OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 82, 3 (2020) 0305–9049
doi: 10.1111/obes.12344

# Wage Differentials and Workers' Effort: Experimental Evidence from Uganda*

JOHN SSERUYANGE and ERWIN BULTE

*Development Economics Group, Wageningen University, Hollandseweg 1, 6700 EW, Wageningen, Netherlands (e-mail: john.sseruyange@wur.nl, erwin.bulte@wur.nl)*

## Abstract

We organize a real-effort field experiment with varying piece rates to assess the impact of wages and social comparisons on productivity. In addition to analyzing how piece rates and social comparisons affect productivity during the 'paid stage' of the experiment, we also consider how they affect effort supply during a voluntary and unpaid follow-up task. Our main results are that effort supply is relatively unresponsive to variation in own earnings, but responds strongly to pay inequality. While we obtain weak support for the hypothesis that positive social comparisons invite extra effort during paid stages of the experiment, our most important finding is that social comparisons matter for voluntary tasks when shirking is cheap. Specifically, positive social comparisons positively affect productivity during unpaid tasks, and negative comparisons have the opposite impact.

## I. Introduction

While conventional economic models assume that people only care about their own income and consumption levels, a rich literature suggests most people also care both about how their income and consumption levels compare to those of others (e.g., Frank, 1985; Leibenstein, 1950; Duessenberry, 1952). The implications of concerns about one's relative position have been studied for a range of outcomes, including happiness and well-being. A related literature considers the behavioural effects of social comparisons, including studies of job choice and effort provision. Social comparisons among workers may occur within firms as well as in the wider labour market, affecting job satisfaction (Card *et al.*, 2012) so that relative pay can act as a compensating differential. Concerns about relative wages may also help explain wage compression (Charness and Kuhn, 2007), secrecy rules regarding earnings, or the sorting of heterogeneous workers across firms. Since within-firm workers make for a more salient reference group than outside workers (e.g. Clark and Senik, 2010), relative pay concerns also affect whether

*We thank three anonymous referees and editor James Fenske for their many helpful comments and suggestions which helped us to improve the quality of the manuscript. Remaining mistakes are our own. Financial support from the Netherlands Organisation for Scientific research (N.W.O, grant # 453-10-001) is gratefully acknowledged.

specific tasks are contracted out or organized within firm boundaries (Nickerson and Zenger 2008).[1]

In this paper, we consider how social comparisons affect effort supply. We are particularly interested in the effect of unequal treatment on effort supply during contracted and non-contracted interactions between the employer and worker. Gift exchange between employers and workers may attenuate moral hazard problems in contexts where complete labour contracts cannot be written. We focus on effort supply in a context where workers interact and can compare their earnings, and study their behavioural response in contexts where withholding effort is costly or not. Social comparisons may help individuals to decide whether they have been treated 'fairly'. The literature on the fair wage-effort hypothesis suggests effort supply is governed by a desire for reciprocity–workers respond to higher wages by working harder. The presence of reciprocal workers affect how labour markets work, and could result in non-competitive wages and involuntary unemployment (e.g., Akerlof, 1982; Akerlof and Yellen, 1990). Reciprocity in the workplace may be rationalized by reference-dependent preferences and social comparisons may help workers to form reference values. If social comparisons shift reference values they are expected to also affect effort supply in a context of imperfect contractibility.

When workers compare their earnings to those of their peers, two potentially conflicting fairness principles may come into play. 'Horizontal equality' prescribes that workers should earn the same wage, and the 'equity norm' says that workers exerting greater effort should receive a higher wage. These principles only coincide when workers are equally productive, else the equity norm implies 'unequal pay for unequal work' (Abeler *et al.*, 2010).[2] The equity norm appears to be more important for guiding behaviour of workers than concerns due to horizontal fairness. In an experiment where two workers were matched to one employer, workers chose greater effort when their individual effort was rewarded (compared to a scenario where workers received an equal wage–perhaps based on average performance). They also created greater surplus for the employer and themselves (Abeler *et al.*, 2010). In other words, workers do not simply compare earnings but also compare effort, and believe that effort differentials should invite compensatory payments. But it is often difficult to evaluate whether the equity norm is satisfied as monitoring the effort or productivity of other workers may be difficult for workers–especially if there are multiple dimensions to performance.

We contribute to the literature on gift exchange, social comparisons and productivity by examining whether absolute and relative earning levels affect effort supply in a real-effort field experiment in Africa. Our study is distinct for three reasons. First, in addition to focusing on effort supplied during the (paid) experimental task, we also elicit willingness to supply effort in a voluntary and unpaid task. This allows us to probe the effect of social comparisons beyond the quid-pro-quo relationship defined by the contract. Second, we

[1] Wage differences may of course create dynamic incentives, as workers may seek to acquire experience and invest in their human capital to increase their productivity and qualify for high wages in the future.

[2] A more complex problem emerges when productivity differences are not (only) due to differences in effort but also due to differences in talent. It is not obvious whether it is always possible to distinguish between effort and talent, and it is also not obvious whether talent should be rewarded. Nevertheless, fairness notions can be both effort-dependent as well as output-dependent. For an analysis of stated preferences over wage distributions and effects on productivity in a context of heterogeneous ability, refer to Cardella and Roomets (2018).

use an experimental design based on exogenous variation in *piece rates*, rather than in the hourly wage as in most other studies. Bracha, Gneezy and Loewenstein (2015) also analyse the implications of varying piece rates, but focus on labour supply (time spent on a task), rather than productivity, or output per unit of time.[3] Third, we use a non-conventional sample of African respondents–mainly smallholders and casual workers–that is distinct from the ones used in existing work. This allows us to verify the robustness of insights obtained earlier, in a different cultural context.

An important reason for studying piece rates in our experiment is the simple fact that piece rates represent the main payment regime in the context of informal or semi-informal (casual) labour markets in developing countries. Moreover, while unequal fixed wages are consistent with the equity norm (in the case of unequal effort), unequal piece rates for the same type of output more obviously violate fairness concerns. A drawback of piece rate treatments, however, is that it is not straightforward for the analyst to identify reciprocal behaviour by workers. Extrinsic motives for effort are determined by the level of the piece rate, so distinguishing between extrinsic and intrinsic motives is difficult. As mentioned by Bellemare and Shearer (2009), gift exchange occurs in a piece rate setting as long as the employer's valuation of the produced output is sufficiently high (exceeding the marginal cost due to the piece rate), but it is difficult to assess whether this condition is satisfied. Therefore, to analyse gift exchange, we analyse effort supply during a voluntary (unpaid) follow-up task.

We obtain three results. First, effort choices are not very sensitive to *own earnings* during the paid stage of the experiment, and we also do not document evidence of gift exchange during the (first) unpaid stage. Second, *social comparisons* matter and relative wages may impact on effort supply. While we do not find that, relative to a control group, negative comparisons reduce effort during the paid stage of the experiment, we do document some evidence that positive comparisons make people more productive. Third, and most importantly, social comparisons affect productivity during follow-up tasks when effort is voluntary. High piece rate workers supply more effort and low-piece rate workers supply less effort during an unpaid task. An explanation for the latter finding is that 'shirking is cheap' because workers do not have to sacrifice their own income if they choose to be unproductive in the unpaid, voluntary stage of the experiment. This finding supports earlier studies based on fixed wages, where withholding effort is also privately cheap. The effect of positive comparisons carries over to settings beyond the contract, which is arguably an important result as many organizations depend on voluntary or badly monitored contributions of their co-workers for (economic) success.

The paper is organized as follows. In the next section, we briefly summarize the existing literature on social comparisons and effort supply, focusing on recent experimental studies. In section III, we introduce our experiment, summarize our data, and outline our identification strategy. Section IV contains our empirical results, focusing both on effort supply for the paid and voluntary tasks. The conclusions and discussion ensue.

[3] See also Cardella and Roomets (2018) for an analysis of unequal piece rates and productivity in a real effort experiment.

## II. Social comparisons and effort supply

Starting with seminal work of Fehr, Kirchsteiner and Rield (1993), many papers have analysed the fair wage-effort hypothesis, postulating that fairness considerations affect the supply of effort by workers if effort is not perfectly contractible. To decide whether she is treated fairly, the worker may compare the wage offer to a reference value. The formation of reference values is an important but not fully understood process. Reference values for fair wages may be based on the status quo (earlier payments; see Clark, Masclet and Villeval, 2010; Cohn et al. 2014) or on expectations about future payments (Abeler *et al.*, 2011). Alternatively, reference values may be based on the vertical distribution of the surplus between worker and employer (Hennig-Schmidt, Rockenbach and Sadrieh, 2010), on horizontal comparisons among peers engaged in the same activity (Abeler *et al.*, 2010),[4] or on a combination of these considerations. Falk and Knell (2004) propose that subjects may choose their reference standards to serve motives of self-improvement and self-enhancement.[5] Regardless of the exact origins of reference standards, own earnings are likely to be evaluated as 'fair' when exceeding the reference level. In contrast, earnings below reference standards may be seen as 'unfair' and invite feelings of disappointment or even anger.

The literature contains several explanations for why fairness evaluations may affect effort supply in the context of fixed wages. Clark *et al.* (2010) framed their results in terms of a concern for status, but observe that inequality aversion would yield similar results. Workers with high earnings supply more effort and incur a greater effort cost to reduce the gap between their 'net earnings' and those of others. Relatedly, Hennig-Schmidt *et al.* (2010) introduce tensions due to cognitive dissonance. To alleviate tensions between an offered wage and reference wage, workers may provide extra effort for wages exceeding the reference value, or shirk in case the reverse is true. An alternative explanation would be a model based on a reciprocity norm prescribing how much effort to provide in response to a certain wage level–where higher wages imply workers should work harder. The salience of such a norm could vary with perceived fairness of the wage level, so workers feel free to ignore the norm when feeling badly treated (see, e.g. the theory of 'motivated reasoning' discussed by Benabou and Tirole, 2016).

The empirical evidence for social comparisons as a determinant of effort is strong, but not overwhelming.[6] Several studies document that workers' effort does not always respond to co-workers' wages in a simple fashion (e.g. Charness and Kuhn, 2007; Hennig-Schmidt *et al.*, 2010; Bartling and Von Siemens, 2011). This may be due to the fact that assessing the appropriate reference value is not always straightforward. For example, fixed wage differentials may be justified when workers are not equally productive. When workers are paid a fixed wage, then paying higher wages to more productive workers appears reasonable. In the context of heterogeneous workers, the equity norm prescribes that wages paid to

[4] In the latter case, as discussed above, horizontal fairness or the equity norm may shape the formation of reference values.

[5] Self-improvement refers to the desire for 'upward comparisons' to improve own performance. In contrast, self-enhancement refers to the desire for 'downward comparisons' to feel good about oneself. Choosing reference standards, according to Falk and Knell (2004), amounts to balancing these opposing motives.

[6] The persistence of reciprocity and gift exchange between employer and worker over time has also been debated (e.g., Gneezy and List, 2006; Bellemare and Shearer, 2009; Kube, Maréchal and Puppe, 2012).

high-productivity individuals do not map automatically on reference values of low-productivity workers.[7]

Nevertheless, the majority of the empirical work supports the hypothesis that social comparisons affect effort supply. This has been established in studies based on observational data (Karnes, 2009; Clark *et al.*, 2010), and in lab games (Clark *et al.*, 2010; Gachter and Thöni, 2010; Bracha *et al.*, 2015).[8] Breza, Kaur and Shamdasani (2018) organize a field experiment in an Indian manufacturing firm, offering different daily wages with or without a clear justification for wage differences based on baseline productivity differences. While absolute wages of groups of workers are similar, reference values are manipulated by exogenous variation in wages of the relevant co-workers and by variation in the observability of productivity across tasks. The empirical evidence suggests negative comparisons–being paid less for the same task than your peer, without a clear rationale or justification–reduce effort supply. In contrast, positive comparisons do not increase effort on average.

This is consistent with experimental evidence by Cohn *et al.* (2014), who study effort supply by pairs of workers in response to wage cuts. Cutting both wages decreases performance (the fair wage-effort hypothesis), but unilaterally cutting only one wage decreases performance of the affected worker more than twice as much and leaves effort of the other worker unaffected. Cohn, Fehr and Goette (2015) also report that the 'removal of perceived unfairness', rather than positive reciprocity, determines performance. Workers who perceive themselves to be underpaid at the base wage increase their performance after their wage is increased, but those who feel fairly paid do not change their performance after a wage increase.

## III. Experimental design and data

We designed and implemented a real-effort field experiment in Kamuli district, Uganda, in November and December of 2015, and collected data in 9 villages. We arrived in the villages one day before the experiment and asked the chief to provide us with a census of the households. After randomly selecting households to participate in the experiment, we invited selected households to send one adult representative. More than 95% of the households complied, and declining households were replaced by another one randomly drawn from the same population. We clearly announced the experiments were for university research purposes, and not part of any electoral campaign starting up in preparation of the 2016 elections in Uganda. Participants were informed there would be an opportunity to earn money. To avoid within-village information spill-overs, we ran several sessions simultaneously, overseen by multiple enumerators (from Makarere University), and not consecutively. To avoid between-village spill-overs, the selected villages were geographically spread out and we never announced in which village, we would collect data subsequently.

[7] Gachter and Thöni (2010) also demonstrate that 'intentions matter more than consequences', or that workers more readily accept wage differentials that appear random than the result of discrimination by the employer. For earlier work on intention-based reciprocity, refer to Dufwenberg and Kirchsteiner (2004).

[8] The design of Bracha *et al.* (2015) is a bit different from the other studies. They do not study how much effort is allocated to a task (per unit of time), but ask how many minutes students want to work on a specific task when receiving a high or low piece rate.

In total, we recruited 600 respondents from a peri-urban environment to engage in the experiment: 400 subjects were randomly assigned to two treatment arms LOW and HIGH (or 200 subjects per arm), and 200 subjects enrolled in the two control arms (100 subjects per control arm) Within experimental arms, we allocated respondents to groups of 10 villagers, participating in the experiment together. Subjects were aware that assignment to experimental arms was random. All participants received a show-up fee of 5,000 Ugandan shillings (1 USD $\approx$ 3,300 UGS). This context is a setting that is intermediate between the conventional 'lab' without interaction between subjects, other than via the game, and the 'field' where subjects are unaware they participate in an experiment. Many subjects in our study spent time together before and presumably after the experiment, which enhances the saliency of social comparisons.

The experiment consisted of two stages (stages 1 and 2) or 4 sub-stages (1a, 1b, 2a, and 2b). In stages 1a and 2a, we asked the subjects to sort beans for 30 minutes, in exchange for payment.[9] They were handed 12 kilograms of mixed dried beans in one large container: 3 kilograms from each of four different types, known by their local names as follows: *nambale short*, *nambale long*, *NAADs*, and *Masavu*. These beans are local varieties, well-known to the great majority of the respondents. During stages 1a and 2a, we paid participants a fixed piece rate per gram sorted based on their type. Subjects from treatment arm LOW were paid a relatively low piece rate of 5 UGS per gram sorted, and subjects from treatment arm HIGH received a high piece rate of 9 UGS per gram. Importantly, we did not provide any explanation for this gap in piece rates. Subjects from the two treatment arms worked in nearby but different rooms, and during the sorting there was no contact between them. Subjects from the control groups also received a high piece rate (CONTROL-HIGH, 100 respondents) or low piece rate (CONTROL-LOW, 100 respondents). At the end of stage 1a sorted output was measured, and individual earnings were computed and announced to individual subjects.

Immediately after stage 1a, we implemented stage 1b, in which we invited subjects to engage in a voluntary task for which they would earn no money. We emphasized that people were free to leave for the break, but that it would be appreciated if they wanted to help by sorting some more beans. Those willing to engage in additional, unpaid bean sorting were invited to sort another type of beans (from another container) for an additional 5-minute interval. Subjects picked different types of beans from the mix for the paid and unpaid stages of the experiment.[10] After stage 1b, and out of sight of the subjects, we mixed all sorted beans for the next experimental session.

The difference between stages 1 and 2 of the experiment is in the information that participants have about their relative earnings. During stages 1a and 1b participants knew only their own piece rate Social comparisons entered in stages 2a and 2b, after players

[9] Hence, our subjects were not free to choose the time spent sorting beans. For example, Fehr and Goette (2007) study whether agents work harder when wages go up, allowing workers to choose the number of hours spent working as well as effort per hour. In their study, the number of hours increased, and effort per hour decreased. Our subjects have fewer degrees of freedom, so our predictions with respect to effort per unit of time are unambiguous. Fehr and Goette (2007) do not study social comparisons, but do document evidence in favour of reference-dependent utility.

[10] Specifically, they picked the *nambale long* type during the paid stage, and the *masavu* type during the unpaid task. Since sorting the latter type is much easier (unlike *nambale long*, it is quite distinct from the other three types of beans in the mix), productivity in the unpaid task is higher, all else equal, so we cannot compare productivity levels across paid and unpaid stages of the experiment.

(presumably) learned about the piece rates earned by their co-workers from the other group. This was accomplished in a quite natural way as follows. During sessions, two groups of subjects were always sorting simultaneously in two adjacent rooms. After stage 1b, we gave participants a 20-minute break and provided them with snacks and refreshments. During this break, participants from the two groups were brought to a common room where they mingled and discussed–recall that participants were from the same village and typically 'knew' each other rather well. We boldly assumed information about earnings would spread naturally across the participants during this break. This was indeed the case, according to informal exit conversations.

Participants from the two treatment arms were always matched with individuals from the other treatment arms – bringing together high- and low-piece rate workers. In contrast, workers from control groups were matched with other control group workers earning the same rate. Hence, social comparisons are only salient for subjects from treatment arms LOW and HIGH, and only in the second stage of the experiment. Specifically, we expect that after discussing their piece rates and earnings with subjects in arm HIGH, subjects in arm LOW learn they have been treated relatively badly even if absolute earnings were considerable—see below. In contrast, subjects from arm HIGH presumably feel privileged. We assume this feeling may translate into a desire to reciprocate, so that subjects in arm HIGH will supply more effort during stages 2a and 2b.[11] We empirically explore whether social comparisons affect productivity in follow-up tasks. An outline of the experiment is provided in Figure 1.

After the break, participants returned to their working environment, and engaged in a second round of 30 minute of bean sorting (stage 2a) and afterwards were again invited to contribute to an unpaid task (2b). During stage 2a, all subjects worked for the same piece rate as before, and sorted the same beans as before. After this second stage, we paid subjects their earnings.

Subjects participated in only one session. Mean earnings in LOW equalled about 1,600 UGS during the first stage, and mean earnings in HIGH equalled 3,000 UGS. During the second stage, LOW respondents earned on average 2,000 UGS and HIGH respondents earned 4,100 UGS. Mean earnings for the full experiment, including the show-up fee, range from 8,400 UGS for the respondents receiving the low piece rate to 12,100 UGS for subjects receiving the high rate. The former amount is roughly a full day of wages for unskilled labour in our study region, so even subjects in our low piece rate regime were paid rather well given that the experiment took less than half a day. This is an important consideration to which we return below. Also, observe that pay in our experiment is more unequal than in, for example, Breza et al. (2018). Comparison across studies is obviously hampered by the fact that we use piece rates and Breza et al. (2018) use daily wages, but it is of interest to note that our high earners, on average, earn some 50% more than their low-earning counterparts. In contrast, the difference in wage levels in Breza et al. (2018) is less than 5%.

We set out to test two hypotheses. First, we want to test whether gift giving occurs in the context that we study, and for this purpose we use data from stage 1. Since high

[11] Observe that, in what follows, we measure *productivity*, and not any 'desire to reciprocate'. Other mechanisms linking positive comparisons to effort may therefore also matter. For example, feeling privileged may lower the cost of effort.

| Treatment LOW (N=200) | Treatment HIGH (N=200) | CONTROL-LOW (N=100) | CONTROL-HIGH (N=100) |
|---|---|---|---|
| *Stage 1a* | | | |
| 30 min of bean sorting for low piece rate | 30 min of bean sorting for high piece rate | 30 min of bean sorting for low piece rate | 30 min of bean sorting for high piece rate |
| *Stage 1b* | | | |
| 5 min of bean sorting (no compensation) | 5 min of bean sorting (no compensation) | 5 min of bean sorting (no compensation) | 5 min of bean sorting (no compensation) |
| Break | | Break | Break |
| *Stage 2a* | | | |
| 30 min of bean sorting for low piece rate | 30 min of bean sorting for high piece rate | 30 min of bean sorting for low piece rate | 30 min of bean sorting for high piece rate |
| *Stage 2b* | | | |
| 5 min of bean sorting (no compensation) | 5 min of bean sorting (no compensation) | 5 min of bean sorting (no compensation) | 5 min of bean sorting (no compensation) |

Figure 1. Outline of the experiment

piece rates affect both extrinsic and intrinsic motives for labour supply we cannot use the productivity data from stage 1a. However, effort during the unpaid voluntary task in stage 1b is not confounded by extrinsic motives and presumably captures a desire to reciprocate. By comparing the productivity levels in stage 1b of subjects with low and high piece rates (in stage 1a), we can examine whether higher wages foster reciprocity:

*Hypothesis 1:* Workers receiving higher piece rates in stage 1a of the experiment will reciprocate and supply more effort during the voluntary follow-up task in stage 1b.

Identification rests on two additional assumptions. First, workers distinguish between piece rates of 5 UGS and 9 UGS, and do not treat these as equal.[12] Second, workers receiving the high piece rate have not reciprocated during stage 1a of the experiment (by working harder), so that they have repaid their mental debt already. We will probe the extent to which this is true by comparing productivity levels in stage 1a.

Our second hypothesis concerns social comparisons, for which we use data from stage 2 of the experiment after workers learned about their relative earnings. The literature suggests negative comparisons will reduce effort levels. But the literature is based on fixed wage payment regimes, where shirking is cheap. In contrast, shirking comes at the cost of reduced earnings during stage 2a of our experiment, which we speculate will attenuate the incentive to reduce effort and work less. Shirking is cheap, however, in stage 2b. If social comparisons matter for effort, high piece rates earners will supply more effort during the unpaid stage

[12] Workers might have difficulty grasping the implications of a specific piece rate for potential earnings (that is as follows: they may find it difficult to predict how many grams they can sort in half an hour, or how much they can potentially earn).

of the experiment, and low piece rate earners will supply less. To identify this effect, and avoid confounding it with gift giving discussed above, we use a difference-in-differences analysis. We do not compare the productivity of high and low piece rate earners to each other, but compare both to the productivity of their relevant control groups; workers earning high or low piece rates but who are uninformed about the earnings of others.

*Hypothesis 2:* Social comparisons imply that during the unpaid stage of the experiment: (i) low piece rate earners will supply less effort than subjects from their control group; and (ii) high piece rate earners will supply more effort than subjects from their control group. In addition (iii) we expect the former effect to be muted during the paid stage of the experiment when 'shirking is expensive'.

To test these hypotheses, we compare productivity levels for relevant subsamples, and analyse the data in a regression framework. To analyse gift giving, we can use data from stage 1b (N = 600) and estimate the following model:

$$y_i = \beta_0 + \beta_1 H_i + \beta_2 X_i + \epsilon_i \tag{1}$$

where $y_i$ captures the quantity of beans sorted by respondent i (in grams), $H_i$ is a dummy variable taking the value of '1' if the subject received a high piece rate (pooling respondents from HIGH and CONTROL-HIGH), $X_i$ is a vector of controls, and $\epsilon_i$ is an error term. We always cluster standard errors at the experimental group level. If higher piece rates invite additional effort during the unpaid task, then we find $\beta_1 > 0$.

We use data from the second stage (stage 2a and 2b) to estimate the following model:

$$y_i = \beta_0 + \beta_1 H_i + \beta_2 SC + \beta_3 SC \times H_i + \beta_4 X_i + \epsilon_i \tag{2}$$

where $SC_i$ is a dummy variable indicating whether the respondent was in one of the treatment arms (i.e. spent the break with subjects earning a different piece rate). This variable takes the value of '1' if subjects were matched with subjects earning a different piece rate, and takes a value of '0' for subjects from the control arms. The estimate coefficient $\beta_1$ again picks up the effect of own wage on effort, $\beta_2$ picks up the effect of unequal payment, and $\beta_3$ picks up any additional effect of unequal payment for high piece rate earners. Average quantities sorted for the various experimental groups during the 2nd stage of the experiment are therefore as follows:

$$\begin{aligned}
&\text{Low piece rate, control group: } \beta_0 + \beta_4 \mathrm{X} \\
&\text{High piece rate, control group: } \beta_0 + \beta_1 + \beta_4 \mathrm{X} \\
&\text{Low piece rate, treatment arm LOW: } \beta_0 + \beta_2 + \beta_4 \mathrm{X} \\
&\text{High piece rate, treatment arm HIGH: } \beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 \mathrm{X}
\end{aligned} \tag{3}$$

The top panel of Table 1 summarizes the raw data across the three experimental arms. A great majority of our subjects are (married) Christian women of the Musoga tribe, with little formal education. The average subject has approximately 4 children and has access to 1.7 acres for farming. Observe that participants are statistically identical for most observables, but that Christians and members of the Musoga tribe are slightly underrepresented in the control group, and that the control group has fewer children. We will control for observables in our regression models, but also use inverse-probability-weights to reweight our data in an effort to improve balance. If the reweighting is successful, then the weighted distribution

TABLE 1

*Descriptive Statistics and Balance Test*

| | *LOW N = 200* | *HIGH N=200* | *CONTROL N =200* | *P-values LOW = HIGH* | *P-values LOW = CONTROL* | *P-values HIGH = CONTROL* |
|---|---|---|---|---|---|---|
| Raw data | | | | | | |
| Male (dummy) | 0.270 | 0.295 | 0.285 | 0.580 | 0.640 | 0.755 |
| Married (dummy) | 0.630 | 0.645 | 0.640 | 0.756 | 0.769 | 0.883 |
| Number of children | 4.075 | 3.840 | 3.605 | 0.495 | 0.041** | 0.304 |
| Education (years) | 2.300 | 2.350 | 2.390 | 0.649 | 0.246 | 0.605 |
| Religion (dummy) | 0.805 | 0.820 | 0.720 | 0.702 | 0.008*** | 0.002*** |
| Tribe (dummy) | 0.930 | 0.900 | 0.880 | 0.283 | 0.031** | 0.386 |
| Land size (acres) | 1.760 | 1.720 | 1.580 | 0.898 | 0.111 | 0.214 |
| Inverse probability weighted data | | | | | | |
| Male (dummy) | 0.2865 | 0.2819 | 0.284 | 0.920 | 0.983 | 0.937 |
| Married (dummy) | 0.642 | 0.638 | 0.637 | 0.932 | 0.989 | 0.920 |
| Number of children | 3.889 | 3.819 | 3.922 | 0.837 | 0.888 | 0.955 |
| Education (years) | 2.353 | 2.3487 | 2.356 | 0.966 | 0.974 | 0.992 |
| Religion (dummy) | 0.784 | 0.7805 | 0.785 | 0.937 | 0.958 | 0.974 |
| Tribe (dummy) | 0.904 | 0.904 | 0.904 | 0.993 | 0.995 | 0.989 |
| Land size (acres) | 1.650 | 1.680 | 1.568 | 0.906 | 0.778 | 0.917 |

*Notes*: $**P < 0.05$, and $***P < 0.01$.

of each covariate should be the same across experimental arms. The bottom panel of Table 1 suggests this is the case. In what follows we present results both for the unweighted and weighted sample.

## IV. Empirical results

Before turning to our hypotheses and regression results, we first report pairwise comparisons of effort during the first paid stage of the experiment, capturing both intrinsic and extrinsic motives for effort supply. Sorted quantities for the various experimental arms during stage 1a are provided in the first row of Table 2.[13] The row contains two series of numbers, and the bottom series of numbers (in italics) are based on the weighted sample. Results are very similar. As expected, subjects in LOW and HIGH, sort the same quantity of beans as their counterparts in their control groups. Randomization in this dimension 'worked' and, in the absence of social comparisons, subjects in LOW sort the same quantity of beans as subjects in the control group receiving the same piece rate (columns 1–2; *P*-values of *t*-test of equality of sample means is reported in column 3). Similarly, subjects in HIGH sort as much as subjects in CONTROL-HIGH (columns 4–5; *P*-values from *t*-test reported in column 6).

More surprisingly, perhaps, is the finding that there is no significant difference in productivity between subjects in LOW and HIGH. While subjects in the latter group sort

[13] Appendix Figure A1 is a graphical representation of the quantities of beans sorted for all treatment arms separately, distinguishing between effort during stages 1a, 1b, 2a and 2b.

TABLE 2

*Productivity in Task 1*

| | *Low piece rate* | | | *High piece rate* | | |
|---|---|---|---|---|---|---|
| | *LOW (N=200)* | *CONTROL-LOW (N=100)* | *P-value (t-test)* | *HIGH (N=200)* | *CONTROL-HIGH (N=100)* | *P-value (t-test)* |
| Output paid task (stage 1a) | 324.5 (6.1) | 320.0 (10.7) | 0.71 | 338.6 (8.4) | 340.1 (11.1) | 0.92 |
| | *323.9 (11.0)* | *325.5 (12.8)* | *0.90* | *340.0 (13.9)* | *336.8 (10.9)* | *0.82* |
| Output unpaid task (stage 1b) | 114.0 (3.3) | 116.1 (3.9) | 0.69 | 118.6 (3.4) | 119.4 (4.4) | 0.89 |
| | *113.3 (5.3)* | *115.5 (4.1)* | *0.64* | *118.8 (5.7)* | *118.5 (4.5)* | 0.95 |

*Notes*: Grams of beans sorted for different experimental groups, standard errors of the means reported in parentheses. *P*-values refer to outcomes of a simple *t*-test. Results reported in italics are based on the inverse probability-weighted sample.

338.6 grams, or some 14 grams (4%) more on average than their counterparts from the low piece rate group, this difference is not statistically significant at conventional levels ($P = 0.19$).[14] Effort does not respond strongly to 'own wages'. See also the results that we obtain in a regression format, controlling for baseline characteristics and clustering standard errors, reported in Appendix Table A1.[15]

What explains the lack of an output response at the offered piece rates? For output to respond to incentives, incentives should be large (relative to variation at a given incentive level) and production should be sensitive to effort choice. While the latter condition is satisfied, it is possible that subjects were uncertain about the marginal benefit of supplying effort and therefore unclear about how hard they should work. Moreover, as argued by Araujo *et al.* (2016), if individuals are intrinsically motivated to exert lots of effort absent extrinsic incentives, then insensitivity in outcomes may be due to a 'ceiling effect'. More specifically, if respondents have very low opportunity costs to supplying effort, for example, because there is nothing else to do and they cannot leave the experimental site then they may simply 'try [their] hardest, with the aim of beating [their] personal best' (Araujo *et al.*, 2016, p.11).[16]

Both incentives and costs matter for effort supply. Goerg, Kube and Radbruch (2019) argue that the true opportunity costs for participants are the sum of implicit and explicit cost, and demonstrate the former is low when respondents have nothing to do but perform their assigned task. Any variation in costs therefore must result from the explicit costs, or the disutility from engaging in the task. Since the incentivized part of the experiment involved a duration of 30 minutes – which is long for a tedious task such as bean sorting – we believe explicit costs in our field experiment matter enough to potentially invite a sensitive output response.[17] This will be explored further below.

[14] The difference in output for high and low piece rate workers in the control group is also not statistically significantly different from zero ($\Delta = 4{,}5$ gram, $P = 0.19$).

[15] We pool data from treatment and control arms to increase statistical power. The HIGH dummy enters weakly significant in the models based on inverse probability weighting but not in the unweighted model.

[16] Goerg, Kube and Radbruch (2019) indeed document that the elasticity of output with respect to extrinsic incentives increases when participants can engage in some other activity (Internet browsing) or leave the premises.

[17] See Bulte, List and van Soest (2019) for supporting evidence that this null result is not driven by ceiling effects. When properly incentivized, subjects are able to provide more effort and are much more productive. Bulte *et al.* (2019)

**Gift exchange during the voluntary task**

In stage 1b of the experiment, we asked subjects to sort additional bags of beans without payment. This provides a natural setting to study reciprocity and gift giving extending beyond the contracted task. There are two different dimensions to reduced reciprocity: the extensive and intensive margin. Subjects may refuse to participate in the task altogether. However, such rejections were extremely rare and refusal rates hover around 1–3% across all experimental arms (refer to Appendix Table A2 for Lee bounds). More importantly from our perspective, subjects may participate in the task but supply little effort.

The second row of Table 2 documents that there is no variation in effort across the experimental arms during stage 1b – statistically these four measures of output are identical.[18] Higher piece rates in a contracted task do not improve productivity for an unpaid voluntary follow-up task. We find the same null result in a regression analysis estimating equation (1) (Table A1), suggesting a nuanced conclusion. The finding that nearly all subjects participated in the follow-up task is consistent with gift giving, but the lack of a correlation between the level of payment and effort in the follow-up task implies we have to reject hypothesis 1. This could reflect that both low- and high-piece rate workers believe they have been well-paid during the first stage.

**Social comparisons and effort choice**

Next, turn to hypothesis 2, or the effect of social comparisons on productivity in paid and unpaid stages of the experiment. We analyse data from stages 2a and 2b, after subjects in LOW and HIGH had the opportunity to compare their earnings to those of others. The average number of grams sorted for payment is provided in the top row of Table 3 (with numbers in italics, indicating, again, average quantities sorted by the weighted sample). Upon comparing the productivity in the top rows of Tables 2 and 3, it is clear that subjects increased their productivity across rounds ($P = 0.00$ according to a simple $t$-test). This applies to workers in both experimental arms as well as in the control groups. Workers have learned to sort more efficiently, or have obtained a better understanding of the monetary stakes during the experiment after receiving a signal about the marginal benefit of supplying effort after stage 1a.

The simple comparison of group mean suggests that productivity levels vary across groups after the break. First, observe that productivity of low piece rate workers does not suffer from negative comparisons. Specifically, for the unweighted sample we find that output in LOW increases by 27%, matching the increment in productivity of CONTROL-LOW ($P = 0.41$).[19] Outcomes are different for subjects in HIGH, who on average increased

also incentivize workers to sort beans and find that leveraging loss aversion in the so-called 'clawback' intervention increases output by some 25–30% for otherwise isomorphic incentives. That is, when subjects are paid up-front but have to return their earnings if they are not sufficiently productive, then they provide more effort.

[18] For this analysis we have included the non-compliers and gave them an output level of zero grams, so the results are akin to an intention to treat analysis. Similar results are obtained when focusing on the subsample of compliers (details available on request).

[19] It could be argued that the absence of a strong negative response in productivity by the low piece rate workers is due to a productivity spill-over of another type. Low piece rates may stimulate workers to catch up and work harder (Bowles and Park, 2005), offsetting the fairness effect pulling in the opposite direction. However, this cannot explain the productivity response for the unpaid stage of the experiment – discussed below.

TABLE 3

*Productivity in Task 2*

| | *Low piece rate* | | | *High piece rate* | | |
|---|---|---|---|---|---|---|
| | *LOW (N=200)* | *CONTROL-LOW (N=100)* | *P-value (t-test)* | *HIGH (N=200)* | *CONTROL-HIGH (N=100)* | *P-value (t-test)* |
| Output paid task | 412.9 (8.2) | 425.2 (13.2) | 0.41 | 461.5 (8.6) | 433.4 (11.3) | 0.05 |
| (stage 2a) | *412.9 (16.2)* | *423.0 (14.1)* | *0.53* | *463.7 (14.3)* | *433.0 (11.2)* | *0.03* |
| Output unpaid task | 123.4 (4.3) | 136.0 (4.5) | 0.07 | 153.3 (3.2) | 127.0 (4.4) | 0.00 |
| (stage 2b) | *122.8 (6.4)* | 136.8 (4.8) | *0.03* | *153.1 (4.4)* | *126.3 (5.4)* | *0.00* |

*Notes*: Grams of beans sorted for different experimental groups, standard errors of the means reported in parentheses. *P*-values refer to outcomes of a simple *t*-test. Results reported in italics are based on the inverse probability-weighted sample.

their output level by no less than 36%. This increase in output exceeds the increase in output of workers in CONTROL-HIGH ($P = 0.05$), as well as that of workers in LOW ($P = 0.00$). For both comparisons we find that the increase in productivity for HIGH workers is 9% points greater. In other words, while negative social comparisons do not significantly reduce productivity in our piece rate experiment, positive comparisons seem to invite an increase in productivity The same is true for a comparison of the weighted data.

The social comparison results for high-earners are not robustly significant in a regression context. (Qualitatively similar results are reported in Appendix Table A2, where we estimate the same model but include a measure of lagged productivity, or the quantity sorted in stage 1.) Consider the first two columns of Table 4, summarizing results for stage 2a for the unweighted and weighted sample. While coefficients $\beta_2$ and $\beta_3$ have the expected signs (i.e. $\beta_2 < 0$ or a negative productivity effect of negative social comparisons and $\beta_3 > 0$ or positive productivity effect of positive comparisons), they are imprecisely estimated. Specifically, $\beta_3$ is only significant at $P = 0.11$ in the model with unweighted data When using weighted data, however, we find that social comparisons increase productivity for high piece rate workers ($P < 0.05$).

The finding that negative social comparisons leave effort of the two experimental groups unaffected contrasts with findings of, for example, Gachter and Thöni (2010), Cohn *et al.* (2014, 2015) and Breza *et al.* (2018). These studies document an asymmetric impact of social comparisons on effort, finding that low-wage workers provide less effort and that productivity of high-wage workers is unaffected. For example, Breza *et al.* (2018) find that, on average, output declines by 22% for a given wage when a worker earns less than his co-workers,[20] and document little support for the hypothesis that performance increases when people earn more than their peers.[21] This divergence in findings is arguably explained by the piece rate regime, implying shirking is privately costly. Extrinsic motives appear sufficiently strong to dominate other concerns. Our findings are consistent with those of Cardella and Roomets (2018) who also study the effect of social comparisons in a real effort

[20] Workers with low relative pay also more frequently miss work days, and give up nearly 10% of their income to avoid a workplace where they earn less than their colleagues.

[21] See also Cohn *et al.* (2015) on differential effort responses to wage increases for underpaid and adequately paid workers and Card *et al.* (2012) on the asymmetric effect of wage inequality on job dissatisfaction.

TABLE 4

*Piece Rates, Social Comparisons and Effort*

| | *Beans sorted for payment (stage 2a)* | | *Beans sorted voluntarily (stage 2b)* | |
|---|---|---|---|---|
| H | 10.051 | 7.81 | −10.20 | −9.77 |
| | (18.25) | (17.05) | (6.91) | (6.46) |
| SC | −6.96 | −8.10 | −13.48 | −13.40 |
| | (16.33) | (15.79) | (7.47)* | (6.19)** |
| SC * HIGH | 36.72 | 42.44 | 40.35 | 40.04 |
| | (25.37) | (20.63)** | (10.47)*** | (8.24)*** |
| Controls | Yes | Yes | Yes | Yes |
| Constant | 423.50 | 421.65 | 120.77 | 120.68 |
| | (16.51)*** | (21.77)*** | (8.82)*** | (8.79)*** |
| Inv. Prob. weighting | No | Yes | No | Yes |
| R-squared | 0.117 | 0.119 | 0.087 | 0.087 |
| Observation | 600 | 600 | 600 | 600 |

*Notes*: Robust standard errors, clustered at the group level, reported in parentheses. Included controls are the variables included in Table 1. ***Coefficient significant at 1%, **Coefficient significant at 5%, *Coefficient significant at 10%.

piece rate experiment, and document a very small effect of comparisons on the amount of output produced.

Are social comparisons are more important when shirking is cheap? To probe the impact of social comparisons when these are not neutralized by extrinsic incentives we analyse productivity during the voluntary work stage 2b. Output levels for the unweighted and weighted samples are reported in the second row of Table 3. Productivity of low piece rate workers is now negatively affected by the social comparison. This result is significant for both the unweighted and weighted samples ($P = 0.067$ and $P = 0.03$, respectively). For voluntary tasks, we find that, compared to the low piece rate members of their control group, workers in LOW sort almost 11% less beans. This is comparable to findings in the literature based on variation in fixed wages: subjects supply less effort when shirking is cheap. The same result emerges in the regression analysis summarized in columns 3 and 4 of Table 4. Negative social comparisons lower voluntary output by 13–14 grams, or by some 10%.

By comparison, positive comparisons in HIGH extend beyond the 'work for payment stages' of the experiment, because these workers sort approximately 20% more beans than high piece rate workers from CONTROL-HIGH ($P = 0.00$ for both samples).[22] It is surprising that the productivity–augmenting effect of positive comparisons is nearly twice as large as the productivity-diminishing effect of negative comparisons. The positive comparison result is also different from the literature. Follow-up research should examine whether this difference is due to cultural factors, or due to the two-stage design used to measure (delayed) reciprocity.

Considering everything, extrinsic motives dominate effort supply decisions during paid stages of the experiment, and social comparisons have a relatively small role to play. However, subjects are aware of whether or not they have been treated 'fairly'. Both positive and negative comparisons affect effort supply during other stages of the work relation

[22] This immediately implies that, for the unpaid task, productivity in HIGH is much higher than in LOW ($P = 0.00$).

– beyond the contracted task – when shirking does not diminish own income. Whether social comparisons matter for the firm or not, depends on the completeness of the contract written with employees. This implies that we cannot reject hypothesis 2.

**Identification challenges**

While the study is based on experimental data with random assignment to wage levels and social comparisons, several identification challenges remain. For example, while subjects mingled during the break, they 'worked' in separate rooms. So subjects may believe that the difficulty of the task varied with the piece rate, which would attenuate incentives for social comparisons. However, the types of beans sorted during the experiment were well-known to the respondents (they are an important part of their daily diet), so during the break they could establish that the tasks were actually identical by discussing the details of their sorting experience.[23]

An important challenge to our interpretation is that the notion of vertical fairness (as a basis for gift exchange) might seem strange because the experiment lacks a real employer. Hypotheses 1 and 2 rest on the assumption that subjects believe they can reciprocate by working harder – that the employer in the experiment (the experimenter) places a value on additional output that is greater than the piece rate. We cannot verify the validity of this assumption, but made sure that the remixing of beans occurred out of sight of the participants so as not to jeopardize this idea.[24] However, if some subjects believe that the experimenter does not value sorted beans, then our findings are a lower bound of the potential impact of gift exchange and social comparisons in settings where vertical fairness is more obvious. Perhaps behavioural responses are to some extent driven by guilt aversion after positive comparisons, or a desire by high piece rate earners to simply 'live up to expectations'.

Another challenge emerges because of the level of wages that we paid. While we pay low and high piece rates to participants, even our low piece rate results in a wage that is 'high' – possibly too high for concerns about unfairness to emerge. On average, subjects in LOW earned 8,400 UGS for half a day or work, and subjects in HIGH earned 12,000 UGS. This seems to compare favourably to the region's wage for unskilled labour or 8,000 UGS per day. However, we hasten to add that this estimate of the opportunity cost of time is an underestimate of the true opportunity cost *during the time we collected our data* – the period just before the presidential elections in February 2016. Political parties organized several rallies in our study region, providing participants with money and in-kind gifts. The incumbent party organized infrastructure projects, serving the dual purpose of improving roads and creating (temporary) employment – temporarily bidding up wages. Hence, we found it very difficult to recruit subjects during pilot sessions when offering less than the

[23] Another potential concern is that the good news (positive comparisons) may spread more easily among the treatment group than bad news to avoid bad feelings among peers. In this case, LOW is less intensively treated than HIGH. This potential concern follows directly from our choice to let knowledge about relative earnings diffuse 'naturally' rather than via announcements of the experimenter.

[24] If subjects believe the experimenter did not care about sorted beans (for example, because the project was announced to be for research purposes), then effort is dictated by extrinsic motives only and gift exchange will not occur. As discussed above, however, this is not consistent with the patterns in our data: people voluntarily sort beans, and we do observe that social comparisons matter when shirking is cheap.

current show-up fee and piece rate. While wages in the experiment are 'high' compared to the daily wage under normal conditions, conditions were far from normal when we collected our data.

This concern therefore refers back to the complex issue of the formation of reference values. The literature provides several suggestions for what might constitute a reference point: status quo levels of assets or wealth (e.g. Kahneman and Tversky, 1979), rational expectations of consumption (Koszegi and Rabin, 2009), the average of lagged outcomes (DellaVigna *et al.*, 2017) or salient targets (Allen *et al.*, 2016). Our experiment was designed such that subjects compared their piece rates to those of the peer group in the adjacent room, building on earlier research demonstrating that experimenters can shift reference points with aid of simple information treatments (e.g. Abeler *et al.*, 2011; Bulte *et al.*, 2019).[25] Because the task at hand is well-defined and rather specific, it seems plausible that many subjects compare their earnings to others in the same position – their experimental peers.[26]

A final issue emerges because of our multi-stage design with delayed reciprocity during the voluntary task. If subjects believe they reciprocated during the paid stages of the experiment, the perceived need for additional reciprocity during voluntary tasks will diminish. Moreover, gift exchange wears off over time (Gneezy and List, 2006), resulting in a distorted reflection of true gift exchange in the paid phase. However, the stage 2 results are based on a diff-in-diff model so that, without delayed reciprocity mediated by social comparisons, subjects from the treatment arms should behave as their peers from the control arms. They do not, suggesting that preferences to reciprocate spill over across tasks temporarily. The experiment did not last long enough for gift exchange to evaporate–gift exchange may wear off after several hours (Gneezy and List, 2006), which is a longer time frame than the one employed in our experiment.

## V. Discussion and conclusions

Employers are willing to pay wages above the market-clearing wage if workers reciprocate by supplying extra effort. When effort is unobservable and non-contractible, such gift-exchange may be sustained as an equilibrium outcome if workers are altruistic (kind), inequality-averse, or wish to respect a reciprocal norm – motives that presumably vary across cultural contexts. In this paper, we analyse productivity in the context of a task where African workers are paid a piece rate, and where we vary piece rates across subjects to invite positive and negative social comparisons. Two innovations of the paper are as follows: (i) that we allow information about piece rate differentials to 'spread naturally' among our sample of workers, rather than artificially priming its salience during the instruction stage; and (ii) that we distinguish between paid and unpaid stages of the experiment to analyse gift exchange (varying the 'cost of shirking').

[25] For example, Abeler *et al.* (2011) shift reference values by manipulating expectations of workers, and Bulte *et al.* (2019) shift reference values by shifting the timing of payment (before or after execution of the task).

[26] But we acknowledge that the relevant reference group may be ambiguous, and that subjects could also consider fellow-villagers participating in a nearby political rally or working on the casual labour market during 'normal conditions.' If so, the low rate we offer may appear generous rather than unfair. To the extent that respondents use multiple reference points, or a weighted average of opportunity cost of time in and out of the experiment, the perspective of negative comparisons may be attenuated. This may explain the muted response to negative comparisons, relative to findings in other studies.

While productivity responses to variation in own piece rates are modest in our experiment, we conclude that social comparisons may be an important determinant of productivity in certain tasks Our results support the hypothesis of gift exchange between workers and employers, and point to social comparisons as an important mediating factor. In the absence of social comparisons, productivity differences between high and low piece rate workers are rather small (perhaps reflecting that payment levels were rather generous across all experimental arms). Introducing social comparisons does not fundamentally alter this finding, and outcomes during paid stages of the experiment appear to be dominated by extrinsic motives. We obtain some support for the hypothesis that positive social comparisons invite extra effort during paid stages of the experiment, but this finding is only marginally significant.

Social comparisons enter much more prominently when extrinsic motives are eliminated–during the unpaid stages of the experiment. Our subjects supply less effort in a voluntary task after discovering their earlier piece rate was relatively low. In contrast, high piece rate workers become more productive following a positive comparison relative to the control group. This suggests the consequences of social comparisons extend beyond 'contractible tasks' for which subjects receive payment, and that both positive and negative comparisons may matter for productivity.

Some of our findings deviate from earlier work, which emphasizes the attenuating effect of negative comparisons on own productivity. While perhaps part of this difference may be attributed to cultural differences across sample populations, we believe the nature of the payment regime helps to explain the divergence. Gift-exchange has been tested mainly in experiments based on hourly wages (and imperfect contracts), where working harder is readily interpreted as an act of altruism or reciprocal behaviour. However, since differences in hourly wages may be justified by underlying productivity differences, fixed wages may send an ambiguous signal about whether payments are fair or not. Instead, we base our study on exogenous variation in piece rates, implying less uncertainty about whether differences in pay are fair or not and also implying that under-supplying effort is privately costly. Our findings are consistent with a scenario where extrinsic motives for effort dominate intrinsic ones. Writing about their results based on variation in fixed wages, Cohn *et al.* (2014) p.897) write '… we cannot translate our findings to a piece rate environment. With such a schedule, workers would still have an incentive to produce output'. This is exactly what we find: any attenuating impact of negative comparisons appears to be dominated by extrinsic motives, but we also find that social comparisons re-enter elsewhere in the labour relationship.

The effect of social comparisons on productivity can be formalized in various ways, or there are multiple candidate mechanisms linking comparisons to the supply of effort. Candidate mechanisms discussed in the literature include inequality aversion, status concerns, kindness, a desire for reciprocity or fairness, or context-specific respect for behavioural norms. Further research is necessary to distinguish between alternative theories.

In the previous section, we discussed several caveats to our experimental design. Notwithstanding these issues, we hope our experimental findings speak to the design of earning structures within organizations. Earlier work suggests pay inequality may reduce overall satisfaction and willingness to cooperate (see Pfeffer and Langton, 1993, but also Bartling and Von Siemens, 2011 for evidence to the contrary). Our results imply that 'low

earners' will undersupply effort for voluntary tasks. Insofar as accurate and timely execution of such voluntary tasks by *all* workers is important to economic success, lowering the morale of part of the workforce implies an organizational risk. If favourable social comparisons invite reciprocal behaviour, it seems better to create a reference group *outside* the organization. If workers collectively compare themselves to workers from other organizations, and feel privileged or well-treated, this should arouse reciprocal behaviour. Interestingly, this implies a cross-organizational externality: one organization's earning structure will affect the morale and productivity of workers in other organizations. Understanding the strategic considerations implied by such outcomes is left for future research.

## Appendix A: Detailed Variables Definition

*Male:* dummy variable taking value '1' if participant is male.
*Married/Engaged:* dummy variable taking value '1' is participant is married or engaged.
*Number of children*: integer indicating number of children in the household.
*Education:* highest grade completed by the respondent.
*Religion (Christian):* dummy taking value '1' if participant religious affiliation is Christianity.
*Tribe (Musoga):* dummy variable taking value '1' if participant belongs to Basoga tribe.
*Land size:* land owned by the participant in acres.
*H* :dummy that takes a value of '1' for participants with high piece rate.
*SC:* dummy that takes a value of '1' if participants were matched with other participants earning a different piece rate.

TABLE A1

*Piece rates and effort*

| | *Beans sorted for payment (stage 1a)* | | *Beans sorted voluntarily (stage 1b)* | |
|---|---|---|---|---|
| High | 15.569 | 15.919 | 4.271 | 4.706 |
| | (12.911) | (8.690)* | (6.244) | (3.735) |
| Controls | Yes | Yes | Yes | Yes |
| Constant | 299.170 | 298.781 | 96.028 | 95.044 |
| | (16.857)*** | (16.888)*** | (7.992)*** | (8.075)*** |
| Inv. Prob. weighting | No | Yes | No | Yes |
| R-squared | 0.041 | 0.043 | 0.023 | 0.025 |
| Observation | 600 | 600 | 600 | 600 |

*Notes*: Robust standard errors, clustered at the group level, reported in parentheses. Included controls are the variables included in Table 1. The mean for the sorted beans (dependent variables) at stage 1 is 331.08 grams and 116.79 grams for the paid and voluntary tasks respectively. ***Coefficient significant at 1%, *Coefficient significant at 10%.

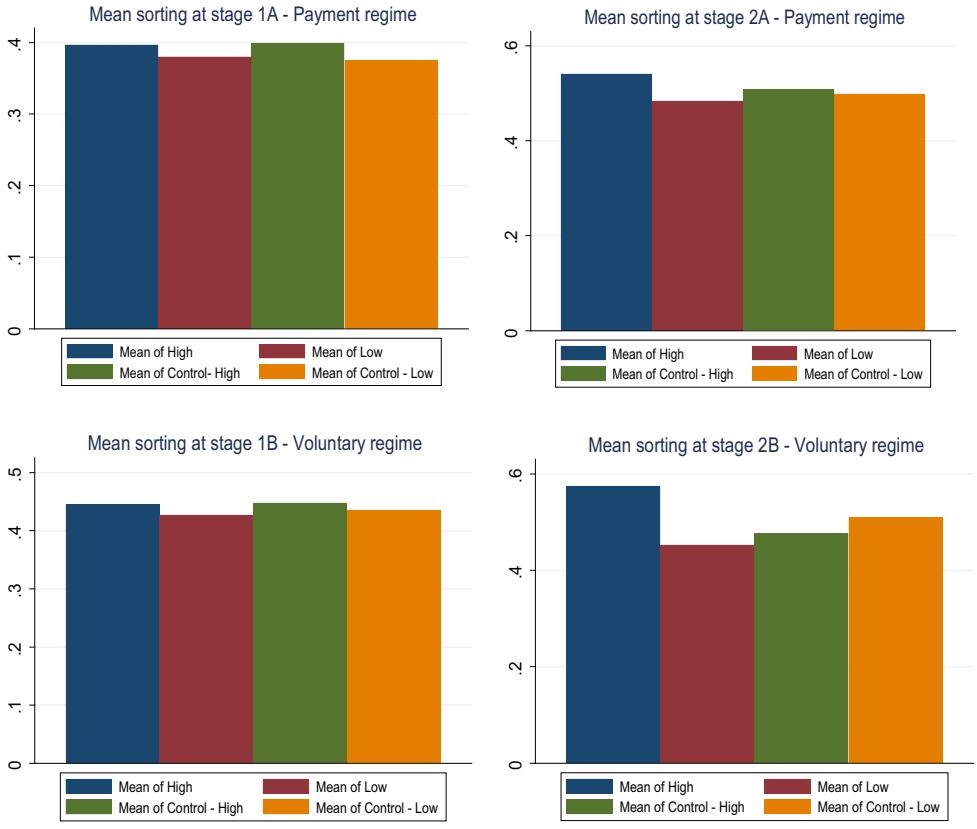


Figure A1. Grams of beans sorted at different stages by different treatment arms (Raw data are standardized by the highest quantity in each regime)

TABLE A2

*Robustness analysis*

| | *Beans sorted for payment (stage 2a)* | | *Beans sorted voluntarily (stage 2b)* | |
|---|---|---|---|---|
| High | 7.94 | −3.79 | −9.94 | −10.96 |
| | (17.07) | (13.14) | (6.46) | (5.98)* |
| SC | −8.03 | −9.10 | −13.48 | −12.36 |
| | (15.78) | (12.38) | (6.19)** | (5.78)** |
| SC * HIGH | 32.84 | 42.61 | 36.52 | 39.15 |
| | (20.41) | (15.67)*** | (8.18)*** | (7.74)*** |
| Sorting at stage 1a | | 0.72 | | |
| | | (0.04)*** | | |
| Sorting at stage 1b | | | | 0.38 |
| | | | | (0.05)*** |
| Controls | Yes | Yes | Yes | Yes |
| Constant | 426.47 | 206.82 | 120.89 | 84.18 |
| | (21.75)*** | (19.85)*** | (8.54)*** | (9.57)*** |
| Inv. Prob. weighting | Yes | Yes | Yes | Yes |
| Lee bounds | Yes | No | Yes | No |
| R-squared | 0.105 | 0.506 | 0.082 | 0.194 |
| Observation | 592 | 600 | 592 | 600 |

*Notes*: Columns (1) and (3) report Lee bounds (trimmed dataset). Columns (2) and (4) include a measure of lagged productivity. Robust standard errors, clustered at the group level, reported in parentheses. Included controls are the variables included in Table 1. ***Coefficient significant at 1%, **Coefficient significant at 5%, *Coefficient significant at 10%.

*Final Manuscript Received: September 2019*

## References

Abeler, J., Altmann, S. Kube, S. and Wibral, M. (2010). 'Gift exchange and workers' fairness concerns: when equality is unclear', *Journal of European Economic Association*, Vol. 8, pp. 1299–1324.

Abeler, J., Falk, A. Goette, L. and Huffman, D. (2011). 'Reference points and effort provision', *American Economic Review*, Vol. 101, pp. 470–492.

Akerlof, G. (1982). 'Gift exchange and efficiency wage theory: 4 views', *American Economic Review*, Vol. 74, pp. 470–492.

Akerlof, G. and Yellen, J. (1990). 'The fair wage efort hypothesis and unemployment', *Quarterly Journal of Economics*, Vol. 105, pp. 79–83.

Allen, E., Dechow, P., Pope, D. and Wu, G. (2016). 'Reference-dependent preferences: evidence from marathon runners', *Management Science*, Vol. 63, pp. 1657–1672.

Araujo, F., Carbone, E. Conell-Price, L. Dunietz, M. Jaroszewicz, A. Landsman, R. Lame, D. Vesterlund, L. Wang, S. and Wilson, A. (2016). 'The slider task: an example of restricted inferrence on incentive effects', *Journal of the Economic Science Association*, Vol. 2, pp. 1–12.

Bartling, B., Von Siemens, F. (2011). 'Wage inequality and team production: an experimental analysis', *Journal of Economic Psychology*, Vol. 32, pp. 1–16.

Bellemare, C. and Shearer, B. (2009). 'Gift giving and worker productivity: evidence from a firm-level experiment', *Games and Economic Behavior*, Vol. 67, pp. 233–244.

Benabou, R. and Tirole, J, (2016). 'Mindful economics: the production, consumption and value of beliefs', *Journal of Economic Perspectives*, Vol. 30, pp. 141–164.

Bowles, S. and Park, Y. (2005). 'Emulation, inequality and work hours: was Thorsten Veblen right?', *The Economic Journal*, Vol. 115, pp. F397–F412.
Bracha, A., Gneezy, U. and Loewenstein, G. (2015). 'Relative pay and labor supply', *Journal of Labor Economics*, Vol. 33, pp. 297–315.
Bulte, E., van Soest, D. and List, J. A. (2019). Towards an understanding of the welfare effects of nudges: evidence from a field experiment in Uganda. *Economic Journal*, In Press.
Breza, E., Kaur, S. and Shamdasani, Y. (2018). 'The morale effects of pay inequality', *Quarterly Journal of Economics*, Vol. 133, pp. 611–663.
Card, D., Mas, A., Moretti, E. and Saez, E. (2012). 'Inequality at work: the effect of peer salaries on job satisfaction', *American Economic Review*, Vol. 102, pp. 2981–3003.
Cardella, E. and Roomets, A. (2018), *Our Bark is Bigger than Our Bite: Stated Preferences Over the Distribution of Wages and the Effect on Productivity*. Texas Tech University, Working Paper
Charness, G. and Kuhn, P. (2007). 'Does pay inequality affect worker effort? Experimental evidence', *Journal of Labor Economics*, Vol. 25, pp. 693–723.
Clark, A., Masclet, D. and Villeval, N.C. (2010). 'Effort and comparison income: experimental and survey evidence', *Industrial and Labor Relations Review*, Vol. 63, pp. 407–426.
Clark, A. and Senik C. (2010). 'Who compares to whom? The anatomy of income comparisons in Europe', *Economic Journal*, Vol. 120, pp. 573–594.
Cohn, A., Fehr, E. and Goette, L. (2015). 'Fair wages and effort provision: combining evidence from a choice experiment and a field experiment', *Management Science*, Vol. 61, pp. 1777–1794.
Cohn, A., Fehr, E., Herrmann, B. and Schneider, F. (2014). 'Social comparison and effort provision: evidence from a field experiment', *Journal of the European Economic Association*, Vol. 12, pp. 877–898.
DellaVigna, S., Lindner, A., Reizer, B. and Schmieder, J. (2017). 'Reference-dependent job search: evidence from Hungary', *Quarterly Journal of Economics*, Vol. 132, pp. 1969–2018.
Duessenberry, J. (1952). *Income, Saving and the Theory of Consumer Behavior*, Harvard University Press, Cambridge, MA.
Dufwenberg, M. and Kirchsteiner, G. (2004). 'A theory of sequential reciprocity', *Games and Economic Behavior*, Vol. 47, pp. 268–298.
Falk, A., and Knell, M. (2004). 'Choosing the Joneses: endogenous goals and reference standards', *Scandinavian Journal of Economics*, Vol. 106, pp. 417–435.
Fehr, E. and Goette, L. (2007). 'Do workers work more if wages are high? Experimental evidence from a randomized field experiment', *American Economic Review*, Vol. 97, pp. 298–317.
Fehr, E., Kirchsteiner, G. and Rield, A. (1993). 'Does fairness prevent market clearing? An experimental investigation', *The Quarterly Journal of Economics*, Vol. 108, pp. 437–459.
Frank, R. (1985). *Choosing the Right Pond: Human Behavior and the Quest for Status*, Oxford University Press, New York.
Gachter, S. and Thöni, C. (2010). 'Social comparison and performance: experimental evidence on the fair wage-effort hypothesis', *Journal of Economic Behavior and Organization*, Vol. 76, pp. 531–543.
Goerg, S., Kube, S. and Radbruch, J. (2019). The effectiveness of incentive schemes in the presence of implicit effort costs. *Management Science*, In Press. https://doi.org/10.1287/mnsc.2018.3160.
Gneezy, U. and List, J. A. (2006). 'Putting behavioral economics to work: testing for gift exchange in labor markets using field experiments', *Econometrica*', 74, pp. 1365–1384.
Hennig-Schmidt, H., Rockenbach, B. and Sadrieh, A. (2010). 'In search of workers' real effort reciprocity – a field and a laboratory experiment', *Journal of the European Economic Association*, Vol. 8, pp. 817–837.
Kahneman, D. and Tversky, A. (1979). 'Prospect theory: an analysis of decision under risk', *Econometrica*, Vol. 47, pp. 263–292.
Karnes, R. E. (2009). 'A change in business ethics: the impact on employer–employee relations', *Journal of Business Ethics*', 87, pp. 189–197.
Koszegi, B. and Rabin, M. (2009). 'Reference-dependent consumption plans', *American Economic Review*, Vol. 99, pp. 909–936.
Kube, S., Maréchal, M. A., and Puppe, C. (2012). 'The currency of reciprocity: gift exchange in the workplace', *The American Economic Review*, Vol. 102, pp. 1644–1662.
Leibenstein, H. (1950). 'Bandwagon, snob, and Veblen effects in the theory of consumer demand', *Quarterly Journal of Economics*, Vol. 64, pp. 183–207.

Nickerson, J., and Zenger, T. (2008). 'Envy, comparison cost, and the economic theory of the firm', *Strategic Management Journal*, Vol. 29, pp. 1429–1449.
Pfeffer, J. and Langton, N. (1993). 'The effect of wage dispersion on satisfaction, productivity and working collaboratively: evidence from college and university faculties', *Administrative Science Quarterly*, Vol. 38, pp. 382–407.