DISCUSSION PAPER SERIES

# How Unjust!
# An Experimental Investigation of Supervisors' Evaluation Errors and Agents' Incentives

Lucia Marchegiani
Tommaso Reggiani
Matteo Rizzolli

December 2011

I Z A

# How Unjust!
# An Experimental Investigation of Supervisors' Evaluation Errors and Agents' Incentives

**Lucia Marchegiani**
*University of Rome 3*

**Tommaso Reggiani**
*University of Bologna, TSE and IZA*

**Matteo Rizzolli**
*Free University of Bozen-Bolzano*

# ABSTRACT

# How Unjust! An Experimental Investigation of Supervisors' Evaluation Errors and Agents' Incentives[*]

In our simple model the supervisor: i) cannot observe the agent's effort; ii) aims at inducing the agent to exert high effort; but iii) can only offer rewards based on performance. Since performance is only stochastically related to effort, evaluation errors may occur. In particular, deserving agents that have exerted high effort may not be rewarded (Type I errors) and undeserving agents that have exerted low effort may be rewarded (Type II errors). We show that, although the model predicts both errors to be equally detrimental to performance, this prediction fails with a lab experiment. In fact, failing to reward deserving agents is significantly more detrimental than rewarding undeserving agents. We discuss our result in the light of some economic and managerial theories of behavior. Our result may have interesting implications for strategic human resource management and personnel economics and may also contribute to the debate about incentives and organizational performance.

JEL Classification:    C91, M50, J50

Keywords:    agency theory, organizational justice, compensation, type I and type II errors, real effort

Corresponding author:

Tommaso Reggiani
University of Bologna
Department of Economics
Piazza A. Scaravilli 2
40126 Bologna
Italy
E-mail: tommaso.reggiani@unibo.it

# 1  Introduction

Rewarding the workforce is a central issue in several research fields that span from personnel economics to organizational behavior, to human resources management. Supervisors - whether managers, teachers, shareholders or otherwise - can manipulate rewards for observable performance as a means of incentivizing unobservable effort. High yet achievable performance goals, remunerated with non-negligible rewards, should induce rational agents to exert high effort. Unsurprisingly, if effort is only stochastically related with performance, evaluation errors occur. Imperfect monitoring can lead to two types of error:

- A supervisor (she) may observe high performance when in fact the agent (he) is exerting low effort. Therefore she may reward the agent undeservingly . This is a Type I error[1]

- A supervisor may observe low performance when in fact the agent is duly exerting high effort and thus she may not reward a deserving agent. This is a Type II error

Whenever individual unobservable effort is only stochastically related to observable performance, any reward system necessarily produces a certain balance of Type I and Type II errors. The supervisor can affect this balance by - for instance - setting different performance goals that the agent must match in order to obtain the reward. A very low goal implies a high probability that an agent exerting low effort is nevertheless rewarded (Type II error). This clearly demotivates the agent from exerting high effort. On the other hand, a very high goal implies a high probability that an agent willing to exert high effort is nevertheless not rewarded (Type I error). Both errors are thus detrimental to effort provision. What, then, is the optimal tradeoff between the two error types? This depends - *inter alia* - upon the elasticities of individuals' willingness to exert high effort to each type of error. Although the literature on the principal-agent relationship also addresses the tradeoff of the errors (Hölmstrom, 1979; Aron and Olivella, 1994) we are not aware of any study that specifically focuses on the relative impact on effort provision of the two types of error in performance evaluation. We tackle this research question through a lab experiment.

In the present paper we first introduce a simplified principal-agent model where we allow also for Type I errors. This model predicts that both errors should be treated equally as they both jeopardize performance by the same token. Then we test this prediction in the lab. We show that there is a substantial gulf between the theoretical predictions and the empirical lab evidence. In particular, failing to reward a deserving agent (Type I error) is significantly more detrimental to effort provision than rewarding an undeserving agent (Type II error). We then tentatively advance some explanations of this result in light of some behavioral and managerial theories. To the best of our knowledge, this is the first experiment where the effects of both Type I and Type II errors on effort provision and performance are studied. We believe that

---

[1]In an ideal contract with perfect monitoring the agent should receive an high remuneration whenever he exerts high effort. The agent's compliance with the prescribed behavior thus can be interpreted as the null hypothesis, so that the supervisor can both incorrectly reject the null and not reward a deserving agent (a Type I error) and incorrectly accept the null and reward an undeserving agent (Type II error).

this could shed more light on the incentive-effort-performance schema, and contribute both theoretically and to managerial practice.

# 2 Literature Review

Several streams of literature, including personell economics (Lazear, 1999), human resources management and organizational studies (Steers et al., 2004) deal with rewards, incentives and performance. The links between employers' effort, evaluation errors and pitfalls of the reward systems are epitomized by the agency model. This paper focuses on errors that are generated by the the stochastic relation between agent's effort provision and observable performance by the principal (Hölmstrom, 1979; Aron and Olivella, 1994; Prendergast, 1999). In this respect asymmetries of information and measurement errors can lead to opportunistic behavior and distortion in performance measurement. This in turn may affect the commitment of employees in accomplishing the tasks (Gibbons, 1998; Baker, 1992). Moreover supervisors evaluation may intentionally and deliberately be less accurate than what is feasible, because: i) supervisors may subordinate the objective of accuracy to other more self-serving and political goals and agenda (e.g., Cleveland and Murphy, 1992); ii) they may be distorted by employees' influence tactics (e.g., Higgins et al., 2003); iii) they may be biased by the social context of the performance evaluation process (e.g., Judge and Ferris, 1993).

Adapting the concept of identity to study trade-offs in supervisory policy, Akerlof and Kranton (2008) show that workers greatly resent supervision. As a result, they exert lower effort and may sabotage production.

In the management literature several streams of research analyze the effects of monetary incentives on employees' motivation and their ability to meet organizational goals (Steers et al., 2004). **Equity Theory** (Adams, 1965) is the first attempt at explaining the link between employees perceptions and organizational outcomes. Being based on the notions of cognitive dissonance, relative deprivation and social comparison (Festinger, 1957), it predicts that individuals are expected to compare their own input-to-output ratio to that of a referent. The referent could be the self, considered at another point in time, or others (other than self) either in the past, at present, or in the expected future. According to equity theory, if the compared ratios are unequal and thus perceived as "unfair", the individual may be motivated to engage in behaviors that will restore the cognitive perception of equality. Equity theory distinguishes between a) positive inequity and b) negative inequity, showing that both kinds negatively impact the employee's motivation, although negative inequity is comparatively more detrimental. Some more recent papers have shown the limits of this formulation, and the need for other complex mechanisms to be taken into account (Pritchard, 1969; Huseman et al., 1987). Among these, **Expectancy Theory** sees behavior as purposeful, goal directed, and largely based on conscious intentions. In this view, employees tend to rationally evaluate various on-the-job work behaviors (e.g., working harder) and then choose those behaviors they believe will lead to their most valued work-related rewards and outcomes (Mitchell, 1982;

Vroom, 1964). **Organizational Justice** perspective (Greenberg, 1987) distinguishes between distributive, procedural, and interactional justice (Konow, 2003; Cropanzano et al., 2007) the perceived fairness of rewarding schemes lays on: a) the allocations of reward, b) the means by which rewards are allocated; and c) the way the employer interacts with the employees. The relationships between employees' perceptions of fairness and organizational justice have been replicated in many empirical studies, in both the laboratory and the field (Viswesvaran and Ones, 2002), showing that the organizational justice framework fairly accurately predicts workers' behavior. In a sense, these approaches complement the Equity Theory in accounting for the possibility that the employee does not take only other employees as referents but values the justice of the outcome with respect to his own perceived input. Nonetheless, these views assume that there is a perfect correspondence between the employee's effort and the employer's assessment of her performance.

An original way of looking at evaluation accuracy is provided by the standard economic theory of deterrence, which suggests that the punishment of an innocent individual (Type I error) is no worse than the acquittal of a guilty one (Type II error), since both types of errors jeopardize deterrence by the same token (Polinsky and Shavell, 2007; Rizzolli and Stanca, Forth.).

Finally laboratory experiments have been successfully designed to test personnel issues, adding valuable insights to the existing body of knowledge (Falk and Fehr, 2003; Charness and Kuhn, 2010). Charness (2000), Falk and Kosfeld (2006) and Charness et al. (2011) show that not imposing a minimum effort level or allowing workers to delegate leads to better outcomes. On the other hand, Schnedler and Vadovic (2011) show that the "hidden costs of control" are mitigated if the principal's control is legitimate. In particular they show that when agents are given an initial endowment they do not resent effort requirements that simply prevent them from "pilfering" the endowment.(Dickinson and Villeval, 2008) show that the principal's decision to monitor the agent can also reduce agents' efforts, but only when the agents 'know' the principal.Fehr et al. (2007)compare performance of three types of contract. In line with the literature (Lazear, 2000) they find that incentive contracts (where deviation from high effort is punished with a certain probability) dominate trust contracts (where the principal simply trusts the agent to exert high effort). However, incentive contracts are dominated by bonus contracts (where the principal makes a promise of a bonus, conditional on the agent's high effort). Bonus contracts constitute the overwhelming majority of contracts offered, and yield higher effort/payoffs to principals. Irlenbusch and Sliwka (2005) show that the principal's choice of piece rates can also reduce agents' efforts.

# 3   Theoretical model

In the following paragraphs we model a simple relation between a supervisor and an agent where the supervisor cannot contract the agent's effort $e_i$ and only final output $q_i$ can be monitored.

The supervisor thus can evaluate performance by observing the level of output generated

by the agent and rewarding him accordingly.

**Agent's disutility of effort.** Let $e$ be a measure of effort. The agent's choice is binary: either he invests a low level of effort ($e_L$) or an high level ($e_H$) such that the set of possible actions is $(e_L, e_H) \in A$. Effort implies disutility for the agent. Following the literature, we define this disutility as a generic function of the level of effort $g(e)$ such that $g(e_H) > g(e_L)$, meaning that higher effort creates more disutility. Each agent has his own disutility of effort for both the low and high level and therefore it is associated with a particular level of $g(e_H)$ and $g(e_L)$. In this model, effort can take up only two values, therefore we can safely assume that $g(e_L) = e_L$ and $g(e_H) = e_H$ (see also Mas-Colell et al., 1995; Bolton and Dewatripont, 2005). We assume that each agent suffers the same disutility of effort $e_L$ and $e_H$. Although this is a simplifying assumption, it does not affect the results of the model[2].

**Agent's utility.** The agent is an expected utility maximizer with a utility function $u_i(w, e) = v_i(w) - e$ where $w$ is the wage and can take the following two values $(w_0, w_r)$ ($w_0$ is the baseline wage and $w_r$ is the rewarding wage) and where $e \in (e_L, e_H)$. The utility function is separable in monetary utility and disutility of effort following the usual assumptions of concavity for the former and convexity for the latter. Note that $v(w_r) - v_i(w_0)$ is the net reward. Each agent is uniquely associated with a value $\Delta_i = v_i(w_r) - v_i(w_0)$ that simply represents his monetary utility of the net reward.

**Supervisor's profit.** Let $q_i$ denote the performance in terms of the project's observable output from the agent as observed by the supervisor. The supervisor is risk-neutral and gains the project's output less any wage payment made to the agent: $q_i - w$ with $w \in (w_0, w_r)$ and with $q_i - w_r > 0$ so that the supervisor's profits are always positive.

## 3.1 Performance and error tradeoff

The agent's effort is not observable, hence it is only stochastically related to $e$. The supervisor decides the performance target/level of observed output $\bar{q}$ which triggers the rewarding wage $w_r$. Notice that the supervisor may commit the following evaluation errors:

Type I error: with probability $\alpha$, the agent that has spent a high level of effort does not meet the performance target and thus he is not rewarded; and

Type II error: with probability $\beta$, the agent that has invested a low level of effort nevertheless meets the performance target and thus he is undeservedly rewarded.

Every agent has a different utility of the monetary reward $\Delta_i$ and therefore, for some levels of $\bar{q}$, $e_L$, $e_H$, $\alpha$ and $\beta$, there will be only a certain proportion of agents willing to exert high effort in return for the utility of the monetary reward. It is reasonable to assume that $1 - \alpha > \beta$, that is to say that the probability of correctly rewarding the performance of the high-effort agent

---

[2]Heterogeneity among agents is preserved in the utility of income. As the relevant variable for us is given by the difference in the utility of income and the disutility of effort, we can simplify $e_H$ and $e_L$ to be the same for all agents.

is larger than the probability of wrongfully rewarding the performance of the low-effort agent. If this were not the case then the evaluation procedure would be equivalent to or worse than tossing a coin.

The derivation of the probabilities of errors from the definition of $q$ and $e_L, e_H$ is outside the scope of the present work. However, it is intuitive to say that the sum of errors $(\alpha + \beta)$ is minimized for some intermediate levels of $\bar{q}$. This is because when the performance target is set very low it is very easy to meet the target both with high and with low effort. Therefore with low $\bar{q}$ we have no Type I errors (i.e. not meeting the target when exerting high effort) and many Type II errors (i.e. meeting the target when exerting low effort). Therefore there is little incentive for the agent to invest high effort. The more the performance target increases, the smaller the probability of Type II error becomes and thus switching to high effort becomes convenient. At some intermediate level of $\bar{q}$ we have a few Type II errors and a few Type I errors. Finally, when the performance target becomes extremely high, the probability of Type II errors (i.e. meeting the target when exerting low effort) becomes virtually nil but at the same time the probability of Type I error (i.e. not meeting the target when exerting high effort) is very high and therefore there is little incentive to exert high effort.

## 3.2 Agent's choice

The payoff for the agent (utility of income less the disutility of effort) of complying with the supervisor's request to exert high effort are thus the following:

$$\alpha \left[ (v_i(w_0) - e_H \right] + (1 - \alpha) \left[ v_i(w_r) - e_H \right] \tag{1}$$

while if low effort is exerted the utility is the following

$$\beta \left[ (v_i(w_r) - e_L \right] + (1 - \beta) \left[ v_i(w_0) - e_L \right] \tag{2}$$

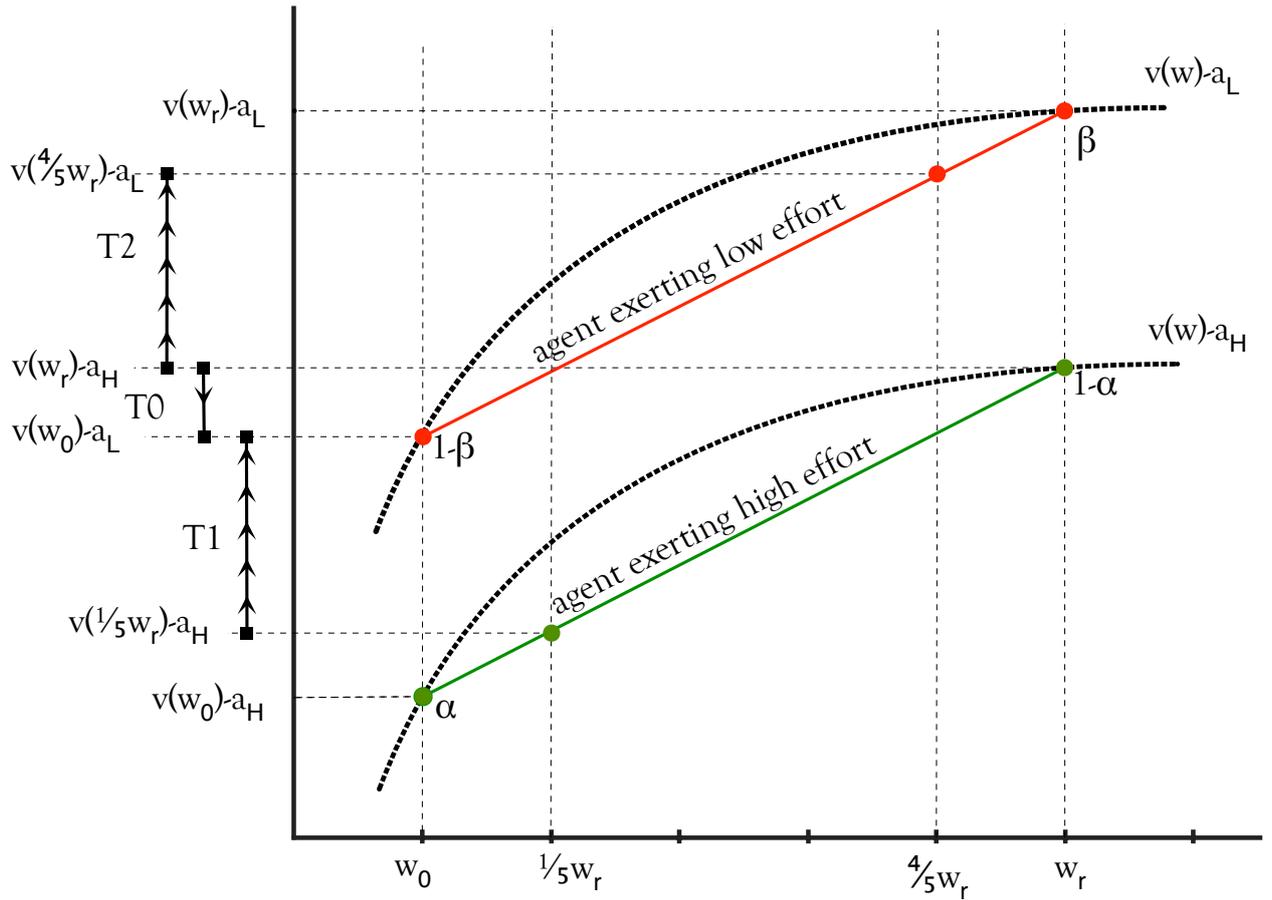The two choices can be graphically depicted as in Figure 1.

Figure 1: Utility function with low and high effort

Figure 1 depicts two standard concave utility functions $v(w) - e_L$ and $v(w) - e_H$. The second utility function is shifted downwards by the non-monetary cost of effort $(e_H - e_L)$ but otherwise it is identical to the first. The two chords represent the linear combination of payoffs given the probabilities of error conditional on exerting high or low effort.

The agent $i$ will comply with the prescribed behavior and choose to invest high effort if the expected payoff of high effort (Equation 1) is higher than the payoff of choosing to exert low effort (Equation 2). The equation simplifies as follows: $(v_i(w_r) - v_i(w_0)) (1 - \alpha - \beta) > e_H - e_L$ which implies the following compliance condition:

$$\bar{\Delta} = \frac{e_H - e_L}{1 - \alpha - \beta} \tag{3}$$

Equation 3 suggests that for each agent there exists a value $\bar{\Delta}$ for which the agent is indifferent between exerting high effort or low effort. Given that the disutility of effort and the probabilities of effort are exogenously determined, the choice of exerting high effort thus depends on whether each individual $\Delta_i \overset{?}{\gtreqless} \bar{\Delta}$.

Note that on the right hand of Equation 3 we have the net disutility of effort for the agent $i$ discounted by both Type I and Type II errors. Note also that on one hand the larger the probability of $\beta$ (being rewarded undeservingly), the larger are the returns from not exerting

effort. On the other hand, however, the larger the probability of $\alpha$ (not being rewarded when deserving it), the larger are the returns of exerting effort. We may define the sum of errors $\alpha + \beta$ as the accuracy of the evaluation process. Accuracy can be kept constant with very different error tradeoffs as long as $\alpha_{low} + \beta_{high} = \alpha_{high} + \beta_{low}$ . We will exploit this implication of the model as our treatment effect in the experiment.

## 3.3 Treatments

To illustrate the implications of the model we devise six treatments described in Table 1.

| | | $w_0$ | $\alpha$ | $\beta$ | $1 - \alpha - \beta$ | $(1 - \alpha - \beta)\left(v_i(w_r) - v_i(w_0)\right)$ |
|---|---|---|---|---|---|---|
| Small | $T0$ - Just | $w_{0-small}$ | $0$ | $0$ | $1$ | $\Delta_{small}$ |
| | $T1$ - Mean | $w_{0-small}$ | $4/5$ | $0$ | $1/5$ | $\frac{1}{5}v(\Delta_{small})$ |
| | $T2$ - Indulgent | $w_{0-small}$ | $0$ | $4/5$ | $1/5$ | $\frac{1}{5}v(\Delta_{small})$ |
| Big | $T0$ - Just | $w_{0-big}$ | $0$ | $0$ | $1$ | $\Delta_{big}$ |
| | $T1$ - Mean | $w_{0-big}$ | $4/5$ | $0$ | $1/5$ | $\frac{1}{5}v(\Delta_{big})$ |
| | $T2$ - Indulgent | $w_{0-big}$ | $0$ | $4/5$ | $1/5$ | $\frac{1}{5}v(\Delta_{big})$ |

Table 1: Table of treatment parameters

**Just Treatment ($T0_{small}$).** In T0 there are no evaluation errors ($\alpha, \beta = 0$) and thus the expected returns are $v(w_{r-small})$ and $v(w_{0-small})$ for exerting high and low effort respectively. The net returns of exerting high effort are thus $\Delta_{small}$. In this treatment the choice of exerting high effort is always rewarded while the choice of exerting low effort never is. This treatment is "just" in the sense that the agent gets what he deserves.

**Mean Treatment ($T1_{small}$).** In T1 there are no Type II errors ($\beta = 0$) but the probability of Type I error is significant ($\alpha = 0.8$).[3] Given the high number of errors, the net returns from exerting high effort are smaller but still positive $\frac{\Delta_{small}}{5}$. In this treatment the high-effort choice is seldom rewarded while the choice of low effort is never rewarded. This treatment is "mean" in the sense that the deserving agent very often does not get what he deserves.

**Indulgent Treatment ($T2_{small}$).** In T2 there is a significant probability of Type II error ($\beta = 0.8$) but there are no Type I errors ($\alpha = 0$). Given the high number of errors, the net returns from exerting high effort are smaller but still positive $\frac{\Delta_{small}}{5}$. In this treatment the high-effort choice is always rewarded while the choice of low effort also is very often rewarded. This treatment is "indulgent" in the sense that the undeserving agent very often gets what he does not deserve.

Moreover we devise three additional treatments ($T0_{big}$, $T1_{big}$, $T2_{big}$) where we increase both the baseline and rewarding wage by an amount $\theta$ so that $w_{r-big} = w_{r-small} + \theta$ and $w_{0-big} = $

---

[3]The choice of $\alpha = 0.8$ was made upon considering this probability high enough to be salient and clear "low" enough to leave space to the realisation of the complementary state-of-the-nature

$w_{0-small} + \theta$. Because of the usual assumption of decreasing marginal returns of income it must be true that $\Delta_{small} = v(w_{r-small}) - v(w_{0-small}) > \Delta_{big} = v(w_{r-big}) - v(w_{0-big})$.

Note that all treatments fulfill the assumption $1 - \alpha > \beta$.

## 3.4    Testable Predictions

For the purpose of the experiment we focus in particular on Equation 3 which describes the condition under which each agent may switch from low to high effort.

**Claim 1. Neglecting due rewards ($\alpha$) decreases agents' effort provision.** Equation 3 shows that Type I errors are detrimental to performance. By increasing the probability of Type I error in the experimental setting the model predicts that the share of agents performing with high effort (those for which $\Delta_i > \bar{\Delta}$) diminishes. In order to test this claim it is thus sufficient to confront $T0$ with $T1$.

**Claim 2. Rewarding undeserving agents ($\beta$) decreases agents' effort provision.** Equation 3 shows that Type II errors also are detrimental to performance. By increasing the probability of Type II error in the experimental setting the model predicts that the share of agents performing with high effort (those for which $\Delta_i > \bar{\Delta}$) diminishes. In order to test this claim it is thus sufficient to confront $T0$ with $T2$.

**Claim 3. Type I and Type II errors are equally detrimental to performance.** A given increase in the probability of Type II error $\beta$, compensated by an equal decrease in the probability of Type I error $\alpha$ leaves the individual indifferent in choosing whether to perform with high effort. The same is true when Type II errors decrease and Type I errors increase. In order to test this claim we will thus confront $T1$ with $T2$.

**Claim 4: Increasing both the baseline and rewarding wage by the same amount decreases agents' effort provision.** When both the baseline and the rewarding wage are increased by an amount $\theta$, the differential utility of exerting effort decreases $\Delta_{big} < \Delta_{small}$ because of the concavity of the utility curve. Therefore the share of agents for which $\Delta_i > \bar{\Delta}$ should diminish once the wages are increased by $\theta$ in all *big* treatments when compared with the *small* treatment. In order to test this claim the number of agents exerting high effort in $T0_{small}, T1_{small}, T2_{small}$ are confronted with the number of agents exerting high effort in, respectively, $T0_{big}, T1_{big}, T2_{big}$.

# 4    The experiment

The use of a lab experiment to test our theoretical predictions provides several important advantages (Falk and Heckman, 2009; Charness and Kuhn, 2010) in comparison with observational datasets that are typically used in labor/personnel economics or managerial case studies;

above all the opportunity to control for all the crucial variables of the economic environment and the possibility to vary *ad hoc* the precise variables of interest (Falk and Fehr, 2003; Falk and Gächter, 2008). On the other hand the external validity of lab findings can be questioned (Gneezy and List, 2006). However, the research question of the present work deals with a variable - evaluation errors - that is basically impossible to observe in the field because of the unobservability of effort and the stochastic relation between performance and effort. In the lab instead we can superimpose an exogenous probability of error in evaluating performance and at the same time we can perfectly observe effort. This ideally allows us to identify precisely the impact of errors on effort provision and thus on performance.

## 4.1 Experimental design

The experimental design is made up of three phases: the preliminary Phase I is used to elicit individuals' risk attitudes via a standard Holt and Laury (2002) incentivized choice of lotteries. This is followed by Phase II, where individual productivity in the default task is measured, and then there is Phase III, where individuals can choose whether to carry out the task. This last phase is our actual main treatment phase. Each individual carries out the task no more than once but choices are elicited under all three treatment conditions: $T0_{small}$, $T1_{small}$, $T2_{small}$ in sessions A and B (*Small* ) and $T0_{big}$, $T1_{big}$, $T2_{big}$ in sessions C and D (*Big*). The experiment adopts a within-subject design (which in its actual implementation is very close to a *strategy method* elicitation mechanism).

To control for any possible ordering effect, the order in which subjects are asked to make their choices under Treatments 0, 1 and 2 is randomized across the different experimental sessions. Feedback information on the outcomes of the lotteries in Phase I and on whether the supervisor-automaton makes an observational mistake are provided at the end of the experimental session. This is to assure full independence of the different treatment phases, free of historical contagion and therefore statistically independent across all subjects.

Between each of the phases, subjects have the opportunity to rest. Common instructions for the subsequent phase are read and described aloud while instructions concerning each treatment are delivered on screen. Control questions for each of the different phases are administered through the computer.

The experiment was programmed and conducted with z-Tree (Fischbacher, 2007) and sessions took place at the Einaudi Institute of Economics and Finance in Rome on April 6, April 8, April 14 and May 2, 2011.We ran a total of four sessions with 84 participants. Subjects were recruited online with ORSEE (Greiner, 2004). Wilcoxon tests indicate that there are no significant differences in the sociodemographic characteristics of the subjects across sessions: mainly undergraduate students with very different backgrounds (humanities, medicine, hard sciences, social sciences). Average age was 22.47 (std 2.16), females 40%, males 60%. Via the strategy method we elicited 84 observations for each treatment. Average payoff was about €10.21.

## Phase I - Risk attitude elicitation

Subjects carry out a standard Holt and Laury (2002) series of lotteries to measure individual risk-aversion. Outcomes of the lotteries are communicated to subjects only at the end of the experiment.

## Phase II - High effort productivity elicitation

Following Abeler et al. (2011), the work of the real effort task consists in counting the number of occurrences of the digit 1 in as many tables as possible, where each table is composed of 50 digits and among these the number of 1s is randomly generated (see Figure 6 in the appendix). This task has several advantages: it does not require any prior specific knowledge; performance is objective and easily measurable; and there is little room for learning effects. At the same time, the task is boring and pointless at least for most of the subjects and thus it can be claimed that the task entails a positive cost of effort. The task is also clearly artificial, and output does not provide intrinsic or extrinsic value to the experimenter. This should rule out any tendency for subjects to use effort provision during the experiment as a way to reciprocate for incentives provided by the experimenter or the possibility that subjects carry out the task for some intrinsic motivation. To elicit the individual productivity, subjects are offered a salient piece-rate compensation scheme. They receive a €0.03 payment for each table correctly processed and €0 for each table incorrectly processed. Furthermore both a countdown timer and a counter reporting the number of tables processed are provided. After 10 minutes subjects receive a summary statistical report concerning the number of tables correctly processed ($q_{phase-II}$), the number of tables incorrectly processed and the total amount of money generated in that phase ($q_{phase-II} \times$ €0.03).

The number of tables counted in this stage is a straightforward measure of subjects' performance under high effort ($q_{phase-II}$). In the following session, and under different treatment conditions (T0, T1 and T2) this measure becomes also the base for calculating the performance target $\bar{q}$ in Phase III. By scaling the individual specific target on individual ability we roughly normalize the individual cost of effort for the task. In other words, individuals with different abilities count different numbers of tables in Phase II. However, by setting individual targets of Phase III proportional on these numbers, the costs of effort for reaching the respective targets should be roughly equivalent.

## Phase III - Experimental treatments

Phase III is 40 minutes long; four times the length of Phase II. The performance target is set on 90% of four times the $q_{phase-II}$ measured for each subject in Phase II
($\bar{q} = q_{phase-II} \times 4 \times 0.9$). The 10% discount is justified by the higher fatigue created by the longer task and at the same time it signals that the task is feasible by exerting an high but not extraordinarily high level of effort[4]. Absent any error (as in T0) the accomplishment of the

---

[4]This is confirmed by the data as all subjects exept one that decide to exert effort eventually match the

task should be rewarded with €6.60 (This amount is proportional to an hourly wage of €10). We implemented the six treatments: $T0_{small}$, $T1_{small}$, $T2_{small}$ and $T0_{big}$, $T1_{big}$, $T2_{big}$

Subjects were presented with the following text common to all treatments:

. In Phase III you have to take a decision. You can decide (i) to undertake the task and meet the target of counting a certain number of tables in 40 minutes; (ii) to skip the task and proceed immediately to the payment phase and leave the lab. You may be rewarded with a <rewarding wage> payment for undertaking the task. The assignment of the reward is subject to errors. Under different situations you might be subject to, the supervisor might provide you the <rewarding wage> payment when you do not undertake or do not accomplish the task properly, conversely it might deny the <rewarding wage> payment when you actually duly accomplish the task.

Then we present the three treatments in succession. The text is tailored to each treatment (T0) [T1] {T2} as follows:

Your task consists in processing $<number\ \bar{q}>$ tables in 40 minutes and you are monitored through an automaton-supervisor. The supervisor *(will not)* *[may]* *{may}* commit an evaluation error.

(i) If you accomplish the task, you will *(certainly receive)* *[not receive with a 80% probability]* *{certainly receive}* the <rewarding wage> amount.

(ii) If you do not accomplish the task, you will *(certainly not receive)* *[certainly not receive]* *{receive with a 80% probability}* the <rewarding wage> amount.

Choose between "I will perform the task" and "I will skip the task".

Each subject is asked to state her choice (i or ii) for each of the three possible scenarios characterized by the treatment. Scenarios are submitted to subjects in random order. However, only one scenario is randomly selected[5] and its parameters applied.

The subject is informed about which scenario is actually randomly implemented (and thus whether the errors will occur), only after she makes her decisions for all scenarios. The design ensures that the subject makes truthful choices for the three scenarios. This is because the choices imply real opportunity costs: if the subject chooses (a) then she must spend 40 minutes in the lab anyway before progressing to the questionnaire and payment phase and if she chooses (b) then the real effort is skipped entirely. Therefore the subject has no reason to misrepresent her true preferences.

If a subject decides not to perform the task in a given treatment, and this treatment is then randomly implemented, she can proceed immediately to the next step - filling the questionnaire

---

performance target

[5]The design is thus a within-subject as we are able to observe the variations of subjects' effort choice across the three treatments and it implements the strategy method as the choices are elicited before one is randomly chosen and implemented. This procedure avoids income effects and also rules out any potential order effect of subjects' choice being influenced by previous decisions.

| Session | | Date | Treatment Order | Baseline Wage | Rewarding Wage | Observations |
|---------|---|------|-----------------|---------------|----------------|--------------|
| A | Small | April 4 | T0,T1,T2 | 0 | €6.60 | 23 |
| B | | April 6 | T0,T2,T1 | 0 | €6.60 | 17 |
| C | Big | April 8 | T0,T1,T2 | €5.28 | €11.88 | 23 |
| D | | May 2 | T0,T2,T1 | €5.28 | €11.88 | 21 |

Table 2: Experimental Sessions

in - and then she is paid and she can leave the lab, making thus alternative use of her time (leaving the lab and doing her own thing) in the case that she chooses not to perform.

We ran a total of four experimental sessions where we randomized the order of presentation and we varied both the baseline and the rewarding wage.Results

**Claim 1. Does neglecting due rewards ($\alpha$) decrease agents' effort provision?** In order to test this hypothesis we compare the share[6] of complying agents (define as $Z$) in $T0$ (just treatment with $\alpha = 0, \beta = 0$) and $T1$ (the mean treatment with $\alpha = 0.8, \beta = 0$). We test the following:

$$H_0 : Z_{T0} > Z_{T1} \text{ - vs - } H_1 : Z_{T0} = Z_{T1}$$

**Sessions *Small* (A & B):** Under perfect monitoring (T0), a share of population equal to 0.825 chooses to exert high effort to accomplish the task while the same share falls to only 0.325 when subjects are exposed to Type I error (T1). This effect is strongly statistically significant at the 0% level, on the basis of a two-sided null hypothesis and 40 independent paired observations (McNemar's[7] $\chi^2 = 20$, $p-value < 0.0001$).

**Session *Big* (C & D):** We repeat the same test for the sessions with higher wages. Under perfect monitoring (T0), a higher share of population (0.909) chooses to exert effort. When exposed to the Type I error (T1) only 0.545 of agents exert effort. This effect is still strongly statistically significant at the 0% level, on the basis of a two-sided null hypothesis and 44 independent paired observations (McNemar's $\chi^2 = 16$ , $p-value < 0.0001$).

Claim 2. Does rewarding undeserving agents ($\beta$) decrease agents' effort provison?

As before, in order to test this hypothesis we compare the share of complying agents in $T0$ (just treatment with $\alpha = 0, \beta = 0$) and $T2$ (indulgent treatment with $\beta = 0.8, \alpha = 0$). We test the following:

$$H_0 : Z_{T0} > Z_{T2} \text{ - vs - } H_1 : Z_{T0} = Z_{T2}$$

**Session *Small* (A & B):** Facing perfect monitoring (T0), 0.825 of agents are willing

---

[6]Or the units of effort exerted (number of tables correctly counted) by each subject in the different treatments.

[7]Our within-subject design enables us to observe the choices of subjects under all the different treatment conditions. The McNemar test fits particularly well with our experimental setting since paired-sample tests are used to assess the differences in the population shares of agents exerting effort under the different treatments. . See (Fehr et al., 2003; Enderer and Manso, 2009; Caplan et al., 2010) Analogous qualitative results on statistical significance in mean differences are replicated adopting a proportions test for differences in proportions.
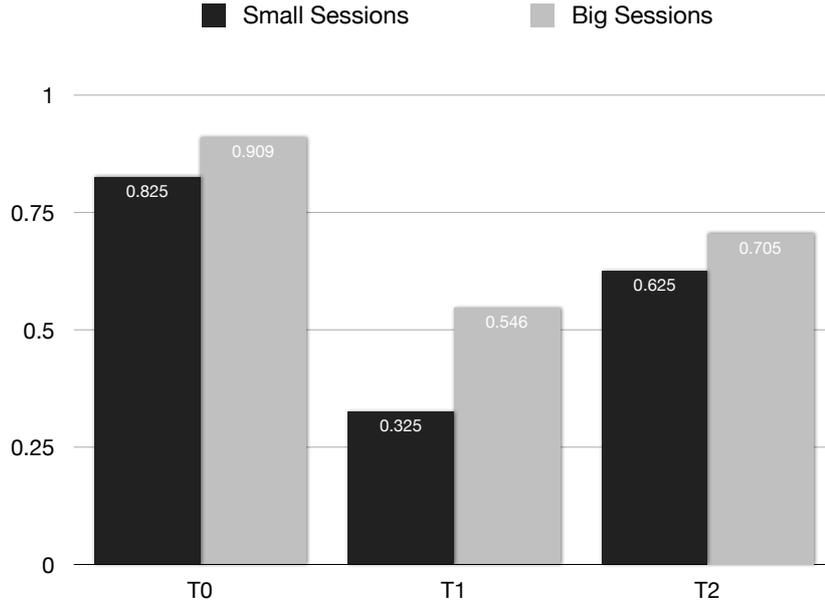
Figure 2: Percentage of population exerting effort in Sessions A & B and in Sessions C & D

to exert effort while the share decreases to 0.625 when subjects are exposed to the scenario characterized by Type II errors (T2). This effect is statistically significant at the 5% level, on the basis of a two-sided null hypothesis and 40 independent paired observations (McNemar's $\chi^2 = 5.33$ , $p - value = 0.022$).

**Session $Big$ (C & D):** In the sessions with higher wages the percentage of subjects exerting high effort drops from 0.909 in T0 to 0.705 in T2. This effect is statistically significant at the 1% level, on the basis of a two-sided null hypothesis and 44 independent paired observations (McNemar's $\chi^2 = 9$ , $p - value = 0.0027$).

## Claim 3. Are Type I and Type II errors equally detrimental to performance?

Claim 3 of the model predicts that neglected rewards to complying agents ($\alpha$) and undeserved rewards to non-complying agents ($\beta$) are identically detrimental for performance. In order to test this hypothesis we compare the share of complying agents in $T1$ and $T2$. We test the following:

$$H_0 : Z_{T1} = Z_{T2} \text{ - vs - } H_1 : \ Z_{T1} \neq Z_{T2}$$

**Session $Small$ (A & B):** Facing a substantial probability of Type I error (T1), only a share of population equal to 0.325 is willing to exert effort, while the share increases to 0.625 when subjects are exposed to a scenario characterized by Type II errors (T2). This effect is statistically significant at the 0% level, on the basis of a two-sided null hypothesis and 40 independent paired observations (McNemar's $\chi^2 = 8$ , $p - value = 0.0047$).

**Session $Big$ (C & D):** In the sessions with higher wages the share of subjects that are willing to exert effort is 0.545 in T1 and 0.705 in T2. This effect is statistically significant at the

|  | T0 | T1 | T2 |
|---|---|---|---|
| Share of population exerting effort in Sessions *Small* (A & B) | 0.825 | 0.325 | 0.625 |
| Share of population exerting effort in Sessions *Big* (C & D) | 0.909 | 0.545 | 0.705 |
|  | T0-T1 | T0-T2 | T2-T1 |
| Share of switchers between treatments in Sessions *Small* (A & B) | 0.5 (***) | 0.2 (**) | 0.3 (***) |
| Share of switchers between treatments in Sessions *Big* (C & D) | 0.364 (***) | 0.205 (***) | 0.159 (**) |

Table 3: Summary of results

1% level, on the basis of a two-sided null hypothesis and 44 independent paired observations (McNemar's $\chi^2 = 7$ , $p - value = 0.0082$). Contrary to the predictions of the standard theory, we provide evidence that Type I and Type II errors do not generate symmetric effects: the detrimental effect of Type I errors is substantially greater than the negative effect entailed by Type II errors.

**Claim 4 Does increasing both the baseline and rewarding wage by the same amount decrease agents' effort provision?**

Claim 4 predicts that fewer people should exert effort, in each treatment, once we move from the Small to the Big sessions. If we focus on T1 and T2, the expected difference in euros between exerting effort and not exerting effort is always the same across all four sessions and equal to €1.32 [8]. However, in the Big sessions subjects receive a guaranteed fixed wage of at least €5.28. This implies that the marginal utility to be extracted from the additional payment of €1.32 should be lower, and this is the reason why we should observe - if anything - a decrease in effort provision once the fixed wage is introduced. An examination of Figure 2 suggests instead that the percentage of population exerting effort in each treatment increases once we move from the small to the big sessions. Statistical tests, however, are less clear cut. We test

$$H_0 : Z_{T0_{small}} = Z_{T0_{big}} \text{ - vs - } H_1 : \ Z_{T0_{small}} \neq Z_{T0_{big}}$$

$$H_0 : Z_{T1_{small}} = Z_{T1_{big}} \text{ - vs - } H_1 : \ Z_{T1_{small}} \neq Z_{T1_{big}}$$

$$H_0 : Z_{T2_{small}} = Z_{T2_{big}} \text{ - vs - } H_1 : \ Z_{T2_{small}} \neq Z_{T2_{big}}$$

and we find that only the difference between $Z_{T1_{low}}$ vs $Z_{T1_{high}}$ is statistically significant at the 5% level, albeit in the wrong direction. Note that Claim 4 predicts that if there exists a difference in effort provision, $Z_{T1_{low}} > Z_{T1_{high}}$

In Figure 2 we observe a qualitative difference in effort provision between Small and Big sessions. A possible behavioral explanation for this result is provided by the theory of gift-exchange. Akerlof's(1982) theory suggests that subjects receiving a share of the compensation pack without any obligation to accomplish the task (in our case the fixed wage in Big sessions) may be induced to reciprocate in kind to the unconditional reward by exerting more effort.

Another explanation for this unexpected finding concern the "legitimacy of control". Schnedler and Vadovic (2011) show that in a principal-agent relation, the "hidden costs of control"(Falk and Kosfeld, 2006) can be relieved to some extent by introducing a fixed wage. Applying their finding to our experimental setting, we can interpret the introduction of the fixed payment in the *Big sessions* as granting the supervisor the right to expose the agent to the erroneous consequences of her control. The agent for his part resents the supervisor less than he does in *Small sessions* for both Type I and Type II errors and thus he is willing to exert more effort.

---

[8]Moreover, treatment T2 in session A&B and treatment T1 in session C&D have the same absolute levels of returns when exerting no effort (€5.28) and when exerting effort (€6.60). T1 and T2 are thus comparable in terms of both relative and absolute returns.

# 5 Discussion

The asymmetry in behavior of subjects under the *mean* (T1) and *indulgent* (T2) treatment is the interesting puzzle of this paper. In the following paragraphs we rule out some potential explanations for the asymmetry and we put forward some tentative interpretations of the results.

## 5.1 Why it cannot be risk aversion

The experiment has been designed in order to rule out potential differences between T1 and T2 in terms of risky choices. To see why, consider the following table of standard generic concave utility functions with separable costs of effort. The subject decides whether to exert high effort whenever the difference in utility (line 3) is positive. Note that the difference in expected utility between T1 and T2 is exactly the same.

|  | $T0$ - Just | $T1$ - Mean | $T2$ - Indulgent |
|---|---|---|---|
| Utility with high effort | $v(w_r) - e_H$ | $\frac{v(w_r)+4v(w_0)}{5} - e_H$ | $v(w_r) - e_H$ |
| Utility with low effort | $v(w_0) - e_L$ | $v(w_0)e_L$ | $\frac{4v(w_r)+v(w_0)}{5} - e_H$ |
| Difference in utility | $v(w_r) - v(w_r) - (e_H - e_L)$ | $\frac{1}{5}v(w_r) - \frac{1}{5}v(w_0) - (e_H - e_L)$ | $\frac{1}{5}v(w_r) - \frac{1}{5}v(w_0) - (e_H - e_L)$ |

Table 4: Risk aversion

On Figure 1 this amounts to confronting the vertical distance between points A-B and C-D which is the same by construction (the two curves have identical concavity defined by $v(w)$ and are distanced by $e_H - e_L$). Whether the attempt to rule out risk aversion by construction can be considered successful depends crucially on the acceptance of the separability of the utility functions in monetary utility and effort (see also Mas-Colell et al., 1995; Bolton and Dewatripont, 2005) and whether we focus on standard risk aversion derived by the decreasing marginal utility of money.

Moreover, in order to control for risk aversion in the data, we have also run an incentivized Holt and Laury (2002) lottery test in Phase I. Correlations between the Holt and Laury measure of risk aversion and the choice of exerting effort both for treatment T1 and T2 is very weak and statistically not significant (respectively Spearman's $\rho - correlation = 0.032$ , $p - value = 0.77$ in T1 and , $p - value = 0.57$ in T2 respectively).

It is however well known that modest-scale risk aversion also can be explained by behavioral biases such as loss aversion and myopic loss aversion (Rabin and Thaler, 2001). We are however skeptical on whether loss aversion can explain our asymmetric result. This is because the net monetary returns of exerting high effort under all three treatments are positive (€6.60 with certainty and €1.32 and €1.32 in expected terms respectively for T0, T1 and T2)

## 5.2 Reciprocity, fairness and inequity

The asymmetry can be modeled by introducing some psychological costs of Type I errors and Type II errors. We first present a model that fits our asymmetry and we then provide interpretations from the literature on management, personnel economics and social psychology.

$w_0-a_H - \alpha\,(w_r-w_0)e^-$

$w_r-a_H - \alpha\,(w_r-w_0)e^-$

$w_r-a_L - \beta\,(w_r-w_0)\,e^+$

$w_0-a_L - \beta\,(w_r-w_0)\,e^+$

high effort

low effort

A

$\alpha$
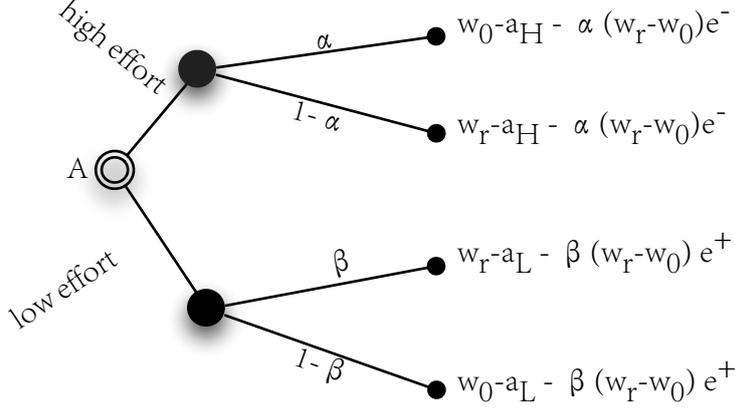
$1-\alpha$

$\beta$

$1-\beta$

Figure 3: Choice of effort with fairness costs

Consider first the subject that exerts high effort. In a *just* scenario with no errors (T0) she would deserve the high wage $v(w_r)$ and would thus get the net reward of $v(w_r)-v(w_0)$. If this is taken as reference, then given the probability of Type I error, $\alpha\,(v(w_r)-v(w_0))$ is the expected value of the reward that is withheld from the subject. Now consider the subject that exerts low effort. In an error-free scenario (T0) she would deserve the low wage $v(w_0)$ with no reward. If this is taken as a reference, then the occurrence of a Type II error generates an amount of undeserved reward equal to $\beta\,(v(w_r)-v(w_0))$ in expected value. There exists thus two types of departure from the reference *just* scenario with no errors. On one hand the agent exerting low effort gets undue rewards and on the other hand the agent exerting high effort does not get due rewards. We extend a model in such a way that both departures represent a cost for the subject although they are weighted differently by the parameters $e^+$ and $e^-$ respectively with $0 \le e^+ \le e^- \le 1$. Figure 3 illustrates the payoff structure of the model extension.

The expected returns of exerting high effort are as in Equation 1 minus the disutility of departing from the reference point. The disutility is proportional to the net reward and to the magnitude of the error so that the expected returns are $\alpha\,[(v(w_0) - e_H] + (1-\alpha)\,[v(w_r) - e_H] - \alpha\,(v(w_r) - v(w_0))\,e^-$. On the other hand the returns of exerting low effort are as in Equation 2 minus the disutility of departing from the reference point. Again, the disutility is proportional to the net reward and to the magnitude of the error so that the expected returns are $\beta\,[(v(w_r) - e_L] + (1 - \beta)\,[v(w_0) - e_L] - \beta\,(v(w_r) - v(w_0))\,e^+$. The new performance condition is defined by the following equation:

$$\bar{\Delta} = \frac{1 + e^- \alpha - e^+ \beta}{1 - \alpha - \beta}(e_H - e_L) \qquad (4)$$

In the following table, Equation 4 is computed with the parameters of our treatments.

| | $T0$ - Just | $T1$ - Mean | $T2$ - Indulgent |
|---|---|---|---|
| $\alpha$ | $0$ | $4/5$ | $0$ |
| $\beta$ | $0$ | $0$ | $4/5$ |
| $\bar{\Delta}$ | $e_H - e_L$ | $(e_H - e_L)(5 + 4e^-)$ | $(e_H - e_L)(5 - 4e^+)$ |

Table 5: Performance condition with fairness under our three treatments

Note that $\bar{\Delta}_{just} > \bar{\Delta}_{indulgent} > \bar{\Delta}_{mean}$ as $e_H - e_L > (e_H - e_L)(5 - 4e^+) > (e_H - e_L)(5 + 4e^-)$. For a given monetary reward $\bar{\Delta}$ the individual is willing to exert a level of effort which is relatively high in T0, low in T2 and lower still in T1. Therefore the performance condition is higher in T0 and lower in T1 and takes an intermediate level in T2. This model is thus compatible with the results we obtain in the lab experiment. How then do we interpret this result?

## 5.3 Equity theory and Inequity Aversion

Equity theory anticipates that not rewarding employees in accordance with their contributions undermines performance (Folger, 2001; Leventhal, 1980). Equity theory (Adams, 1965) distinguishes between negative and positive inequity. Negative inequity happens in situations such as our T1: when individuals exerting high effort are not deservingly rewarded. In our model the costs of negative inequity are $\alpha\,(v(w_r) - v(w_0))\,e^-$, therefore proportional to the magnitude of the Type I error and to the net reward as well.

In T2 all subjects exerting high effort are rewarded. Higher effort provision (as compared to T1) can still be explained by equity theory in terms of positive inequity. Perception of unfairness persists when unfair distributions of outcomes are in favor of the employee. Though employees report being proud of their performance even when their success is the result of cheating, they also tend to feel guilty for their unfair behavior (Krehbiel and Cropanzano, 2000). In the model the costs of positive inequity is represented by $\beta\,(v(w_r) - v(w_0))\,e^+$: they are proportional to the probability of Type II error and to the magnitude of the net reward as well.

There is also a prolific stream of reseach in behavioral economics dealing with inequity aversion, described as a recurring preference for fairness and aversion to distributive inequality. Inequity aversion has been incorporated into several formal models of decision (Rabin, 1993; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). These models all encompass two common features: subjects dislike inequitable outcomes and suffer more from disadvantageous inequality than from advantageous inequality.

Inequity aversion can thus be fitted into our context, although in a looser sense. In fact in our context there is no interaction between agents and the supervisor is an automaton. Thus there cannot be inequality in outcomes arising from the choices of the agent. However, inequity may arise with respect to the reference *just* treatment and thus it resembles more closely an issue of organizational and procedural justice. If the subject measures the equity of the treatment she is subject to against the one she thinks ought to apply under the *just* treatment, then inequity can take the form of both disadvantageous/negative inequity (as in the *mean* treatment) and

advantageous/positive inequity (as in the *indulgent* treatment). Following this literature, it is sound and compatible with our results to assume that $e^+ < e^-$, as favorable although unfair outcomes are preferred (less costly) to unfavorable and unfair outcomes. However, note that the result $\bar{\Delta}_{just} > \bar{\Delta}_{indulgent} > \bar{\Delta}_{mean}$ holds also when $e^+ = e^-$ .

## 5.4   Gift-exchange and Reciprocity

Another stream of literature that is of use in interpreting the result is Akerlof's theory (1982) of gift-exchange. According to this theory the employer offers a gift to the employee in the form of an above-market-equilibrium salary in return for the worker's gift of high non-observable effort for the firm. Within the gift-exchange framework, our *indulgent* treatment (T2) represents a positive departure from the *just* reference (T0) and therefore it could induce the individual to exert more effort (even above the optimal level as defined by Equation 3). On the contrary, our *mean* treatment (T1) is a negative departure from the *just* reference (T0) that may induce him to exert no effort at all (even below the optimal level as defined by Equation 3). A broader perspective on gift-exchange is given also by the reciprocity literature (Bruni et al., 2008; Falk and Fischbacher, 2006; Fehr and Schmidt, 2003). Reciprocating individuals typically respond in kind to others' actions: they are *mean* with the mean and *indulgent* with the indulgent. The model above can also fit the reciprocity approach. In fact, the larger is the departure from the just reference, the larger is the willingness for the subject to sustain costs in order to either exchange gifts with the employer in the case of indulgent behavior $(\beta\left(v(w_r) - v(w_0)\right)e^+)$ or negatively reciprocate in the case of mean behavior $\alpha\left(v(w_r) - v(w_0)\right)e^-$.

Our results seem to fit such behavior, as the propensity to exert high effort is ostensibly higher in the *indulgent* treatment than in the *mean* one. The trouble in interpreting our result in terms of reciprocity and gift-exchange comes from the fact that in our experiment there is no real supervisor (our supervisor is represented by a passive automaton) to be indulgent or mean with. A well-established result in the literature is that, facing an automaton instead of a human subject playing as a proposer in an ultimatum game, respondents are less apt to reject unfair offers (Blount, 1995). There are two common and complementary interpretations of this result: a) a low proposal offered by an automaton cannot be associated with the intention of an actual proposer that it is possible to harm by refusing the offer (negative reciprocity) and b) an unbiased random draw is perceived as a fair procedure. In our experimental setting the supervisor is clearly an automaton and the choice of the applied treatment is random. Therefore subjects should have roughly the same inclination to exert high effort under both the unjust mean and the unjust indulgent treatment. In other words, subjects do not have a real supervisor with which to negatively reciprocate in the case of mean treatment or with which to exchange gifts in the case of indulgent treatment. Even the alternate explanation that subjects are actually positively (negatively) reciprocating with the experimenter in the indulgent (mean) treatment seems to be weak. In fact it should be noticed that the effort task is clearly purposeless and therefore subjects can anticipate that the experimenter does not gain anything from having more effort exerted. Moreover they might also think that the

experimenter is hurt by having to pay more. Therefore a good way to reciprocate positively (negatively) would be to exert low (high) effort in T2 (T1) so that the experimenter pays on average lower (higher) payoffs.

## 5.5  Organizational justice and procedural fairness

A final thought goes to the literature on procedural fairness, which has been addressed both in management and economics. In management, organizational justice is a further elaboration of equity theories that focuses not only on distributive but also on procedural and interactional justice (Cropanzano et al., 2007).

In behavioral economics, too, procedural justice has gained traction in recent years. The procedural fairness literature claims that subjects do not primarily focus on the "ex-post allocative fairness" (Kagel and Roth, 1995) between the mean/indulgent supervisor and the deserving/undeserving subject but rather to the "ex-ante procedural fairness" (Bolton et al., 2005) of the evaluation process. The question of what plays the main role concerns the process used to implement a decision rather than the actual outcome of the decision *per se*. In this respect, our T1 and T2 present an unjust compensation procedure for rewarding agents, although in T1 the subject suffers the mean treatment while she enjoys the indulgent treatment in T2.

# 6  Implications and conclusion

The experiment finds strong support for the existence of an asymmetric impact of errors on agents' willingness to exert high effort. In particular an agent exposed to evaluation errors is more sensitive to Type I errors. This result is not predicted by the model even when considering risk aversion within the expected utility framework. Further work is needed to explain the result, which seems to be robust against some preliminary treatments manipulations. From a theoretical perspective, the experiment sheds new light on the relation between reward systems and motivation that should inform agency theory, organizational behavior and personnel economics. From an organizational perspective, our result expands the notion of organizational justice: Departures from the just treatment can be both advantageous and disadvantageous even in absence of distributional implications with third parties as assumed by equity theories; subjects react differently when they suffer (enjoy) disadvantageous (advantageous) injustice. Our results can also be interpreted as further evidence of gift-exchange behavior in the indulgent treatment and negative reciprocity in the mean treatment.

Although the experimental method has limited external validity, this particular result may have direct practical implications in real-world contexts. Since intangibles are increasingly important in business organizations and knowledge-intensive jobs are difficult to assess, errors in evaluating employees' performance may well be a relevant phenomenon. Our research seems to suggest that, when a perfect assessment of employees' performance is not viable, it may be wise for the supervisor to be cautious when neglecting rewards and - in general - have a

pro-employee bias in conducting her assessment, as this may well be beneficial for employees' motivation and effort provision in the longer term.

# Appendix

# 7 Instructions

We report here the instructions used for the *big* treatments with baseline wage = €5.28 and the rewarding wage = €11.88. In *small* treatments instructions differ only in that the baseline wage = 0 and the rewarding wage = €6.60.

## $T0_{Big}$ Sessions C&D

### *** Situation A ***

In Situation A you will receive a guaranteed fixed payment of €<5.28>

Your task (to count <goal> tables in 40 minutes) is supervised by an automatic supervisor

In this situation, the automatic supervisor does not commit any error of observation

- If you accomplish the task (that is to correctly count the number of "1" in at least <goal> tables), the supervisor will certainly (probability 100%) commit no evaluation error and it will assign to you the payment of €<11.88>

- If instead you do not accomplish the task (that is to correctly count the number of "1" in at least <goal> tables) the supervisor will certainly (probability 100%) commit no evaluation error and it will not assign you the payment of €<11.88>

### # CONTROL QUESTIONS

- In this situation, if you do accomplish the task, you will receive the €<11.88> payment with a probability of [ please, provide the answer -_____%- ] (Correct answer is 100)

- In this situation, if you do not accomplish the task, you will receive zero payment ii) with a probability of [ please, provide the answer -_____%- ] (Correct answer is 100)

- In this situation, you will receive a fixed payment Y of €<5.28> with a probability of [ please, provide the answer -_____% - ] (Correct answer is 100)

### # PARTICIPATION/EFFORT CHOICE

If Phase III of the experiment corresponds to SITUATION A as just described, will you perform the task or will you skip the task?

Remember that

- If you press the "I will perform the task" button and Phase III corresponds to SITUATION A, you will have to wait 40 minutes anyway before proceeding to the questionnaire phase

- If you press the "I will skip the task" button and Phase III corresponds to SITUATION A, you will proceed directly to the questionnaire phase

[ "I will perform the task" ] [ "I will skip the task" ]

**\*\*\* Situation B \*\*\***

In Situation B you will receive a guaranteed fixed payment of €<5.28>

Your task (to count <goal> tables in 40 minutes) is supervised by an automatic supervisor

In this situation, the automatic supervisor might commit an error of observation

- If you accomplish the task (that is to correctly count the number of "1" in at least <goal> tables),

    – the supervisor with a probability of 80% will commit an evaluation error and it will not assign to you the payment of €<11.88>

    – the supervisor with a probability of 20% will commit no evaluation error and it will assign to you the payment of €<11.88>

- If instead you do not accomplish the task (that is to correctly count the number of "1" in at least <goal> tables), the supervisor will certainly (probability 100%) commit no evaluation error and it will not assign to you the payment of €<11.88>

# **CONTROL QUESTIONS**

- In this situation, if you do accomplish the task, you will receive the €<11.88> payment with a probability of [ please, provide the answer -_ _ _ _ _%- ] (Correct answer is 20)

- In this situation, if you do not accomplish the task, you will receive the €<5.28> payment ii) with a probability of [ please, provide the answer -_ _ _ _ _%- ] (Correct answer is 100)

- In this situation, you will receive a fixed payment of €<5.28> with a probability of [ please, provide the answer -_ _ _ _% - ] (Correct answer is 100)

# **PARTICIPATION/EFFORT CHOICE**   If Phase III of the experiment corresponds to SITUATION B as just described, will you perform the task or will you skip the task?

Remember that

- If you press the "I will perform the task" button and Phase III corresponds to SITUATION B, you will have to wait 40 minutes anyway before proceeding to the questionnaire phase

- If you press the "I will skip the task" button and Phase III corresponds to SITUATION B, you will proceed directly to the questionnaire phase

[ "I will perform the task" ] [ "I will skip the task" ]

**\*\*\* Situation C \*\*\***

In Situation C you will receive a guaranteed fixed payment of €<5.28>

Your task (to count <goal> tables in 40 minutes) is supervised by an automatic supervisor

In this situation, the automatic supervisor might commit an error of observation

- If you accomplish the task (that is to correctly count the number of "1" in at least <goal> tables), the supervisor will certainly (probability 100%) commit no evaluation error and it will assign to you the payment of €<11.88>

- If instead you do not accomplish the task (that is to correctly count the number of "1" in at least <goal> tables),

  - the supervisor with a probability of 80% will commit an evaluation error and it will assign to you the payment of €<11.88> €

  - the supervisor with a probability of 20% will commit no evaluation error and it will not assign to you the payment of €<11.88>

# # CONTROL QUESTIONS

- In this situation, if you do accomplish the task, you will receive the €<11.88> payment with a probability of [ please, provide the answer -_____%- ] (Correct answer is 100)

- In this situation, if you do not accomplish the task, you will receive the €<5.28> payment ii) with a probability of [ please, provide the answer -_____%- ] (Correct answer is 20)

- In this situation, you will receive a fixed payment of €<5.28> with a probability of [ please, provide the answer -_____% - ] (Correct answer is 100)

# # PARTICIPATION/EFFORT CHOICE    If Phase III of the experiment corresponds to SITUATION C as just described, will you perform the task or will you skip the task?

Remember that

- If you press the "I will perform the task" button and Phase III corresponds to SITUATION C, you will have to wait 40 minutes anyway before proceeding to the questionnaire phase

- If you press the "I will skip the task" button and Phase III corresponds to SITUATION C, you will proceed directly to the questionnaire phase

[ "I will perform the task" ] [ "I will skip the task" ]
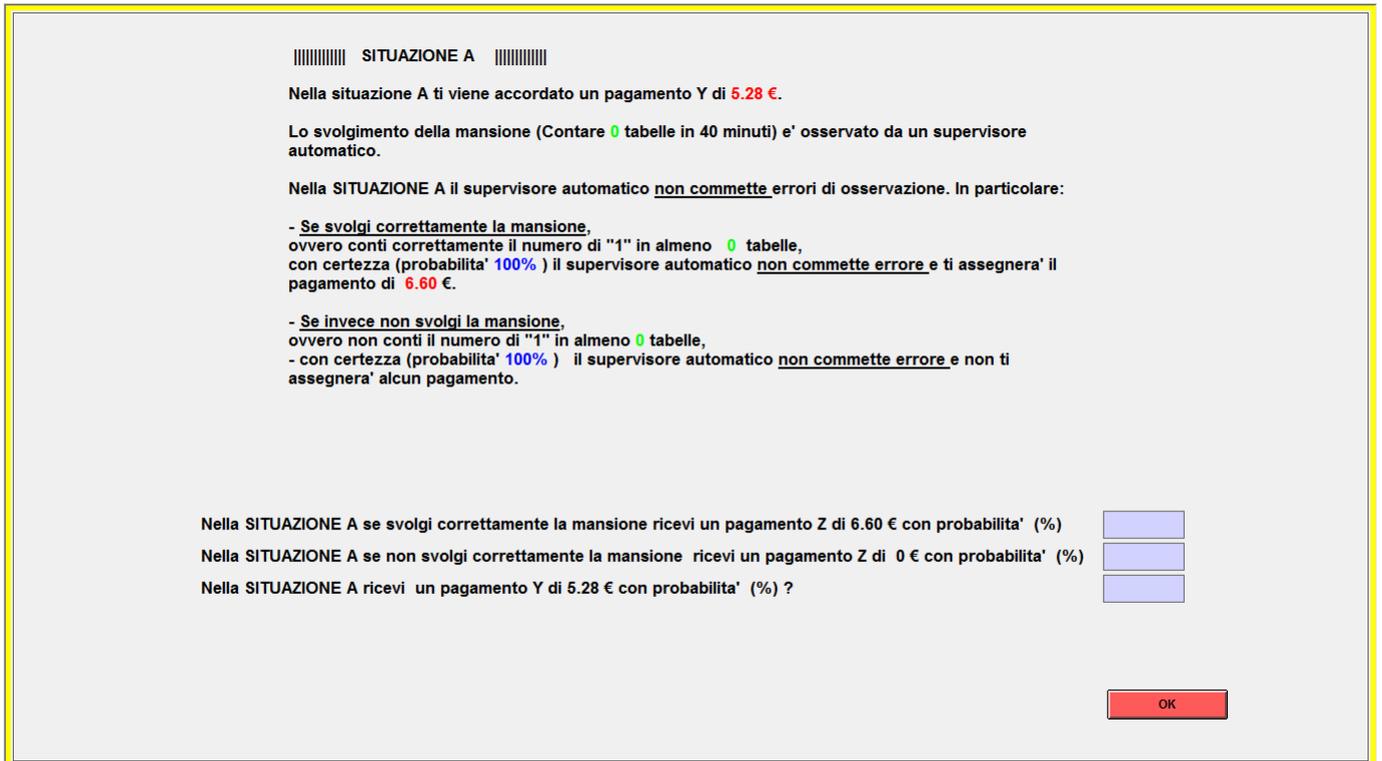
Screenshots

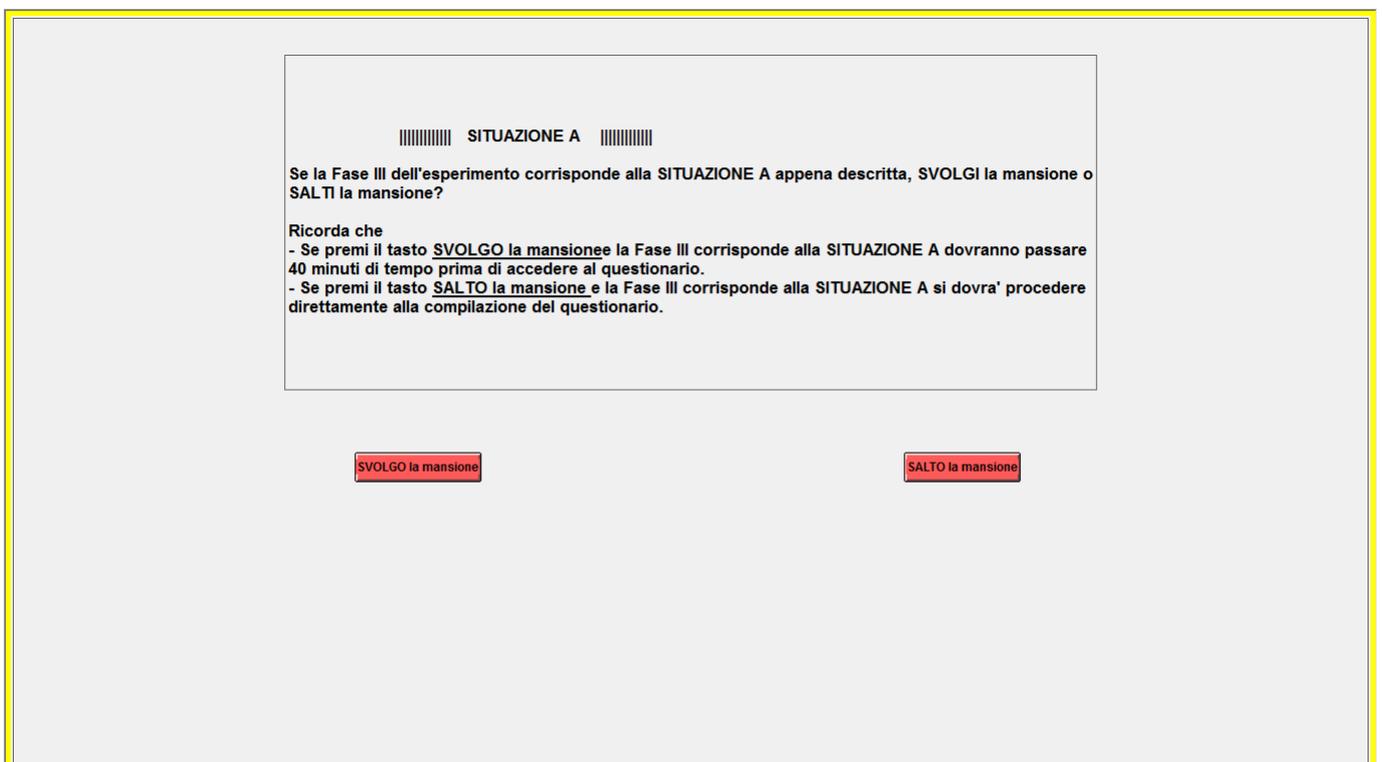Figure 4: Screenshot of the SITUATION presentation and Control Questions



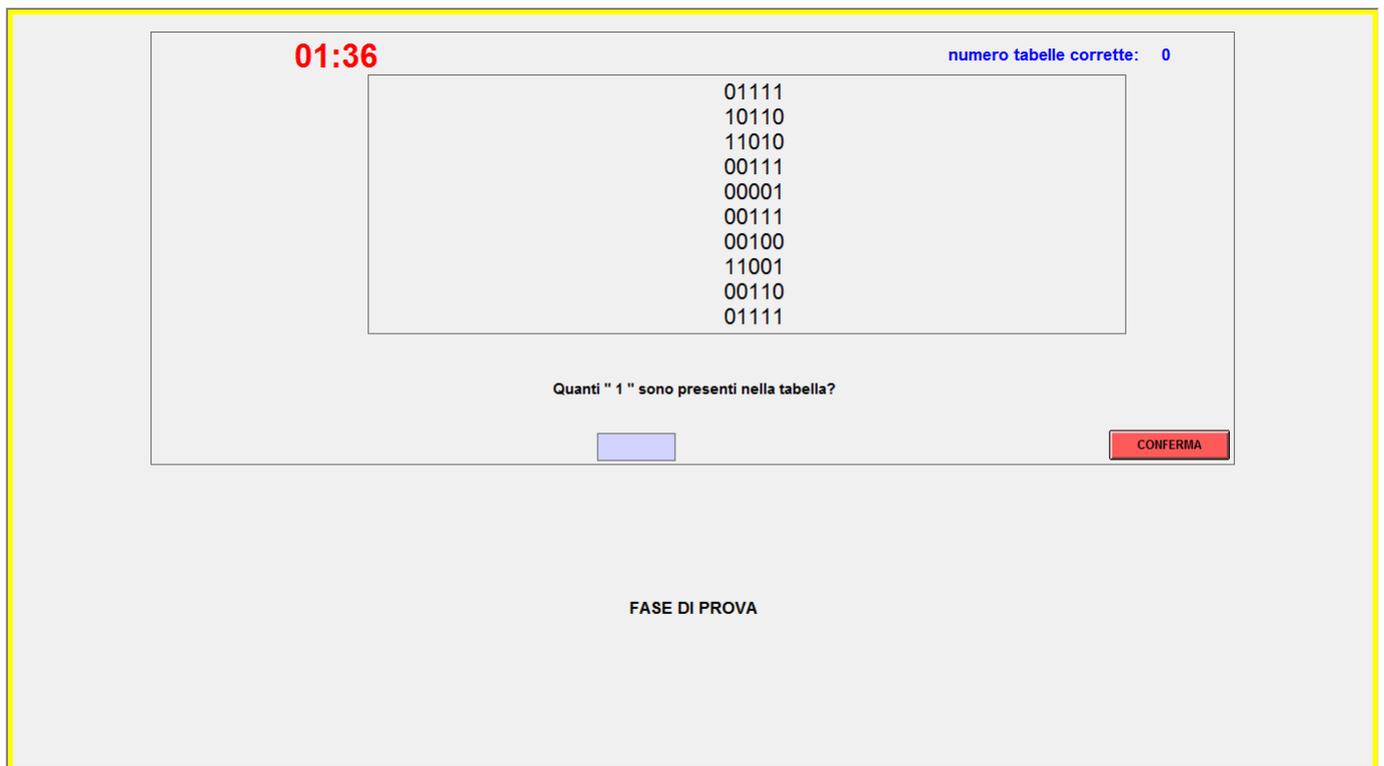Figure 5: Screenshot of the Participation/Effort Choice

Figure 6: Screenshot of the Effort task

# References

**Abeler, J., A. Falk, L. Goette, and D. Huffman**, "Reference Points and Effort Provision," *American Economic Review*, 2011, *101* (2), 470–492.

**Adams, J.S.**, "Inequity in Social Exchange.," *Advances in Experimental Social Psychology*, 1965, *2*, 267–299.

**Akerlof, G.**, "Labor Contracts as Partial Gift Exchange," *The Quarterly Journal of Economics*, 1982, *97* (4), 543–569.

**Akerlof, G.A. and R.E. Kranton**, "Identity, Supervision, and Work Groups," *American Economic Review*, 2008, *98* (2), 212–217.

**Aron, D. and P. Olivella**, "Bonus and Penalty Schemes as Equilibrium Incentive Devices, with Application to Manufacturing Systems," *Journal of Law, Economics, & Organization*, 1994, *10* (1), pp. 1–34.

**Baker, G.P.**, "Incentive contracts and performance measurement," *Journal of Political Economy*, 1992, *100* (3), 598–614.

**Blount, S.**, "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences," *Organizational Behavior and Human Decision Processes*, August 1995, *63* (2), 131–144.

**Bolton, G. and A. Ockenfels**, "ERC: A Theory of Equity, Reciprocity, and Competition," *The American Economic Review*, 2000, *90* (1), 166–193.

_ , **J. Brandts, and A. Ockenfels**, "Fair Procedures: Evidence from Games Involving Lotteries," *The Economic Journal*, 2005, *115* (506), 1054–1076.

**Bolton, P. and M. Dewatripont**, *Contract Theory*, The MIT Press, 2005.

**Bruni, L., M. Gilli, and V. Pelligra**, "Reciprocity: theory and facts," *International Review of Economics*, 2008, *55* (1), 1–11.

**Caplan, A.J., D. Aadland, and A. Macharia**, "Estimating Hypothetical Bias in Economically Emergent Africa: A Generic Public Good Experiment," *Agricultural and Resource Economics Review*, 2010, *39* (2), 344–358.

**Charness, G.**, "Responsibility and effort in an experimental labor market," *Journal of Economic Behavior and Organization*, 2000, *42* (3), 375–384.

_ **and P. Kuhn**, "Lab labor: What can labor economists learn from the lab?," *NBER Working Paper*, 2010.

_ , **R. Cobo-Reyes, N. Jiménez, J.A. Lacomba, and F. Lagos**, "The Hidden Advantages of Delegation: Pareto-improvements in a Gift-exchange Game," *American Economic Review, forthcoming*, 2011.

**Cleveland, J.N. and K.R. Murphy**, "Analyzing performance appraisal as goal-directed behavior," *Research in personnel and human resources management*, 1992, *10*, 121–185.

**Cropanzano, R., D.E. Bowen, and S.W. Gilliland**, "The Management of Organizational Justice," *The Academy of Management Perspectives*, 2007, *21* (4), 34–48.

**Dickinson, D. and M.C. Villeval**, "Does monitoring decrease work effort?:: The complementarity between agency and crowding-out theories," *Games and Economic Behavior*, 2008, *63* (1), 56–76.

**Enderer, F. and G. Manso**, "Is Pay-For-Performance Detrimental to Innovation?," *UC Berkeley: Department of Economics, UCB*, 2009.

**Falk, A. and E. Fehr**, "Why labour market experiments?," *Labour Economics*, 2003, *10* (4), 399 – 406. Special Issue on Labour Market Experiments.

_ **and J. J. Heckman**, "Lab Experiments Are a Major Source of Knowledge in the Social Sciences," *Science*, 2009, *326* (5952), 535–538.

_ **and M. Kosfeld**, "The Hidden Costs of Control," *The American Economic Review*, 2006, *96* (5), 1611–1630.

___ **and S. Gächter**, "Experimental Labour Economics," *The New Palgrave Dictionary of Economics, Basingstoke: Palgrave Macmillan*, 2008.

___ **and U. Fischbacher**, "A theory of reciprocity," *Games and Economic Behavior*, 2006, *54* (2), 293–315.

**Fehr, E., A. Falk, and U. Fischbacher**, "On the Nature of Fair Behavior," *Economic Inquiry*, 2003, *41* (7).

___ **, A. Klein, and K.M. Schmidt**, "Fairness and contract design," *Econometrica*, 2007, *75* (1), 121–154.

___ **and K.M. Schmidt**, "A theory of Fairness, Competition, and Cooperation," *Quarterly journal of Economics*, 1999, *114* (3), 817–868.

___ **and ___** , "Theories of fairness and reciprocity: Evidence and economic applications," in "Advances in economics and econometrics: Theory and applications, Eighth World Congress," Vol. 1 2003, pp. 208–257.

**Festinger, L.**, "A theory of cognitive dissonance.," 1957.

**Fischbacher, U.**, "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 2007, *10* (2), 171–178.

**Folger, R.**, "Fairness as deonance," *Theoretical and cultural perspectives on organizational justice*, 2001, pp. 3–33.

**Gibbons, R.**, "Incentives in organizations," *The Journal of Economic Perspectives*, 1998, *12* (4), 115–132.

**Gneezy, U. and J. List**, "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments," *Econometrica*, 2006, *74* (5), 1365–1384.

**Greenberg, J.**, "A taxonomy of organizational justice theories," *Academy of Management review*, 1987, *12* (1), 9–22.

**Greiner, B.**, "The online recruitment system ORSEE 2.0-A guide for the organization of experiments in economics," *University of Cologne, Working paper series in economics*, 2004, *10.*

**Higgins, C.A., T.A. Judge, and G.R. Ferris**, "Influence tactics and work outcomes: a meta-analysis," *Journal of Organizational Behavior*, 2003, *24* (1), 89–106.

**Hölmstrom, Bengt**, "Moral Hazard and Observability," *The Bell Journal of Economics*, 1979, *10* (1), pp. 74–91.

**Holt, C.A. and S.K. Laury**, "Risk Aversion and Incentive Effects," *The American Economic Review*, 2002, *92* (5), 1644–1655.

**Huseman, R.C., J.D. Hatfield, and E.W. Miles**, "A new perspective on equity theory: The equity sensitivity construct," *Academy of Management Review*, 1987, *12* (2), 222–234.

**Irlenbusch, Bernd and Dirk Sliwka**, "Transparency and reciprocal behavior in employment relations," *Journal of Economic Behavior & Organization*, 2005, *56* (3), 383 – 403.

**Judge, T.A. and G.R. Ferris**, "Social context of performance evaluation decisions," *Academy of Management Journal*, 1993, *36* (1), 80–105.

**Kagel, J. and A. Roth**, *The Handbook of Experimental Economics*, Princeton University Press, 1995.

**Konow, James**, "Which Is the Fairest One of All? A Positive Analysis of Justice Theories," *Journal of Economic Literature*, 2003, *41*, 1188–1239.

**Krehbiel, P.J. and R. Cropanzano**, "Procedural justice, outcome favorability and emotion," *Social justice research*, 2000, *13* (4), 339–360.

**Lazear, E.P.**, "Personnel economics: past lessons and future directions," 1999.

_ , "The future of personnel economics," *The Economic Journal*, 2000, *110* (467), 611–639.

**Leventhal, G.S.**, "What Should Be Done with Equity Theory? New Approaches to the Study of Fairness in Social Relationships.," in K. Gergen, J. Greenberg, and R. Willis, eds., *Social exchanges: Advances in theory and research*, Plenum Press New York 1980, pp. 27–55.

**Mas-Colell, A., M.D. Whinston, and J.R. Green**, *Microeconomic Theory*, Vol. 981, Oxford Univiversity Press, 1995.

**Mitchell, T.R.**, "Expectancy-value models in organizational psychology," *Expectations and actions: Expectancy-value models in psychology*, 1982, pp. 293–312.

**Polinsky, A.M. and S. Shavell**, "The theory of public enforcement of law," *Handbook of law and economics*, 2007, *1*, 403–454.

**Prendergast, C.**, "The Provision of Incentives in Firms," *Journal of Economic Literature*, 1999, *37* (1), pp. 7–63.

**Pritchard, R.D.**, "Equity theory: A review and critique," *Organizational Behavior and Human Performance*, 1969, *4* (2), 176–211.

**Rabin, M.**, "Incorporating fairness into game theory and economics," *The American Economic Review*, 1993, pp. 1281–1302.

**Rabin, Matthew and Richard H. Thaler**, "Risk Aversion," *Journal of Economic Perspectives*, 2001, *15* (1), 219–232.

**Rizzolli, M. and L. Stanca**, "Judicial errors and deterrence: theory and experimental evidence," *Journal of Law and Economics*, Forth., *Forth.*

**Schnedler, Wendelin and Radovan Vadovic**, "Legitimacy of Control," *Journal of Economics & Management Strategy*, 2011, *20* (4), 985–1009.

**Steers, R.M., R.T. Mowday, and D.L. Shapiro**, "Introduction to special topic forum: The future of work motivation theory," *The Academy of Management Review*, 2004, *29* (3), 379–387.

**Viswesvaran, C. and D.S. Ones**, "Examining the construct of organizational justice: A meta-analytic evaluation of relations with work attitudes and behaviors," *Journal of Business Ethics*, 2002, *38* (3), 193–203.

**Vroom, V.H.**, "Work and motivation.," 1964.