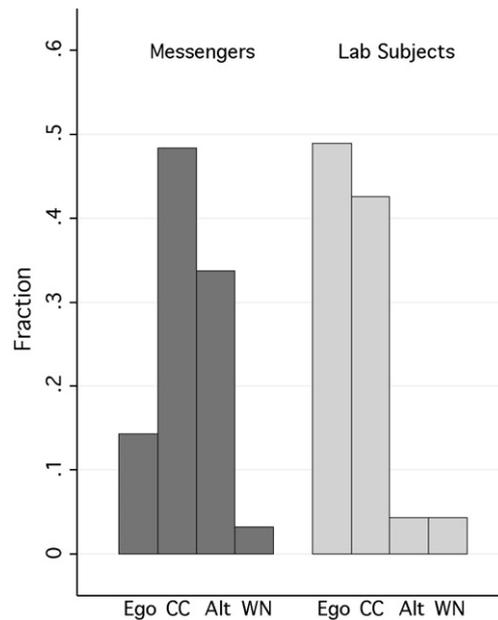


Fig. 1. The distribution of types by field or lab implementation.

optimistic about how second-movers would respond to a cooperative first-mover. On average, they expected that 67% of the other players would cooperate in the second-mover role if the first-mover cooperated. In fact, these expectations are not optimistic enough because 86% of the messengers cooperated in the second-mover role.

Performance pay is by far the norm among our participants; 82% were paid on commission. Even more common than performance pay is allocated dispatch. Only 6% of our participants were dispatched by free call.⁸ The average tenure at the current messenger job was a little more than three years. While 39% of messengers work full time (more than 34 h a week), many others work only a few shifts a week (the average is 27.48 h) and the average annual earnings of a messenger are \$22,411. As alluded to at the beginning of this section, our sample is relatively well balanced by location; 55% of the respondents were Swiss. Being a courier is a male-dominated occupation. In our sample 85% of the respondents were men. In terms of other standard demographics, our average participant was 30 years old, high school educated, unmarried, and one third were students.

5. Results

In this section, we report the results from the experiment described above. We first compare the messengers to 47 lab subjects to get a benchmark of how much messengers' behavior differs from that of lab subjects. This serves to examine whether, in an anonymous interaction, behavior is affected by whether one interacts with fellow students in an experiment or with a fellow worker. We then turn to the main result of the paper and examine how behavior in the experiment differs between differently organized firms. Finally, we offer a way to try to disentangle different interpretations of how compensation schemes can affect worker behavior through selection of different types of workers or through preference changes.

5.1. Differences between messengers and students

Fig. 1 displays the choices for the second-movers in our experiment based on the types that we described above. Thankfully, there are few wingnuts (3% of the messengers and 2% of the students), so we omit them from the analysis from this section onward. However, the histogram of the types also shows a very strong difference between messengers and lab subjects. There are far more egoists (35% more) and far fewer altruists (about 30% less) in the lab subject population. A Fisher's exact test confirms that the difference seen in the graph is statistically significant ($p < 0.001$).⁹ Thus, the subjects in the lab context are significantly more selfish.

⁸ We think these percentages reflect the underlying population of messengers. That is, we did not get the impression that we over-sampled commission couriers or under-sampled free call couriers.

⁹ When relevant, we discuss the results of two-tailed tests.

Table 2
Differences in experimental behavior and beliefs between lab subjects and field subjects.

		(1)	(2)	(3)
		Preferences		
Egoists:	Lab subject	0.354*** (0.076)	0.368*** (0.079)	0.261** (0.127)
Altruists:	Lab subject	−0.300*** (0.042)	−0.317*** (0.051)	−0.292*** (0.070)
Wingnuts:	Lab subject	−0.011 (0.024)		
		Beliefs (percent of Bs cooperating if A does)		
	Lab subject		−0.327*** (0.038)	−0.364*** (0.040)
Controls		No	No	Yes
Switzerland only		No	Yes	Yes
R ² (behavior)		0.056	0.092	0.116
R ² (beliefs)			0.189	0.306

Notes: The upper panel, labeled “Preferences,” displays the marginal effects from a multinomial logit model, treating conditional cooperators as the omitted category. The bottom panel, labeled “Beliefs,” presents the coefficient estimates for the beliefs using OLS with robust standard errors. *, **, *** indicate significance at the 10, 5 and 1 percent levels, respectively.

This raises the question of whether the messengers misunderstood the game, whether demographic differences between the two groups caused the difference in behaviors, or whether the difference in the context of the experiment (interaction with unknown student or with unknown coworker) is responsible for the difference.

There is no evidence that the messengers behave differently because they did not fully understand the instructions. As mentioned above, both populations generate small numbers of wingnuts, and regressions suggest that these fractions are not different. In the upper panel of Table 2, we present the formal estimates from a multinomial logit model, using the conditional cooperators as the reference category. For convenience, we present the marginal effects (i.e., the changes in the predicted frequency of a certain outcome) rather than the underlying coefficient estimates. The estimates in column (1) show that there is no difference in the frequency of wingnuts between messengers and lab subjects.

In our study, the Swiss messengers provide a clean opportunity to examine whether demographics or context explain the differences in the behavior, as many of them are also students and are thus demographically very similar to the subjects in the lab context. In column (2) of Table 2, we replicate the baseline estimate with only the Swiss subjects and, in column (3), add demographic controls. The point estimates of the marginal effects are virtually identical to those in column (1), showing that the Swiss messengers behave much like the US messengers. Most importantly, we control for whether a messenger is enrolled at a university. We then add gender, age, and income, as the messengers tend to be male and older than the students. For the sake of brevity, the table presents only the marginal effects of being a lab subjects (as opposed to being a messenger). As can be seen, the large differences in behavior remain, even when we control for demographic differences.

There are also substantial differences in beliefs between messengers and lab subjects. In Fig. 2, we display the histogram of the beliefs about the second-mover’s propensity to cooperate if the first-mover cooperates separately for the two groups.

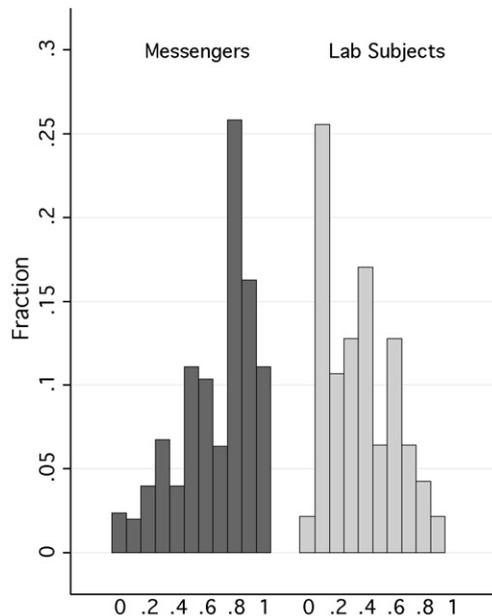


Fig. 2. The distribution of beliefs about how many second-movers will cooperate if the first-mover does in the sequential prisoner’s dilemma by field or lab implementation.

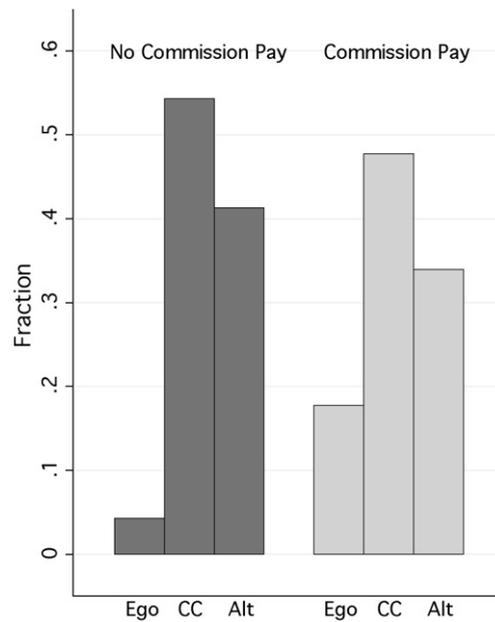


Fig. 3. The distribution of types in the sequential prisoner's dilemma by messenger compensation scheme.

The messengers have much more optimistic beliefs than the students. We present OLS regressions (with robust standard errors) to test formally for a difference in the beliefs in the lower panel of Table 2. The estimates confirm the qualitative picture from Fig. 2. There is a highly significant difference between the two groups. As with the second-mover behavior, the differences cannot be explained by demographic differences between the two groups.

We conclude that the messengers behave differently because of the work place context of the experiments, not because of their demographics. The difference can be seen in preferences (the second-mover behavior) and in the beliefs about other's preferences. As in Carpenter et al. (2005a), this suggests that the workplace frame matters and that it tends to make behavior more pro-social.

5.2. Differences between compensation schemes

We now ask the more subtle question of whether the contractual arrangements within a firm affect the behavior of workers in the experiment. Fig. 3 shows the histogram of second-mover types broken down by whether a messenger's firm uses commission pay or not. The figure shows a difference in the distribution of types between commission and non-commission firms.¹⁰ A Fisher's exact test rejects the hypothesis of equality of the two distributions ($p = 0.052$).

Visual inspection of the figure suggests that the distributions primarily differ in the number of egoists in the sample. Table 3 examines in more detail how the two distributions are different. The upper panel shows the marginal effects from a multinomial logit model with conditional cooperators as the reference category. As can be seen, the estimates indicate that the commission rate firms differ from the others only in the fraction of egoists present. Our estimates indicate that there are 12.5% more egoists in commission rate firms. The estimates also show that this effect is highly significant.

In column (2) of Table 3, we also include demographic controls. These controls are potentially important if commission rate firms differ in the retention or the age of the messengers (as explained in Section 3, revenue sharing firms in San Francisco hire only experienced messengers, for example). Importantly, the controls also include a country indicator. This is to ensure that the result is not driven by country differences in social preferences and firm organization. However, our results remain unchanged when we control for these differences. The marginal effect on egoists is virtually the same and still statistically significant.

In the previous analyses, we lumped together messengers paid by revenue sharing and messengers with fixed pay because both face the incentive to shirk. It is interesting to see whether the differences we find stem primarily from commission pay messengers being different than the rest or whether there are also differences between revenue sharing and hourly pay messengers. In columns (3) and (4), the reference group is now messengers on commission rate pay, and we examine how messengers on fixed pay and revenue sharing are different from that reference group. Whether we control for demographics or not, we find the same result: The difference in egoism is due to whether the messengers are paid a commission rate, and there is no difference in egoism between hourly paid messengers and messengers on revenue sharing ($p = 0.83$).

¹⁰ As noted before, we now omit the 8 wingnuts from the analysis.

Table 3
The effect of different pay structures on experimental behavior and beliefs.

		(1)	(2)	(3)	(4)	(5)	(6)
		Preferences					
Egoist: (0.042)	Commission	0.125*** (0.041)	0.123*** (0.042)			0.035 (0.069)	0.119* (0.070)
	Revenue sharing						
	Hourly pay						
	Tenure*commission			-0.108** (0.043)	-0.107** (0.042)		
	Tenure			-0.103** (0.044)	-0.118*** (0.038)		
						0.000 (0.003)	0.000 (0.003)
						-0.000 (0.003)	-0.000 (0.003)
Altruists:	Commission	-0.045 (0.081)	-0.054 (0.085)			0.035 (0.115)	-0.011 (0.126)
	Revenue sharing			0.066 (0.092)	0.101 (0.098)		
	Hourly pay			-0.002 (0.092)	-0.090 (0.098)		
	Tenure* commission						
	Tenure					-0.002 (0.003)	-0.001 (0.003)
						0.000 (0.002)	-0.001 (0.002)
		Beliefs					
	Commission	-0.073* (0.039)	-0.097*** (0.037)				
Controls		No	Yes	No	Yes	No	Yes
R ² (behavior)		0.012	0.055	0.014	0.065	0.018	0.056
R ² (beliefs)		0.012	0.076				

Notes: The upper panel, labeled “Preferences,” displays the marginal effects from a multinomial logit model, treating conditional cooperators as the omitted category. The bottom panel, labeled “Beliefs,” presents the coefficient estimates for the beliefs using OLS with robust standard errors. *, **, *** indicate significance at the 10, 5 and 1 percent levels, respectively.

We also examine whether the difference in behavior between the two types of firms arises immediately or whether it interacts with tenure. The results are displayed in columns (5) and (6): There is no significant interaction with tenure at the firm.

As a final step, we examine whether the difference in preferences that we documented is also reflected in the beliefs of the messengers in the two types of firms. Fig. 4 displays the histogram of the beliefs about the second mover’s likelihood of cooperating if the first-mover cooperated. The figure suggests that non-commission firms have somewhat higher beliefs than commission rate firms. The difference is significant, as can be seen in the estimates in the bottom panel of Table 3, and is robust to including a rich set of controls. The beliefs are approximately 10% lower in commission rate firms. This difference, again, is consonant with the difference in preferences between the two types of firms.

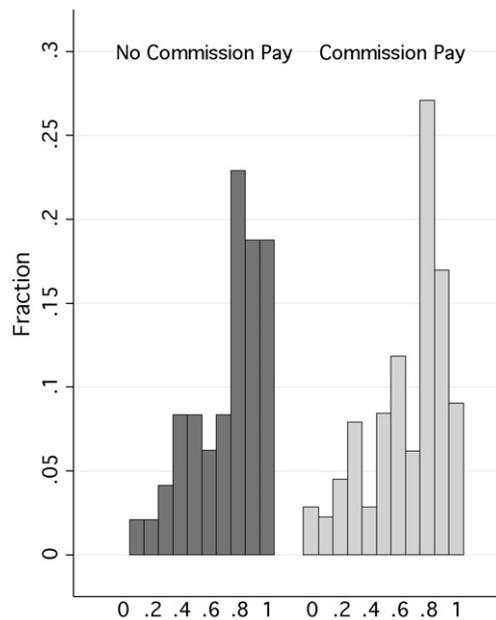


Fig. 4. The distribution of beliefs about how many second-movers will cooperate if the first-mover does in the sequential prisoner’s dilemma by messenger compensation scheme.

Table 4

The effect of different pay structures on experimental behavior and beliefs, restricted sample.

		(1)	(2)
		Preferences	
Egoist:	Commission	0.115** (0.055)	0.116** (0.053)
Altruists:	Commission	−0.020 (0.111)	−0.016 (0.115)
		Beliefs	
	Commission	−0.164*** (0.044)	−0.168*** (0.047)
Controls		No	Yes
R ² (Behavior)		0.009	0.095
R ² (Beliefs)		0.049	0.171

Notes: The upper panel, labeled “Preferences,” displays the marginal effects from a multinomial logit model, treating conditional cooperators as the omitted category. The bottom panel, labeled “Beliefs,” presents the coefficient estimates for the beliefs using OLS with robust standard errors. *, **, *** indicate significance at the 10, 5 and 1 percent levels, respectively.

5.3. Treatment or selection?

This difference in preferences and beliefs raises the question of how they should be interpreted. We consider three plausible mechanisms that could produce this outcome. First, it is possible that egoists prefer the commission rate firms, so endogenous selection based on the contract type could explain our results. Second, it is possible that non-commission rate firms screen more intensely along preference dimensions because these firms may be particularly vulnerable to egoistic opportunism. Thus, the difference should be interpreted as differential screening by the firms. Third, different payment schemes may change the workers' social preferences; cherry-picking and competition may simply make workers less considerate of each other.

In discussing these three potential mechanisms, we should make clear at the outset that our empirical setup is insufficient to rule out the first two reasons completely. In order to do this, we would need to assign messengers randomly to firms. Needless to say, no firm would be willing to agree to such an experiment. Nevertheless, our setup allows us to check the plausibility of two of these explanations.

We address the first issue (self-selection by messengers based on contract types) by considering only messengers who could not choose between differently organized firms. For example, in Switzerland, there are only commission rate firms in Zurich, Basel, and Luzern. By contrast, in Bern, the only bike messenger firm uses revenue sharing. Thus, living, for example, in Zurich, a messenger has no choice but to work in a commission rate firm. Our identification strategy is valid if not having the preferred compensation scheme does not deter a potential messenger from joining a firm. A sufficient condition for this to hold is to assume lexicographic preferences over jobs and then compensation schemes.¹¹ In San Francisco, rookie messengers can find employment only at large firms, which are all commission pay firms. Thus, by restricting the sample to Swiss messengers and rookie messengers from San Francisco, we can reexamine how behavior in the experiment differs between the differently organized firms, knowing that we have ruled out selection by messengers on contract types.

The results are presented in Table 4. Column (1) displays the multinomial logit results without demographic controls in place. There is still a significant difference in egoism between commission rate and non-commission rate firms. The point estimate of the effect (11% more egoists) is virtually identical to what we obtained using the full sample, although it is estimated somewhat less precisely due to the smaller sample. Column (2) of the table adds the control variables, but, as in the previous results, they do not matter very much. We still find a significant difference of virtually the same magnitude. The second panel in Table 4 shows the results for the beliefs. Similarly, we find a somewhat stronger effect using this restricted sample than using the full sample. The belief that the second mover cooperates in response to cooperation decreases by about 16 percentage points at commission rate firms. Compared to our result using the overall sample, the decrease is slightly stronger. Again, control variables make no difference, as can be seen in the second column.

The fact that these estimates for preferences and beliefs are so similar to the estimates we obtain from the full sample leads us to conclude that the differential choice of contract by the messengers does not seem to drive our results.

It is more difficult to distinguish between the other two candidates to explain the pattern we found: conditioning and screening. If non-commission rate firms screen applicants differently from commission rate firms, this introduces a form of selection that we are unable to address by comparing different subsamples of our data. We can, however, provide additional evidence that the two remaining explanations generate our results. We use our survey to show that there is a marked difference between firms in terms of work norms and norm enforcement that depends on the contract type. We asked the

¹¹ This assumption seems valid given 86% of our messengers state that “Working as a messenger is more than just a job”. There is also the possibility that these individuals chose to work as messengers in a different city in order to work in the preferred contract type. There are a number of reasons why this sort of selection is just implausible. One important reason is that travel times and distances between the cities are substantial, so the transportation costs would consume a large portion of the income from the mostly part-time work as a messenger.

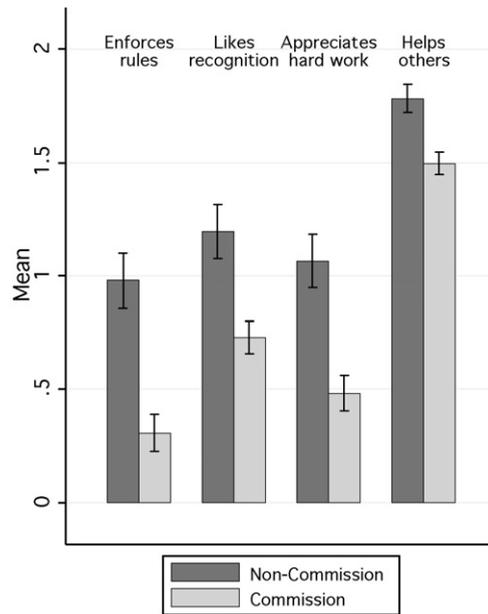


Fig. 5. Mean responses to surveyed behavior (measured on a -2,-1,0,1,2 Likert scale) by messenger compensation scheme. *Notes.* The following specific wording was used: “Enforces Rules”: “When another messenger at my firm doesn’t play by the rules, I let him or her know I don’t like it.” “Likes Recognition”: “I like other messengers to see that I am good at my job.” “Appreciates hard work”: “I let messengers who work hard know that I appreciate them.” “Helps others”: “I try my best to help out other messengers at work.”

messengers to respond to a battery of 14 statements related to their attitudes toward working hard, enforcing rules in their firm, and helping others. For each statement, the messengers had to indicate whether or not they agreed or disagreed on a scale from “disagree completely” (coded as -2) to “agree completely” (coded as 2), with “no opinion” coded as zero. We test for differences between commission rate firms and others, after adjusting the *p*-values for multiple hypothesis tests. Fig. 5 shows the average score for the two categories for the four items that were significant after correcting for multiple hypothesis tests (using the Holm, 1979 method).

The results show that there are differences in the items that measure enforcement of social norms (Question 7), susceptibility to social norms (Question 5), showing social approval (Question 13), and the propensity to help others (Question 1). As can be seen in the figure, there is a pronounced difference in the role social norms play between the two types of firms. The norms can be broadly interpreted as high-effort norms, and the evidence shows that these are more strongly present in non-commission rate firms and that they are more strongly enforced. In Table 5, we present ordered probit estimates, using the five answering categories of the questions as the dependent variable, without and with demographic controls. The controls are the same as in Table 3. The estimates show that, again, the differences are not explained by demographic differences across firms.

The pattern we document is consonant with Levitt and List, who argue that different environments (in this case, different contract types) cue different norms that affect behavior. However, it is also consistent with the interpretation that differential screening by firms generated this pattern, as it would be advantageous for non-commission firms to have employees who enforce high-effort work norms that act as a substitute for monetary incentives to exert effort. Thus, while we are unable to distinguish between the two, we provide supporting evidence that these two explanations most likely produce the differences we find and conclude that self-selection on the part of messengers is probably less important. Future research will have to sort out the relative importance of these two forces.

Table 5

Association between survey questions and working in a commission or non-commission firm.

	Enforces Rules	Likes Recognition	Appreciates Hard Work	Helps Others
Commission (no controls)	-0.642*** (0.176)	-0.541*** (0.180)	-0.582*** (0.178)	-0.611*** (0.221)
Commission (controls)	-0.695*** (0.183)	-0.550*** (0.186)	-0.638*** (0.185)	-0.705*** (0.233)
<i>R</i> ² (no controls)	0.018	0.014	0.015	0.019
<i>R</i> ² (controls)	0.044	0.025	0.047	0.041

Notes: Coefficient estimates from ordered probit models. The following specific wording was used: “Enforces Rules”: “When another messenger at my firm doesn’t play by the rules, I let him or her know I don’t like it.” “Likes Recognition”: “I like other messengers to see that I am good at my job.” “Appreciates hard work”: “I let messengers who work hard know that I appreciate them.” “Helps others”: “I try my best to help out other messengers at work.”

6. Concluding remarks

We conducted a sequential strategic-form prisoner's dilemma field experiment with workers (bicycle messengers) who work under social dilemma incentives and two different compensation schemes. Our goal was to see if social preferences are endogenous to whether the compensation scheme involved performance pay. In addition, we ran a separate control experiment with students to link our results to the laboratory literature on social preferences.

First, we find that messengers do exhibit preference heterogeneity, and that they are considerably more cooperative than students in our sequential prisoner's dilemma experiment. As in Carpenter et al. (2005a), this suggests that the workplace frame matters. Second, we find that the compensation scheme correlates with cooperative behavior. Messengers at firms that have adopted performance pay are less cooperative both in the game and also according to survey measures of on-the-job and off-the-job behavior, as compared to messengers at firms paying an hourly wage or a share of total revenues. Third, and perhaps most interestingly, we find evidence that the preferences of messengers may be endogenous to the employer's choice of compensation scheme. Working under performance pay may increase egoism. In fact, our regression analysis suggests that this "conditioning" effect is large, robust, and dominates any self-selection by egoists into performance pay firms. The effect is also immediate (within the limits of our survey-based measure of tenure length), which suggests that its character could be closer to cuing a behavior-guiding norm than to gradually adopting a new preference. These results must be qualified, however, by our inability to control for the possibility of firms (especially smaller coops and those that pay an hourly wage) differentially screening applicants by type.

The last finding, that the effect of the compensation scheme on behavior appears quickly, is consistent with another interpretation of our results. Levitt and List construct a model that is, in some sense, more complicated than the one we have in mind; instead of preferences determining behavior directly, in the Levitt and List model, the link between preferences and behavior may be attenuated by environmental cues. An alternative explanation of our data that is consistent with the Levitt and List model is that our participants have stable underlying preferences, but their behavior depends both on these preferences and normative cues from the workplace. In our case it is reasonable to think that the coops cue different norms than the piece rate firms. In the absence of more information on messenger preferences, which will be hard to disentangle from behavior, or more information on the strength of the environmental cues, it is impossible to differentiate one explanation from the other.

What are the potential implications for human relations in the workplace and, ultimately, for firm profitability? Rotemberg (1994) demonstrated that altruism could evolve in the workplace if there are strategic complementarities among workers, as is the case with our messengers. Our field experiment is consistent with this theoretical possibility, because the amount of pro-social behavior that we see is correlated with the pay scheme. We do not, however, find evidence of the sort of enlightened self-interest that is implied by Rotemberg's model (p. 686). That is, we see no evidence that our messengers consciously change their preferences.

At this point we can only speculate about the implications of social preferences for the overall profitability of firms, as we do not have on-the-job performance data for the subjects of the current study. While the empirical evidence offered in Lazear (2000) unambiguously suggests that performance pay increases profits when individual effort is identifiable and strategic complementarities among workers are weak, it is as yet unknown what happens when complementarities are more important and social preferences are present. It could be that performance pay "spoils" the pool of social preferences, so that when commission pay is introduced, between-worker cooperation may fall compared to hourly pay firms or revenue sharing cooperatives. This cost would reduce the net benefit of using performance pay. Indeed, if the complementarities in the production process are important enough, this could raise the possibility that in some settings, performance pay results in lower net profits for social preference reasons.

References

- Alchian, A., Demsetz, H., 1972. Production, information costs, and economic organization. *American Economic Review* 62, 777–795.
- Bandiera, O., Barankay, I., Rasul, I., 2005. Social preferences and the response to incentives: evidence from personnel data. *Quarterly Journal of Economics* 120, 917–962.
- Barnard, C., 1938. *The Function of the Executive*. Harvard University Press, Cambridge.
- Bewley, T., 1999. *Why Wages Don't Fall During A Recession*. Harvard University Press, Cambridge.
- Bowles, S., 1998. Endogenous preferences: the cultural consequences of markets and other economic institutions. *Journal of Economic Literature* 36, 75–111.
- Brandts, J., Charness, G., 2000. Hot vs. cold: sequential responses and preference stability in experimental games. *Experimental Economics* 2, 227–238.
- Carpenter, J., Burks, S., Verhoogen, E., 2005a. Comparing students to workers: the effects of social framing on behavior in distribution games. In: Carpenter, J., Harrison, G., List, J. (Eds.), *Field Experiments in Economics (Research in Experimental Economics)*. JAI/Elsevier, Greenwich, CT, and London, pp. 261–290.
- Carpenter, J., Harrison, G., List, J., 2005b. Field experiments in economics: an introduction. In: Carpenter, J., Harrison, G., List, J. (Eds.), *Field Experiments in Economics (Research in Experimental Economics)*. JAI/Elsevier, Greenwich, CT, and London, pp. 1–16.
- Carpenter, J., Connolly, C., Knowles Myers, C., 2008. Altruistic behavior in a representative dictator experiment. *Experimental Economics* 11 (3), 282–298.
- Falk, A., Fehr, E., 2002. Psychological foundations of incentives. *European Economic Review* 46, 687–724.
- Fehr, E., Fischbacher, U., 2002. Why social preferences matter—the impact of non-selfish motives on competition, cooperation and incentives. *Economic Journal* 112, 1–33.
- Fehr, E., Gächter, S., 2002. Do incentive contracts undermine voluntary cooperation? University of Zurich, IEW Working Paper, (34).
- Harrison, G., List, J., 2004. Field experiments. *Journal of Economic Literature* 42, 1009–1055.
- Holm, S., 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6, 65–70.
- Lazear, E., 2000. Performance pay and productivity. *American Economic Review* 90, 1346–1361.
- Levitt, S., List, J., 2007. What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives* 21 (2), 153–174.

- List, J., 2004. Young, selfish and male: field evidence of social preferences. *The Economic Journal* 114, 121–149.
- Manski, C., 2002. Identification of decision rules in experiments on simple games of proposal and response. *European Economic Review* 46, 880–891.
- Marshall, A., 1890. *Principles of Economics*. Macmillan, London.
- Mayo, E., 1933. *The Human Problems of an Industrial Civilization*. Macmillan, New York.
- Roethlisberger, F., Dickson, W., 1939. *Management and the Worker*. Harvard University Press, Cambridge.
- Rotemberg, J., 1994. Human relations in the workplace. *Journal of Political Economy* 102, 684–717.
- Roy, D., 1952. Quota restriction and goldbricking in a machine shop. *American Journal of Sociology* 57, 427–442.