

NBER WORKING PAPER SERIES

WORKER EMPOWERMENT AND SUBJECTIVE EVALUATION:  
ON BUILDING AN EFFECTIVE CONFLICT CULTURE

W. Bentley MacLeod  
Victoria Valle Lara  
Christian Zehnder

Working Paper 27857  
<http://www.nber.org/papers/w27857>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
September 2020

We very much appreciate comments from Björn Bartling, Ernst Fehr, Maria Guadalupe, Michael Kosfeld, Ola Kvaloy, Jean-Robert Tyran, Giorgio Zanarone and seminar participants at Workshop on Behavioral Public Economics, Vienna, Workshop on Organizational Economics, University of Konstanz, Brownbag Seminar, University of Lausanne, Workshop on Incentives and Motivation, Stavanger, SIOE, Stockholm, Organizational Economics Workshop, Berkeley, NBER Organizational Economics Workshop, and Relational Contracts Workshop, Moscow. We also gratefully acknowledge financial support from the Swiss National Science Foundation (Grant No. 100018\_152903i). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by W. Bentley MacLeod, Victoria Valle Lara, and Christian Zehnder. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Worker Empowerment and Subjective Evaluation: On Building an Effective Conflict Culture  
W. Bentley MacLeod, Victoria Valle Lara, and Christian Zehnder  
NBER Working Paper No. 27857  
September 2020  
JEL No. D02,D03,J33,J41,M5,M52

### **ABSTRACT**

Although conflicts typically lead to a waste of resources, organizations may still benefit from a corporate culture that tolerates or even encourages conflicts. The reason is that coordinated conflicts may help to enforce informal contracts and foster cooperation. In this paper we report results of a series of laboratory experiments designed to explore whether and under what conditions an efficiency-enhancing conflict culture can emerge. Using a principal-worker setup with subjective performance evaluation, we show that establishing a functional conflict culture is a delicate matter. If conflicts are encouraged in a careless, hands-off manner, the destructive side of conflicts is likely to dominate. To be successful a conflict culture requires a careful management of fairness norms. In our experiment we find that conflicts have positive net effects on efficiency only if an explicit code of conduct is established and conflicts are institutionalized through a grievance process. Thus, providing workers with more power may be a necessary but not sufficient condition for improving productivity when performance evaluations are subjective.

W. Bentley MacLeod  
Department of Economics  
Columbia University  
420 West 118th Street, MC 3308  
New York, NY 10027  
and NBER  
wbmacleod@wbmacleod.net

Christian Zehnder  
Faculty of Business and Economics  
University of Lausanne  
Quartier UNIL-Chamberonne  
Internef 612  
CH-1015 Lausanne  
Switzerland  
christian.zehnder@unil.ch

Victoria Valle Lara  
UNIL-Chamberonne  
Internef  
CH-1015 Lausanne  
Lausanne, not CH-1015  
Switzerland  
vikivallelara@gmail.com

*“A good manager doesn’t try to eliminate conflict; he tries to keep it from wasting the energies of his people. If you’re the boss and your people fight you openly when they think that you are wrong—that’s healthy.”*

*Robert Townsend (2007)*

## 1. INTRODUCTION

A corporate culture that tolerates or even encourages conflicts in case of perceived injustice entails both opportunities and dangers. On one hand, credible threats of conflict may motivate otherwise non-complying parties to respect social norms. On the other hand, an escalation of conflict may lead to waste of resources. In this paper we study conflicts in the context of employment relations with subjective performance evaluation. We report evidence from a series of laboratory experiments illustrating that the formation of a healthy conflict culture is a delicate matter. If conflicts are managed in a careless, hands-off manner, their destructive side is likely to dominate. Productive conflicts require a coordination of norms. In our experiment conflicts have positive net effects only if an explicit code of conduct is established and conflicts are institutionalized through a grievance process.

Ample evidence shows that perceived unfairness is an important trigger of conflicts. Many people engage in costly retaliation in response to violations of implicit agreements that are anchored by norms. Famous examples of such negative reactions to norm violations include rejections of small offers in ultimatum games (Güth et al., 1982; Kahneman et al., 1986), punishment of free-riders in public goods experiments (Gächter and Fehr, 2000; Ostrom et al., 1992), or third-party interventions in distribution games (Leibbrandt and López-Pérez, 2012; Gächter and Fehr, 2000). Although such confrontations are destructive and inefficient in the short term (Dreber et al., 2008; Egas and Riedl, 2008), they can be ultimately efficiency-enhancing because—once established—the threat of retaliation may induce norm breakers to comply and cooperate (Gächter et al., 2008).

Existing studies reporting an efficiency-enhancing effect of conflicts have predominantly focused on deterministic set-ups with symmetric information (see, Chaudhuri, 2011, for a survey).<sup>1</sup> These findings may therefore not generalize to more complex settings like employment relationships, which are characterized by hidden actions, stochastic input-output links, and noisy information. Probabilistic outcomes and information asymmetries may considerably complicate the role of conflicts because misaligned views on what constitutes fair behavior may trigger lots of destructive misunderstandings.

---

<sup>1</sup>Grechenig et al. (2010) and Ambrus and Greiner (2012) show that peer-punishment technologies are less beneficial in public good games in which players have imperfect information about contributions of others.

Our experiments build on a principal-agent game with subjective performance evaluation. The Worker’s effort is private information and creates a stochastic output that neither the Principal nor to the Worker can observe. Instead, the Principal and the Worker each get a private, subjective, and non-contractible output signal. These signals are noisy and imperfectly correlated so that the Principal and the Worker can have contradicting information. After observing the subjective signal, principals decide whether or not to pay a bonus. The Worker then can decide to retaliate. To study the impact of conflicts on efficiency, we compare two treatments: a no conflict treatment that excludes conflict by design and a conflict treatment in which workers can engage in costly conflict behavior. We implement conflicts in the form of costly punishment to reflect the view that the immediate consequence of conflict is typically negative for all involved parties, but mostly so for the party under attack.

Under the self-interest assumption, subgame-perfection implies low effort in both treatments: conflict and no conflict. And neither bonus payments nor costly conflicts ever occur on the equilibrium path. Efficiency, in contrast, requires that the Worker exerts high effort. Intuitively, the presence of fairness considerations might improve matters in this context. Even in the absence of conflicts, a fair-minded Principal who believes that her Worker provided high effort can use the bonus payment to return the favor (gift-exchange, see Akerlof, 1980). Moreover, if conflicts are present, a fair-minded Worker can discipline the Principal through the threat of punishment so that even a selfish Principal may have an incentive pay the bonus (Fehr and Fischbacher, 2004, 2003). However, a complicating issue is that it is not obvious what fair compensation entails in our set-up.

Previous work on bargaining games shows that perceived fairness is both context specific and shaped by strategic considerations (Prasnikar and Roth, 1992; Binmore et al., 1993) and these effects are even more pronounced under uncertainty (Kagel et al., 1996). In our environment uncertainty manifests itself in the stochastic link between the Worker’s input and output and may imply that multiple fairness norms co-exist.<sup>2</sup> There are two norms with a particular intuitive salience in our context. The first fairness norm requires that agents are compensated for exerting high effort (the pay-for-input norm). The second norm, in contrast, requires that workers should be rewarded only if output turns out to be high (the pay-for-output norm).<sup>3</sup> For a conflict culture to function it is crucial that the Worker and the Principal agree on the fairness norm upon which their implicit agreement builds. If trading partners (unknowingly) disagree about the definition of norm-abiding behavior,

---

<sup>2</sup>MacLeod and Malcomson (1998) discuss the existence of multiple equilibrium norms in labor markets within the context of self-enforcing relational contracts.

<sup>3</sup>The labels we use for fairness norms refer to Lazear’s (1986) categorization of compensations schemes. However, whereas Lazear’s work aims at identifying determinants of different compensation forms, our interest is in exploring the efficiency consequences of norm enforcement in employment relationships.

misunderstandings will be interpreted as breach of agreement and may lead to wasteful conflicts. Moreover, for efficiency it not only matters that parties coordinate on a norm, but it is also important which particular norm they coordinate on.

The pay-for-input (PI) norm prescribes that the Worker provides high effort and the Principal reciprocates with bonus pay. Workers who enforce the PI norm ignore their subjective output signal and view any non-payment of the bonus as breach of agreement. If there are sufficiently many norm enforcers, the threat of conflict will induce principals to pay the bonus irrespective of their own subjective performance signal. The advantage of implicit agreements that build on the PI norm is that cooperation does not require any conflicts on the equilibrium path. Principals are disciplined by the threat of conflict alone. The downside is that the incentive structure established by the PI norm fails to motivate selfish agents to exert high effort. In fact, unconditional bonus payments invite selfish agents to invade the system and to free-ride on the norm enforcement of others. If there is a large fraction of self-interested Workers, the PI norm will not lead to a large increase in performance.<sup>4</sup>

The pay-for-output (PO) norm prescribes that the Worker exerts high effort, but feels entitled to a bonus payment if and only if his own subjective signal suggests that output is high. Likewise, the Principal pays a bonus if, and only if, he observes a high subjective performance signal. The upside of the PO norm is that it deters the shirking of self-interested agents by connecting performance to pay. If a sufficiently large population of agents enforce the PO norm, principals have an incentive to make signal-contingent bonus payments and all agents (including those who are selfish) have an incentive to provide high effort. However, since the Worker's signal of performance is imperfectly correlated with the Principal's signal, the two parties may receive conflicting signals. In particular, the Principal might observe a low signal, and not pay the bonus, while the Agent observes a high signal and expects to get a bonus. In this case, the PO norm implies that the Worker will punish the Principal. Thus, there exists conflict on the equilibrium path. However, as long as the probability of contradicting signals is not too high, the damage created by occasional conflicts is largely dominated by the efficiency gains from full effort provision.

Behavioral contract theory typically studies how social preferences shape behavior and contract design (Fehr and Schmidt, 1999; Fehr et al., 2011; Koszegi, 2014). This research takes the population of workers and their preferences as given and tries to adjust contracts and institutions such that performance is high. In this paper, in contrast, we take the view that preferences are partially endogenous in that they may depend on the institutional environment. Specifically, we assume that the details of the context may affect the selection

---

<sup>4</sup>Existing evidence documents strong heterogeneity in the degree to which people care about fairness. Whereas some people make considerable sacrifices to enforce fairness norms, others behavior is best described as approximately selfish (see, e.g., Cooper and Kagel, 2016, for a recent survey of the literature).

of fairness norms that trading parties use as the basis of their implicit agreements. We explore the coordination on fairness norms within the context of two treatment conditions—conflict and no conflict. Thus the first question we ask is whether or not the potential for conflict can enhance organizational performance. Second, we compare these treatments across varying contexts, while holding the relationship between actions and payoffs fixed. In standard behavioral contract models, where preferences are defined exclusively over payoffs and the distribution of payoffs between the two parts is fixed, context should have no effect on behaviour. If it does (and we find that they does), then this is evidence that context (e.g. in the form of “corporate culture”) can affect behavior.

Our experiments aim at exploring whether and under what conditions an efficiency-enhancing conflict culture based on the pay-for-output norm can emerge. In our first experiment (henceforth called “baseline experiment”) participants spontaneously interact in our subjective evaluation setup without any further intervention. Participants play the game for 15 periods with changing partners (random matching protocol). We observe that participants fail to build a functioning conflict culture in the baseline experiment. The conflict treatment increases the frequency of bonus payments and leads to a small, but statistically insignificant increase in effort relative to the no conflict treatment. However, overall, the damage inflicted by the emergence of conflicts is more important than the small increase in productivity, so that total surplus decreases in the presence of conflicts. The underlying problem is that participants fail to coordinate on the pay-for-output norm. About half of the workers in the conflict treatment induce costly sanctions, but many of those norm enforcers exhibit sanction patterns that are more in line with the pay-for-input norm. The multiplicity of norms leads to uncoordinated conflicts with only limited motivating effects on workers.

Our baseline experiment teaches us an interesting lesson: a careless, hands-off implementation of a conflict culture can easily backfire. Uncoordinated conflicts are costly and have only a very limited positive impact on motivation.

At the same time, one might argue that the lack of any communication makes coordination “unnaturally” complicated. In most real-life settings, communication is available and organizations can encourage conversations on conflict to foster the development of broadly accepted fairness norms. Our second experiment (the “communication experiment”) investigates the extent to which communication helps the trading parties to coordinate on an efficiency-enhancing use of conflicts. To this purpose, we added an additional stage at the beginning of every period of the experiment. In this communication stage principals and workers were able to announce their planned strategy through the selection of a message from a pre-defined set of messages.

Our second experiment reveal that communication does not necessarily facilitate the establishment of an efficiency-enhancing conflict culture. As in the baseline experiment, average effort in the conflict treatment is only slightly higher than in the no conflict treatment and overall surplus decreases in the presence of conflicts. All in all, it seems that communication further reinforces rather than mitigates the problems observed in the baseline experiment. In particular, communication leads to an even stronger use of the pay-for-input norm than in the baseline experiment and the overall negative impact of conflicts on efficiency remains.

The communication experiment confirms that self-coordination on an efficiency-enhancing fairness norm is difficult in our setting. As a next step we study the impact of a direct external appeal to coordinate on the pay-for-output norm. Our third experiment (“the agreement experiment”) explores the effectiveness of a code of conduct in our set-up. We begin each period with a stage in which we ask the Principal and the Worker to electronically sign a non-binding agreement in which they confirm their intention to follow the pay-for-output norm. Failure to sign the agreement by at least one party implies that trade does not occur and an unattractive outside option is implemented. Surprisingly, this intervention has only very weak effects. Despite the fact that the vast majority of participants sign the agreement, most players do not feel obliged to respect the rules of the code of conduct. As a consequence, the results of the agreement experiment are similar to those of the baseline experiment.

Our fourth and final experiment (“the grievance experiment”) adds a decisive feature to the agreement experiment: institutionalized conflicts. The trading parties sign the same code of conduct as in the agreement experiment. The only difference is that the initiation of conflict requires a formal grievance process. If workers intend to engage in a conflict with their principal they need to file a complaint in which they explicitly confirm that the Principal has violated the code of conduct. It is important to emphasize that in our experiment the institutionalization of the conflict is just a re-framing of the exact same action space that we use in the agreement experiment. In particular, there is no verification process after a worker files a complaint and there is no material consequence to lying.

Nevertheless, we observe that outcomes in the grievance experiment are substantially different from those in our other experiments: the trading parties follow more frequently the pay-for-output norm and conflicts are much more coordinated. Moreover, the conflict patterns also induce the principals to make their bonus payments contingent on their subjective signal so that even selfish workers have a monetary incentive to exert high effort. As a consequence, there is a significant increase in the average effort level and higher total surplus in the conflict treatment relative to the no conflict treatment.

Our results have implications for conflict management in organizations. The management literature recognizes that conflicts are inevitable and many authors recommend tactics and

procedures to handle conflicts constructively (Coleman et al., 2014; De Dreu et al., 2008; Tjosvold et al., 2014). Our findings suggest that even conflicts that appear to be purely wasteful may be ultimately efficiency-enhancing because they establish an implicit incentive system. The challenging part for management is the creation of broadly shared fairness norms on which implicit agreements within the organization are built. We believe that the effective choice of managerial interventions that coordinate people on specific norms is part of what creating a corporate culture means. We share the view of identity economics (Akerlof and Kranton, 2000, 2010) that preferences should be modeled as a function of social context, because it is the context that defines the norms with which people identify (see also Krupka et al., 2017). This perspective also relates to insights in the sociology and law literatures. Already Weber (1947) recognized the importance of norms and legitimacy for modifying individual behavior and argued that parties are more likely to follow “legitimate” norms. The literature on the expressive role of law discusses how rules that are viewed as legitimate affect behavior (see McAdams, 2017).

Our work also extends the literature on norms in two ways. On the theory side, we introduce the notion of a norm equilibrium. This concept allows us to establish the stability of a fairness norm with respect to invasion from players following another norm (including players who follow the rational, self-interest “norm”). As detailed above, there are three natural norms in our experiment: pure self-interest, pay-for-input, and pay-for-output. In the no conflict treatment only the pure self-interest norm is an equilibrium. In the conflict treatment, in contrast, both the pure self-interest norm and the pay-for-output norms are equilibria and pay-for-output is the payoff-dominant equilibrium. This result motivates our quest for contextual conditions that favor coordination on the pay-for-output norm.

On the empirical side, we demonstrate that a feature of our environment is that outcomes can be represented with a binary tree. If subjects are assumed to choose a norm with error, then the probability distribution of the end nodes of the tree can be represented by a multi-nomial logit. Hence adherence to a norm can be estimated using an off-the-shelf, multi-nomial regression. This approach not only allows us to measure whether a particular norm gets better or worse at describing the data as we move across environments, but it also enables us to use a non-nested likelihood ratio test to formally test which norm does the best job of fitting the data within a given treatment. We see this as a contribution to the development of empirical methods for identifying and measuring social norms and cultures.

The remainder of our paper is organized as follows. The next section introduces our framework and defines the norm equilibrium. Section 3 presents the experimental results by environment. In Section 4 we introduce a method to measure the prevalent social norm in a given experiment. Section 5 concludes.



## 2. FRAMEWORK

**The Subjective Evaluation Game.** Our work builds on a simplified version of the subjective evaluation model introduced in MacLeod (Section III). In particular, we consider the following contracting problem: The Principal hires a Worker and offers a contract  $(w, B)$ , where  $w$  is a contractible base wage and  $B$  is a non-enforceable bonus that the Principal can pay at her discretion. The Worker can choose an effort  $e \in \{0, 1\}$  at cost  $c(e) = c \times e$ , where  $c > 0$ . The Worker's effort induces a stochastic return  $r \in \{r_H, r_L\}$ , where  $r_L \geq 0$ ,  $\delta \equiv r_H - r_L > 0$ ,  $Prob(r = r_H|e) = \gamma_e$ , and  $1 > \gamma_1 > \gamma_0 > 0$ . The return is not observable, neither for the Principal, nor for the Worker. However, both parties receive private and subjective performance signals. The Principal's subjective signal  $s_P \in \{0, 1\}$  depends on the return and is determined by the following conditional probabilities:

$$\begin{aligned} Pr \{s_P = 1|r = r_H\} &= p, \\ Pr \{s_P = 1|r = r_L\} &= 1 - p, \end{aligned}$$

where  $p \in (0.5, 1)$ . The Worker's subjective signal  $s_A \in \{0, 1\}$  depends on the Principal's signal and the corresponding conditional probabilities are:

$$\begin{aligned} Pr \{s_A = 1|s_P = 1\} &= q. \\ Pr \{s_A = 1|s_P = 0\} &= 1 - q, \end{aligned}$$

where  $q \in (0.5, 1)$ .

After having observed her signal, the Principal decides whether or not she pays the bonus ( $b \in \{0, B\}$ ). In response to the bonus decision the Worker then decides whether or not to initiate a conflict  $d \in \{0, 1\}$  that reduces the Principal's payoff by an amount  $f(d) = dF$ , where  $F > 0$ . Engaging in a conflict also has a material cost for the Worker  $k(d) = dK$ , where  $0 < K \leq F$ .

Thus, the Principal and the Worker realize the following payoffs:

$$\begin{aligned} \pi_P &= r_L + \gamma_e \delta - w - b - dF, \\ \pi_A &= w + b - c(e) - dK. \end{aligned}$$

The decisive parameters of this game are  $\gamma_1$ ,  $\gamma_0$ ,  $p$ , and  $q$ . These parameters govern the stochastic relation between input (effort) and output (return), the degree to which the Principal's signal is distorted, and the subjectiveness of the players' performance evaluation (i.e., the correlation between the Principal's and the worker's signal). If  $q = 1$  the signals of the Principal and the Worker are always perfectly aligned and performance evaluation is no longer subjective. If  $p = 1$  the Principal has perfect information about the worker's

output (thus, the special case  $p = q = 1$  corresponds to the standard principal-agent model). Finally, if  $\gamma_1 = 1$  and  $\gamma_0 = 0$ , input and output are perfectly correlated (in this case, a simple output-contingent contract implements first-best effort).

To study the impact of conflicts on outcomes, we implement two versions of this game in our experiment. In the no conflict condition, the game ends with the Principal’s bonus choice. In the conflict condition, the full game is played.

**Norm Guided Behavior.** The economic approach to studying behavior in such an environment typically relies on some version of a Nash equilibrium. Each party chooses a strategy that maximizes their own material payoff given correct expectations of the other party’s behavior. Decades of research have shown that individuals do respond to material incentives, much as predicted by the theory. However, the evidence also reveals two important qualifications: i) Individuals respond imperfectly: people make mistakes, do not consider all the options, and process information incompletely (see, e.g., Simon, 1997; Rubinstein, 1998). ii) Material payoffs are not the only driving force behind human behavior: many identify with and respect social norms that are prevalent in their environment (see, e.g., Akerlof and Kranton, 2000, 2010; Cooper and Kagel, 2016).

We believe that there is an interesting link between bounded rationality and norm abidance. A fundamental problem that complicates individuals’ decision making process is that the strategy space they face is huge.<sup>5</sup> Even though our experiment is a simplified and highly stylized version of employment, our setup still allows for a wide variety of behaviors, so that it is not obvious what optimal or “appropriate” behavior is. We argue that norms may mitigate this problem, because they increase the salience of certain strategies and therewith reduce the perceived set of relevant options. As discussed in the introduction, there are two fairness norms that seem natural in our environment. We borrow the terminology introduced by Lazear (1986) and call the norms *pay-for-input* (PO) and *pay-for-output* (PI).

Under the PI norm, the Worker agrees to provide high effort in exchange for generous pay. This norm essentially corresponds to a version of gift-exchange (see Akerlof, 1980; Fehr and Schmidt, 1999; Falk et al., 2003, , for example). When following this norm, the Worker provides high effort in expectation of getting the bonus. If the Principal decides not to pay the bonus, the Worker feels unfairly treated, and will, if given the chance, retaliate by punishing the Principal. Formally, let  $\sigma_W^{PI}$  and  $\sigma_P^{PI}$  be the strategies in the game described above corresponding to adherence to the PI norm for the Worker and the Principal.

The PO norm is closer to what classical agency theory predicts: if effort cannot be observed, then the optimal contract should be one in which the Worker is rewarded for observed performance measures. The twist in our experiment (relative to the standard agency model)

---

<sup>5</sup>This is well known since the seminal work of Simon (1982).

is that performance signals are *subjective* and not observed by the other party. Under the PO norm, the Worker therefore agrees to provide high effort in expectation of a bonus payment for (subjectively perceived) high output. The Principal pays a bonus, if and only if, she observes a high subjective signal herself. If the Worker and the Principal receive the same signal, there will never be a conflict. If both signals are low, the Principal does not pay the bonus, but the Worker accepts it. If both signals are high, the Principal pays the bonus and the Worker takes no action. However, if the Worker's signal is positive, but the Principal's signal is negative, the Worker will not get the bonus although he feels that he deserves it. In this case, the Worker will engage in conflict and punish the Principal. Conflicting signals in the reversed direction are not causing conflicts because in those cases the Worker simply gets a bonus although he feels that he does not necessarily deserve it. Let  $\sigma_W^{PO}$  and  $\sigma_P^{PO}$  denote the strategies corresponding to adherence to the PO norm.

Finally, we also consider the benchmark of purely cynical/self-interested players. In our experiment it is costly to initiate conflict and, hence, a cynical/self-interested Worker would never punish the Principal for being unfair. Given this, the cynical/self-interested Principal would never face consequences for keeping the bonus for herself instead of paying it to the Worker. Thus, the Worker will never be rewarded for high effort and, hence, will be better off shirking. Under standard assumptions, this strategy combination is the unique sub-game perfect Nash equilibrium of this game. It is obvious that the hypothesis that play can be predicted by the sub-game perfect Nash equilibrium concept is soundly rejected by the data, but it serves as a useful benchmark. Let  $\sigma_W^{SI}$  and  $\sigma_P^{SI}$  be the strategies in the game corresponding to *self-interested* (SI) choices.

In terms of organizing the results we propose to think in terms of a "norm" equilibrium, motivated in part by Axelrod (1981) and Akerlof (1980). Instead of assuming that parties can follow *any* strategy, let us suppose that cognitive/behavioral effects imply that parties focus upon one of the three norms (PI, PO or SI). Hence, the strategy space for each player is assumed to be given by:

$$\begin{aligned}\Sigma_P &= \{\sigma_P^{PI}, \sigma_P^{PO}, \sigma_P^{SI}\}, \\ \Sigma_W &= \{\sigma_W^{PI}, \sigma_W^{PO}, \sigma_W^{SI}\}.\end{aligned}$$

Let  $U_P(\sigma_P, \sigma_W)$  and  $U_W(\sigma_P, \sigma_W)$  be the corresponding payoffs for the Principal and the Worker. Given this game, the notion of a Nash equilibrium is well defined in "norm space". The intuition is that a norm is an equilibrium if no party has an incentive to unilaterally switch to another norm (i.e., equilibrium norms are robust to an "invasion" of another norm).

It is easy to see that in our game *self-interest* is the only equilibrium in the no conflict treatment. The reason is that in the absence of a sanctioning mechanism the Principal

always has an incentive to switch to SI norm and keep the bonus for herself (this holds for both, the PI and the PO norm).

In the conflict treatment the SI norm remains an equilibrium because the purely self-interested strategies obviously are mutual best responses for both players. The PI norm is not an equilibrium. Whereas the Principal would not have an incentive to switch norms if the agent plays the PI strategy (conflicts enforce bonus payments), the unconditional bonus payments imply that the Worker has an incentive to switch to the SI norm and shirk on effort. The PI norm is therefore vulnerable to an "invasion" of self-interested agents. The PO norm, in contrast, is an equilibrium in the conflict treatment. The logic is that the punishment strategy of the Worker motivates the Principal to pay the bonus, and the contingent bonus payments of the Principal motivate the Worker to provide high effort. The PO norm in the conflict treatment is therefore the only equilibrium in which the worker provides high effort. Notice though that the PO norm does not yield the first-best outcome, because it implies that there are conflicts on the equilibrium path (if conflicting signals imply that the Worker does not get a bonus although he feels that he deserves it).

In practice it is unrealistic to expect that play will be characterized by a single norm. However, we can measure how close behavior is to a particular norm. Notice that players face a sequence of binary choices, that in turn can be viewed as a probability vector in a simplex. For the Principal, she has a single choice - how much bonus to pay after having observed her signal ( $s_P \in \{0, 1\}$ ). Hence, the strategy of the Principal is given by:

$$\sigma_P = (p_0, p_1) \in \Delta^2 = \{(p_0, p_1) \mid 1 \geq p_0, p_1 \geq 0\},$$

where  $p_0$  is the probability of paying a bonus when the signal is low ( $s_P = 0$ ), and  $p_1$  is the probability of paying a bonus when the signal is high ( $s_P = 1$ ). With this notation it follows:

$$\sigma_P^{PI} = (1, 1), \sigma_P^{PO} = (1, 0), \sigma_P^R = (0, 0).$$

The Worker has a more complex decision. He needs to choose high or low effort and then choose to initiate conflict depending upon his bonus and his observed signal. Suppose that  $p_e$  is the probability of high effort. Let his information set be given by  $I = \{B1, B0, 01, 00\}$ , where  $B1$  means a bonus was paid and he observes a high signal,  $01$  is no bonus and low signal and so on. Thus the strategy of the Worker can be given by vector of probabilities of effort and imposing a cost upon the Principal:

$$\sigma_W = (p_e, p_{B1}, p_{B0}, p_{01}, p_{00}) \in \Delta^5.$$

The three types of norm behaviors are:

$$\begin{aligned}\sigma_W^{PI} &= (1, 0, 0, 1, 1), \\ \sigma_W^{PO} &= (1, 0, 0, 1, 0), \\ \sigma_W^{SI} &= (0, 0, 0, 0, 0).\end{aligned}$$

Notice that the extensive form game tree is a binary tree with three players: Nature, Principal and Worker. Since it is a binary tree we can estimate the strategies using a multi-nomial logit. Specifically, each norm implies a binomial distribution at each node, which in turn implies a probability distribution over the final nodes in the game that can be represented by a logit model. We can then estimate how well the data is represented by a particular combination of norms. The details are provided in Section 4.

### 3. THE EXPERIMENTS

With our experiments we explore the conditions under which an efficiency-enhancing conflict culture emerges. In the first experiment (the baseline experiment) participants play the subjective evaluation game without any further intervention or clarification from the experimenter. We identify the causal effect of the opportunity to engage in conflicts by comparing a treatment in which conflicts are excluded by design (no conflict treatment) to a treatment in which the conflict option is available (conflict treatment). As we illustrate later, in this first experiment the participants failed to use the conflict option in a beneficial way so that efficiency was lower in the conflict treatment than in the no conflict treatment. In response to the results of the first experiment, we decided to run a series of additional experiments to explore different factors that potentially facilitate the emergence of a coordinated conflict norm. In total, we run three additional experiments that explore the impact of structured communication (the communication experiment), codes of conduct (the agreement experiment), and institutionalized procedures (the grievance experiment).

Each experiment stands on its own and includes at least 10 independent sessions comparing a conflict and a no conflict treatment. Moreover, each experiment is based on data from about 200 participants (in all four experiments together we had 852 participants in 42 sessions). As each experiment was conducted in a different time period, we present the results of each experiment in isolation and do not pool our data in this section. We establish our main results using two-sided, non-parametric tests that use session-level averages from the same experiment as independent observations. We present additional results in Section 4 on norm adherence across experiment. For this part we will use a pooled data set that includes observations from all experiments together. Doing so allows us to apply more complex statistical models.

We next describe the subjective evaluation game underlying all our experiments and provide details regarding choice structure and parameters. Thereafter we describe each of our four experiments separately. In particular, for each experiment we outline the specific design characteristics, describe the procedures, and finally report the results.

**Implementation of the Subjective Evaluation Game.** All our experiments are based on the same implementation of the subjective evaluation game. In this section we outline a step-by-step account of the game and describe each player’s strategy space and decision sequence in detail. Participants play the experiment for a total of 15 periods. Principals and workers are randomly re-matched at the beginning of every period. Each period consists of the following steps:

- (1) *Matching and Contract*: Each worker is randomly assigned to a principal. The Worker gets a standardized contract. The contract guarantees a fixed wage  $w = 100$ , which the Worker gets with certainty. In addition, the Principal has the possibility to pay a bonus,  $b \in \{0, B = 50\}$ , after she has received a private signal  $s_P$  about the output  $r$  produced by the Worker.
- (2) *Worker’s Effort Choice*: The Worker can choose between a low,  $e = 0$ , and a high effort level,  $e = 1$ . Choosing the high effort level is associated with a cost  $c = 10$ . The effort exerted by the Worker produces a return that can either be high,  $r_H = 350$ , or low,  $r_L = 150$ . The effort choice defines the probability with which the output is high or low:  $Prob(r = r_H|e = 1) = Prob(r = r_L|e = 0) = 0.85$  and  $Prob(r = r_L|e = 1) = Prob(r = r_H|e = 0) = 0.15$ .
- (3) *Output Determination and Signals*: After the Worker has chosen his effort level, a computerized random device determines the realized output. Neither the Principal nor the Worker observe the output. Instead, each one of them receives a private signal  $s$  indicating either a high output,  $s = 1$ , or a low output,  $s = 0$ . The Principal’s signal depends on the realized output  $r$ :  $Prob(s_P = 1|r = r_H) = Prob(s_P = 0|r = r_L) = 0.75$ . The Worker’s signal depends on the Principal’s signal  $s_P$ :  $Prob(s_P = 1|s_P = 1) = Prob(s_P = 0|s_P = 0) = 0.75$ .<sup>6</sup>
- (4) *Principal’s Bonus Payment*: After the Principal observes the signal,  $s_P$ , she makes a decision regarding the bonus payment  $b$ .
- (5) *Worker’s Initiation of Conflict*: Finally, the Worker observes the private signal  $s_A$  and the Principal’s bonus choice  $b$  and decides whether to initiate conflict,  $d \in \{0, 1\}$ . When initiating a conflict the Worker determines the intensity of the conflict  $F \in$

---

<sup>6</sup>Obviously, it would be possible to write down an equivalent structure of signals in which both signals depend on output. However, we chose this way of presenting the signals to the participants because it makes the imperfect correlation of the two signals very transparent to the participants.

$[0, 100]$ . Conflict harms the Principal by reducing her payoff by  $f(d) = dF = F$  and creates a cost  $k(d) = dFK$  for the Worker, where  $K \in \{0.1, \infty\}$  depending on the treatment.

**Payoffs.** In this section, we explain in detail the material payoffs and show that for this employment problem there are two norm equilibria - either all self-interested (SI) or all pay-for-output (PO). Instead, the pay-for-input norm (PI) is never an equilibrium since the workers face an incentive to shirk on effort.

Table 1 shows the social surplus of the conflict treatments for each combination of norms played by the Principal and the Worker. Notice that the highest surplus is achieved when both parties follow the PI strategy. This is because the PI entails high effort and no conflict costs. If workers and principals were trustworthy, then this would be the outcome all would agree upon. The smallest surplus is achieved when the Worker follows the rational strategy regardless of the strategy followed by the principal. This is because the Worker never exerts high effort, which decreases greatly the expected output produced for the principal.

TABLE 1. Social Surplus in Conflict Treatment

		<b>Worker</b>		
		Self-Interest ( $e = 0$ )	Pay for Input ( $e = 1$ )	Pay for Output ( $e = 1$ )
<b>Principal</b>	Self-Interest	180	200	236
	Pay for Input	180	310	290
	Pay for Output	180	274	290

As we mentioned before, the highest surplus is achieved when both parties follow the pay-for-input strategy. However, this is not a stable strategy because the agent has high incentives to shirk. Table 2 normalizes the surplus to the equilibrium utility and the gain from deviation. In the table, movement along a column corresponds to changes in the Principal’s payoff, while movement along a row corresponds to changes in the Worker’s payoff relative to the norm for that row. The table illustrates that when the Principal and the Worker choose to play the combination of “SI/SI” or “PO/PO”, there are no gains to deviation, hence, these are stable norms. Instead, if they follow the PI norm, the Worker has a net gain of 10 by moving to the rational norm and, hence, this is not a stable norm. However, as we shall see there is evidence that some players would like to coordinate upon the PI norm.

In the no conflict treatments the SI norm is the only stable norm. The reason is that in the absence of conflicts, principals always have an incentive to play the SI strategy.

### Experiment 1: Baseline.

TABLE 2. Gain to Deviation from Norm in Conflict Treatment

		<b>Worker</b>					
		Self-Interest ( $e = 0$ )		Pay for Input ( $e = 1$ )		Pay for Output ( $e = 1$ )	
		Principal	Worker	Principal	Worker	Principal	Worker
<b>Principal</b>	Self-Interest	0	0	-50	-20	-15	-16.75
	Pay for Input	-50	10	0	0	-16.25	-1.88
	Pay for Output	-16.25	-5.63	-16.25	-1.38	0	0

**Design.** In our first experiment the participants played the game exactly as described in section 3. We implemented the following two treatments:

- **No Conflict Treatment:** In this treatment the interaction between the Principal and the Worker ends after the bonus payment of the Principal. Formally, this corresponds to a version of the subjective evaluation game in which conflicts are prohibitively costly for the Worker,  $K = \infty$ . In this treatment Step 5 (the conflict stage) of the above procedure does not exist.
- **Conflict Treatment:** In this treatments all 5 steps of the game are played and the cost of conflict is  $K = 0.1$ .

**Data Collection and Procedural Details.** We conducted the study at the behavioral laboratory of HEC Lausanne (LABEX). Participants were recruited from the regular subject pool, covering all fields of study. We used ORSEE (Greiner, 2015) for the recruitment and z-Tree (Fischbacher, 2007) for programming the experiment. We ran 10 sessions with a total of 204 subjects and we conducted the sessions in November and December 2012. Sessions were randomly allocated to treatments and within sessions participants were randomly assigned to roles. We aimed at 24 subjects per session but some sessions were smaller due to no-shows. Five sessions had 16 subjects only and the other sessions had either 18 or 20 subjects. Sessions lasted for 50 to 80 minutes including the reading of the instructions and the final cash payments. Subjects received a show-up fee of 10 CHF and experimental points were converted at a rate of 70 points per CHF. Average total earnings were 23.8 CHF (24.2 CHF for subjects in the role of worker, and 23.3 CHF for subjects in the role of employer). The roles in the experiment were labeled as “worker” and “employer”. Role assignments, choices, and earnings were completely anonymous.

**Results.** We first illustrate how the conflict treatment affects the main variables of interest (effort and surplus). Subsequently, we explore the underlying mechanisms by investigating how conflict strategies affect bonus payments and effort choices.

**Result 1** (Baseline - Effort and Surplus). *The opportunity to engage in costly conflict leads to a small and statistically insignificant increase in worker’s effort. However, conflict costs*



dominate the efficiency gains produced by a higher effort, so that total surplus in the conflict treatment is lower than in the no conflict treatment.

The top row of Figure 1 compares the relative frequency of high effort (left-hand side) and the average total surplus (right-hand side) in the no conflict and conflict treatment of the baseline experiment. The figure shows that the rate with which workers choose high effort rises from 40.4% in the no conflict treatment to 47.5% in the conflict treatment. This corresponds to an increase of 18%, but this increase is not statistically significant (RS:  $p = 0.420$ ).<sup>7</sup> Moreover, the figure also reveals that the increase in the frequency of high effort does not translate into a higher total surplus. On the contrary, total surplus decreases insignificantly from 233.8 in the no conflict treatment to 221.6 in the conflict treatment (RS:  $p = 0.222$ ).

Figure 2 illustrates the reason we find a lower surplus in the conflict treatment. The figure contrasts the total cost of conflict (i.e., the Worker’s cost of initiating conflict plus the damage imposed on the Principal) with the additional gains from trade that result from the increased effort level. We find that average conflict costs (20.0) are larger than average gains from higher effort (7.8) resulting in a negative effect of conflict on efficiency.

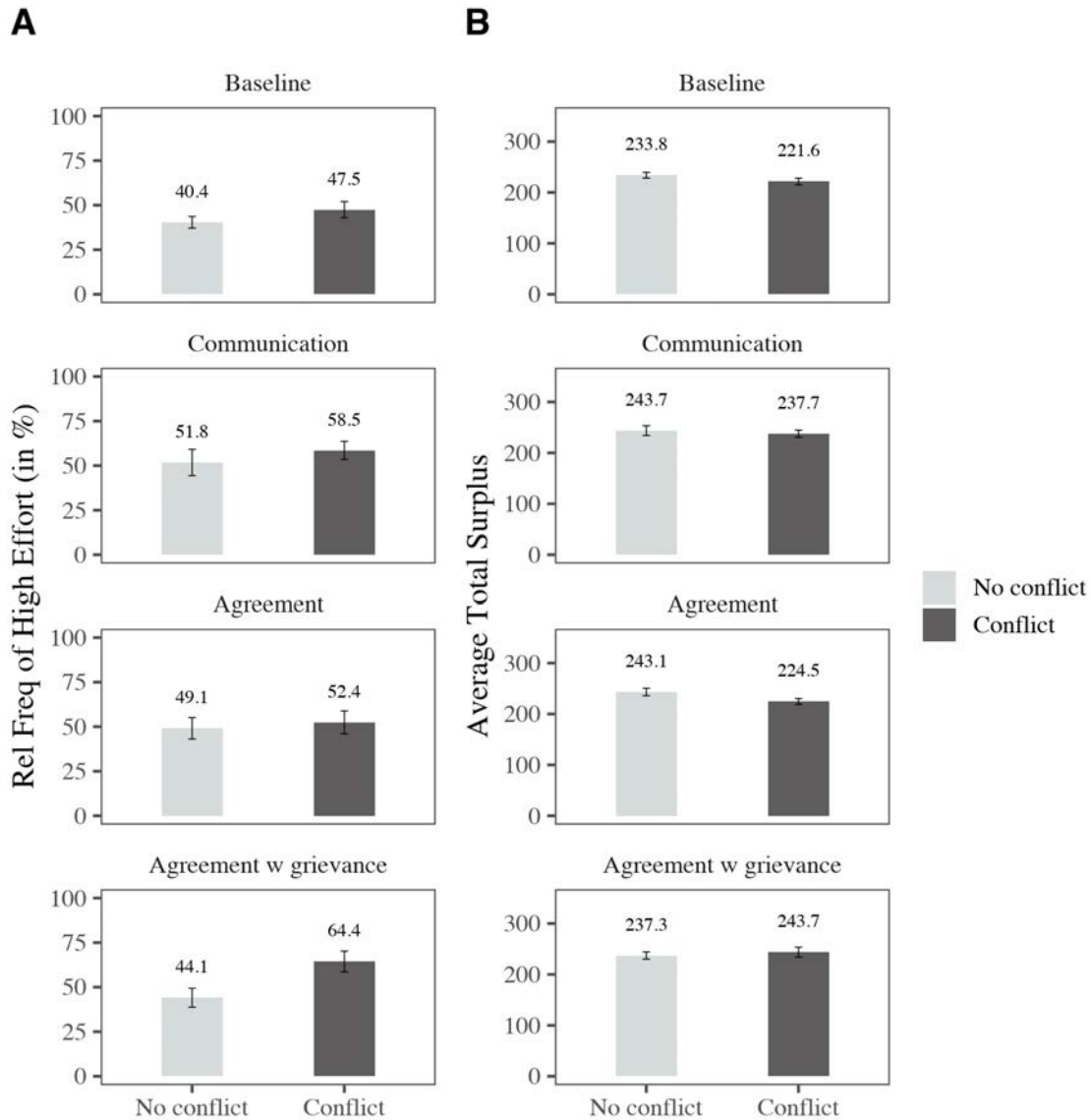
As our participants play the game for 15 periods with different partners, learning effects might occur and it is important to look at the dynamics. Table 3 presents the relative frequency of high effort and average total surplus over time (in bins of 5 periods). The table reveals that effort and surplus are subject to a downward trend in both treatments, but in particular in the absence of conflict. The frequency of high effort drops by 19% in the conflict condition (SR:  $p = 0.188$ ) and by 35.6% in the no conflict condition (SR:  $p = 0.063$ ). Total surplus decreases by 10% in the conflict condition (SR:  $p = 0.063$ ) and by 13.4% in the no conflict condition (SR:  $p = 0.063$ ). Although the presence of conflict slows down the decrease in effort provision over time compared to the no conflict treatment, the difference between effort levels across treatments remains insignificant, even in the final periods of the experiment (RS:  $p = 0.310$ ). Moreover, total surplus keeps being superior in the absence of conflicts. Hence, Result 1 holds throughout the experiment.

TABLE 3. Baseline - Development of effort and surplus over time

Periods	High Effort (in %)			Average Surplus		
	1-5	6-10	11-15	1-5	6-10	11-15
Conflict	53.06	46.53	42.86	235.31	217.74	211.65
No conflict	52.83	34.34	33.96	259.43	216.76	225.10
p-values (RS)	0.753	0.151	0.421	0.222	1.000	0.310

<sup>7</sup>If not explicitly stated otherwise reported p-values are based on two-sided, non-parametric rank sum (RS) or signed rank (SR) tests using session averages as independent observations.

FIGURE 1. Effort and surplus

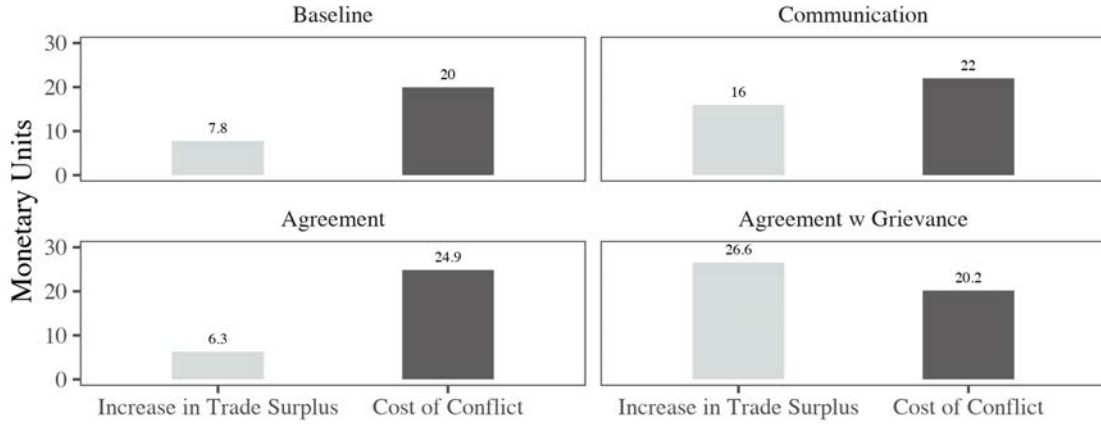


Note: Error bars represent plus/minus one standard error of the mean, clustered at the session level.

To understand the reasons for why conflicts fail to be efficiency-enhancing, it is instructive to take a careful look at the conflict pattern that emerges in our first experiment.

**Result 2** (Baseline - Conflict Pattern). *Most workers in the conflict treatment either refrain from engaging in conflict or follow a conflict pattern that aims at enforcing the pay-for-input norm. As a consequence, the Worker's subjective performance signal has a very low impact on the conflict rate. If conflicts are initiated at all, they most frequently occur when workers provide high effort, but do not receive a bonus.*

FIGURE 2. Gains and costs of conflict



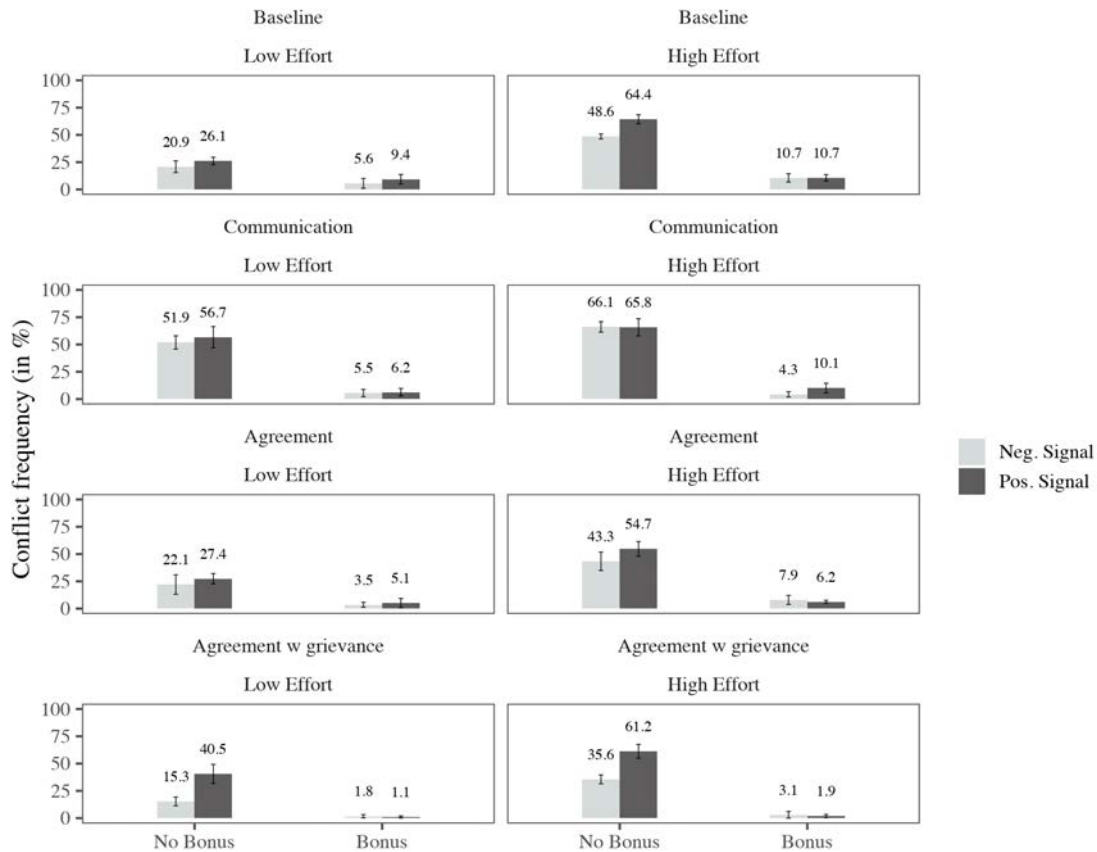
Note: Error bars represent plus/minus one standard error of the mean clustered at session level.

The top row of Figure 3 presents the relative frequency of conflict initiation contingent on three determinants, the Worker’s effort level, the Worker’s subjective performance signal and the Principal’s bonus payment. The left-hand panel covers low effort observations, the right-hand panel covers high effort observations. Two insights emerge from the observed pattern in this figure: First, the punishment decision of a large set of workers is guided by the pay-for-input norm (PI): the conflict rate is highest if workers who decide to exert high effort are not rewarded with a bonus. Second, the pay-for-output (PO) norm matters only for a minority of workers as the subjective performance signal of workers has a weak impact on the decision to engage in conflict.

Figure 3 shows that the worker’s effort choice in combination with a lack of a bonus payment is the most important determinant of conflict. The highest conflict rate is observed if the Worker chooses high effort, gets a positive subjective signal, but does not receive a bonus (64.4%). Keeping everything else constant (high effort, positive worker signal) the conflict rate drops to 10.7% if the Principal pays the bonus, a decrease of 83% (SR:  $p = 0.063$ ). Likewise, if we compare the highest conflict rate (64.4%) to the corresponding low effort case (keeping the absence of a bonus and the positive worker signal constant) the conflict rate drops to 26.1%, a decrease of 59.4% (SR:  $p = 0.063$ ). At the same time, the figure also reveals that the Worker’s signal has a weaker impact on conflict initiation. For example, if we compare the highest conflict rate (64.4%) to the case in which the Worker’s subjective

signal is negative (keeping the high effort and the absence of the bonus payment constant), the conflict rate only drops from 64.4% to 48.6%, a decrease of 24.5% (SR:  $p = 0.063$ ).<sup>8</sup>

FIGURE 3. Determinants of conflict



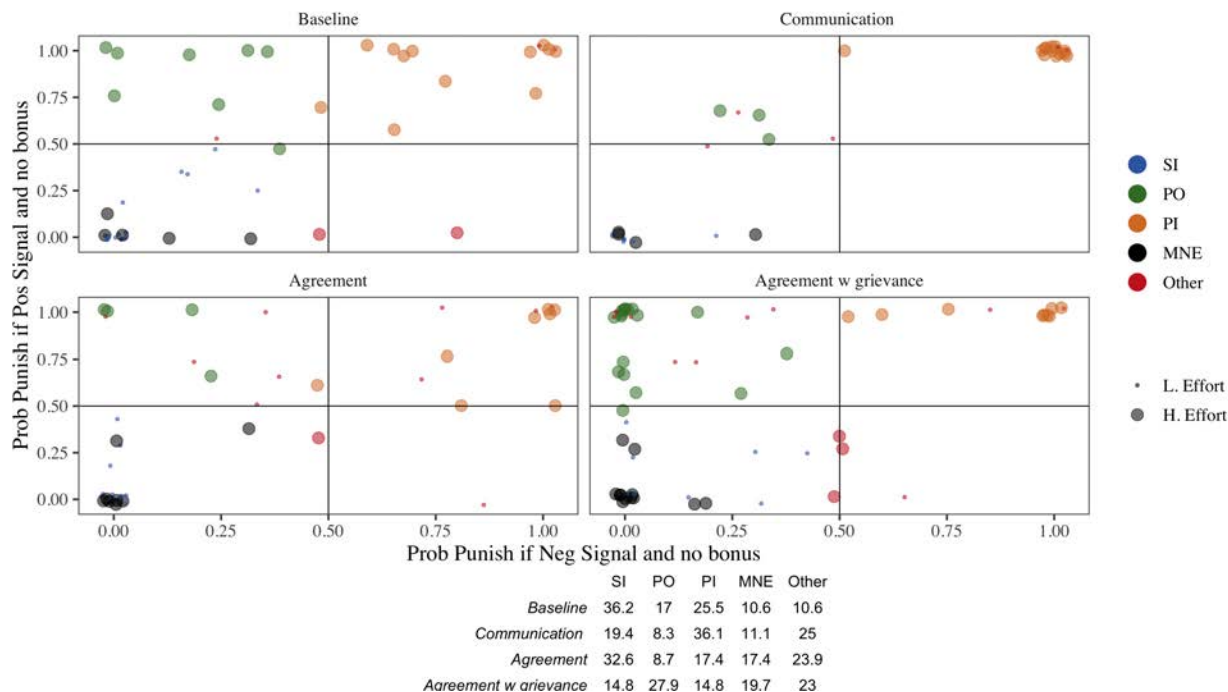
Note: Error bars represent plus/minus one standard error of the mean, clustered at the session level.

Figure 4 provides a different perspective on the workers' conflict strategies. In this figure, instead of averaging behavior across subjects, the figure shows different distributions of punishment patterns taking individual workers as the unit of observation, i.e., each dot in the figure represents a single worker. Moreover, as conflicts emerge predominantly when the Worker does not receive a bonus, the figure focuses only on conflict initiation in the cases in which the Principal decides not to pay a bonus. The figure illustrates the extent to

<sup>8</sup>The finding that the conflict rate is predominantly determined by bonus payments and effort choices receives further support when we focus the analysis on the cases in which the Worker receives a negative signal. If the Worker exercises high effort, the conflict rate decreases from 48.6% to 10.7% if the Principal decides to pay the bonus (SR:  $p = 0.063$ ). Similarly, if we keep the absence of a bonus payment constant, the conflict rate drops from 48.6% to 20.9% if the Worker decides to exert low effort (SR:  $p = 0.125$ ).

which the decision to initiate conflict depends on the worker's subjective signal (given that no bonus payment has been made). On the x-axis (y-axis), we measure the probability of initiating conflict if the signal is negative (positive). In addition, the figure also reveals each worker's effort level: small (large) dots represent workers that mostly picked a low (high) level of effort. Finally, the figure uses colors to categorize the workers. We distinguish four types of workers: pay-for-output norm enforcers (PO), pay-for-input norm enforcers (PI), self-interest non-enforcers (SI) and motivated non-enforcers (MNE). PO enforcers pick high effort and only engage in conflict if they do not receive a bonus after having observed a positive subjective signal. These workers appear as large dots in the top-left quadrant. PI norm enforcers also choose high effort, but they ignore their signal and engage in conflict whenever they do not receive the bonus. These workers appear as large dots in the top-right quadrant. MNE are workers who exert high effort, but do not engage in norm enforcement (i.e., stay away from conflicts). They appear as large dots in the bottom-left quadrant. SI, finally, pick low effort and do not engage in conflict. They appear as small dots in the bottom-left quadrant.

FIGURE 4. Worker's Effort and Conflict Decisions (Individual Level)



Note: In this graph each dot represents one worker in each treatment. The dots give information on (1) the level of effort, represented by size of the dot; and (2) the conflict behavior, represented by the position of the dot in the graph. The space of the graph is determined by the probability of conflict if the Worker does not receive a bonus and the signal is positive (y-axis) and the probability of conflict if the Worker does not receive a bonus and the signal is negative (x-axis). Furthermore, the colors of the dots help to emphasize the norm behavior that each worker most closely follows.

The figure shows that the worker’s population fails to coordinate on any particular norm. To begin with, only 42.5% of all workers can be defined as norm enforcers (PI or PO). Moreover, there are more PI enforcers (25.5%) than PO enforcers (17.0%). The majority of non-enforcers (46.8%) are SI and exert mostly low effort (36.2%) and MNE are rare (10.6%). The remaining workers (10.6%) show profiles that do not correspond to either of the four types described above.

In Table 4 we present the development of the conflict rate over time. The table reports conflict initiations for cases in which no bonus has been paid contingent on the worker’s effort level. Most interesting is the development of the conflict rate in the situations in which the Worker exerts high effort. The table reveals that in the first 10 periods of the experiment, the conflict rate is only slightly higher when the Worker’s subjective performance signal is positive than when it’s negative and the difference is not statistically significant. In the final five periods, in contrast, we observe a substantial increase in the conflict rate when the worker’s signal is positive (from roughly 60% in periods 1-10 to more than 80% in periods 1-15). No such increase is observed for the cases in which the worker’s signal is negative. As a consequence, the conflict rate in the final five periods is marginally significantly higher for a positive worker signal than for a negative one (SR: one-sided  $p = 0.063$ ). However, despite the fact that the dynamic analysis suggests that the pay-for-output norm may increase in importance over time, it is also important to emphasize that the conflict rate with negative signals remains roughly constant and at a high level throughout the experiment (45-52%). This observation confirms that workers fail to coordinate on the enforcement of the pay-for-output norm (even after having played the game for many periods).

TABLE 4. Baseline - Development of conflict rate over time (only if no bonus received)

Effort Choice Periods	Conflict Rate (in %)					
	High Effort			Low Effort		
	1-5	6-10	11-15	1-5	6-10	11-15
Positive Signal	59.52	59.46	81.81	18.52	26.92	31.43
Negative Signal	47.83	52.17	45.83	19.15	25.93	17.74
p-values (SR)	0.20	0.79	0.06	0.50	0.44	0.64

From a theoretical point of view, the workers’ failure to coordinate on the pay-for-output norm is problematic because the somewhat erratic conflict patterns displayed in Figures 3 and 4 does not create well-aligned financial incentives for principals to use their bonus payments in a way that rewards those workers who are perceived to be productive. A more detailed investigation of principals’ bonus payments confirms the empirical relevance of this concern.

**Result 3** (Baseline - Bonus Payments). *Principals in the conflict treatment pay the bonus significantly more often than principals in the no conflict treatment. However, the bonus payments in the conflict treatment are only partially consistent with the pay-for-output norm. Although the principals pay the bonus more often when their subjective signal is positive, bonus payments also occur quite frequently when the Principal’s subjective signal is negative.*

The opportunity to engage in conflict has a strong positive impact on the overall frequency with which principals pay a bonus to the Worker: the bonus rate increases from 14.3% in the no conflict treatment to 42.6% in the conflict treatment (RS:  $p = 0.008$ ). Figure 5 shows the frequency of bonus payments as a function of the Principal’s private signal in both treatments. The pattern observed in the bonus payments is only partially in line with the pay-for-output norm; although in both treatments principals are more likely to pay the bonus if their subjective signal is positive, bonus payments occur frequently when the signal is negative. In the conflict treatment the bonus rate is 57.4% if the Principal’s subjective signal is positive and 29.0% if the signal is negative (SR:  $p = 0.063$ ). The corresponding numbers in the no conflict treatment are 24.7% and 5.4% (SR:  $p = 0.063$ ).

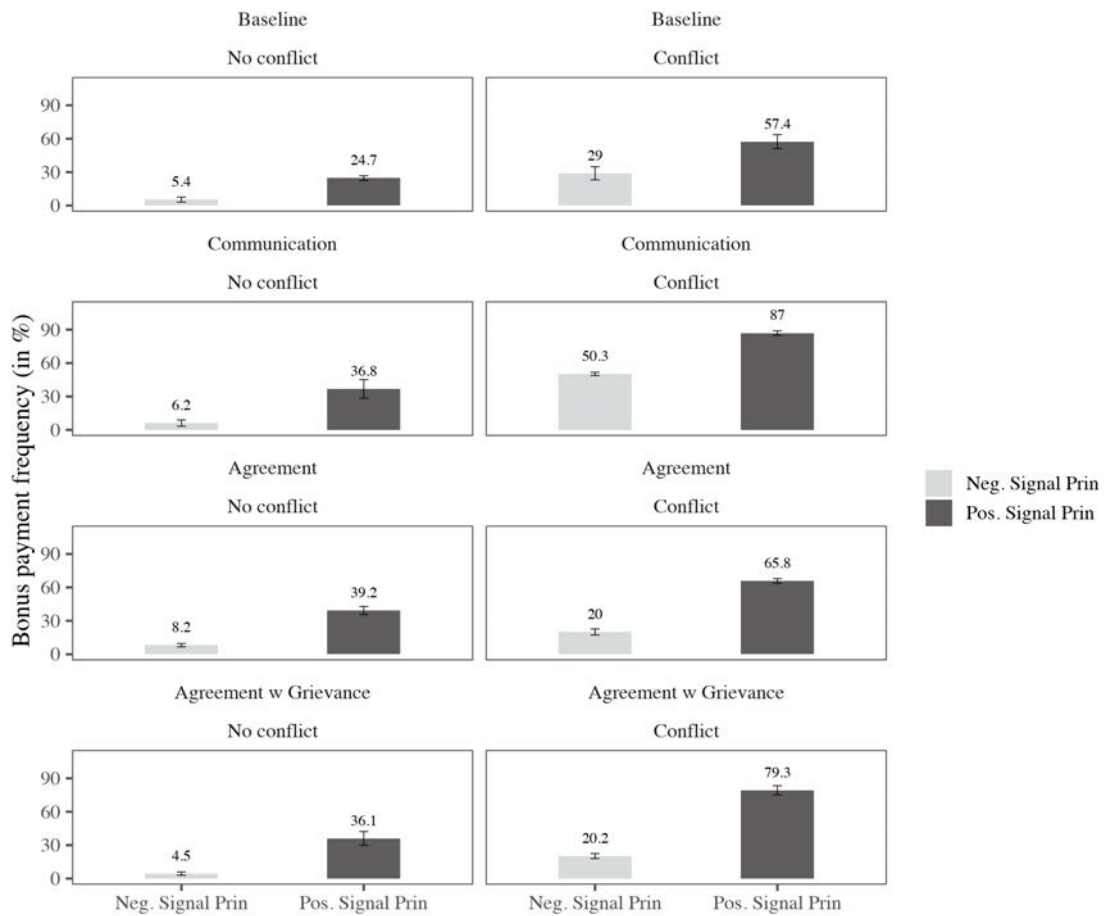
Table 5 shows bonus payments in both treatments over time contingent on the signal observed by the Principal. In the conflict treatment the observed patterns remain stable over time. In the no conflict treatment bonus payments show a decreasing trend when the signal is high (this development occurs in parallel with the observed decrease in effort displayed in Table 3).

TABLE 5. Baseline - Development of bonus payments over time

Principal’s Signal Periods	Bonus Payments (in %)					
	High Signal			Low Signal		
	1-5	6-10	11-15	1-5	6-10	11-15
Conflict	54.20	60.18	58.33	30.70	28.03	28.47
No conflict	32.14	23.90	16.81	4.80	6.41	4.80
p-values (RS)	0.02	0.01	0.01	0.04	0.03	0.04

Figure 6 displays distributions of bonus payment patterns in the conflict treatment at the individual level; each dot in the figure represents one principal. The only information that principals have when deciding about the bonus payment is their subjective performance signal. The Figure, therefore, displays the frequency with which principals pay the bonus as a function of their signal. The horizontal axis displays the bonus rate in response to a negative signal and the vertical axis the bonus rate in response to a positive signal. Similar to Figure 4, this figure allows to identify norm compliance, but this time for the principals instead of workers. Principals who follow the pay-for-output norm and pay bonuses predominantly

FIGURE 5. Determinants of bonus



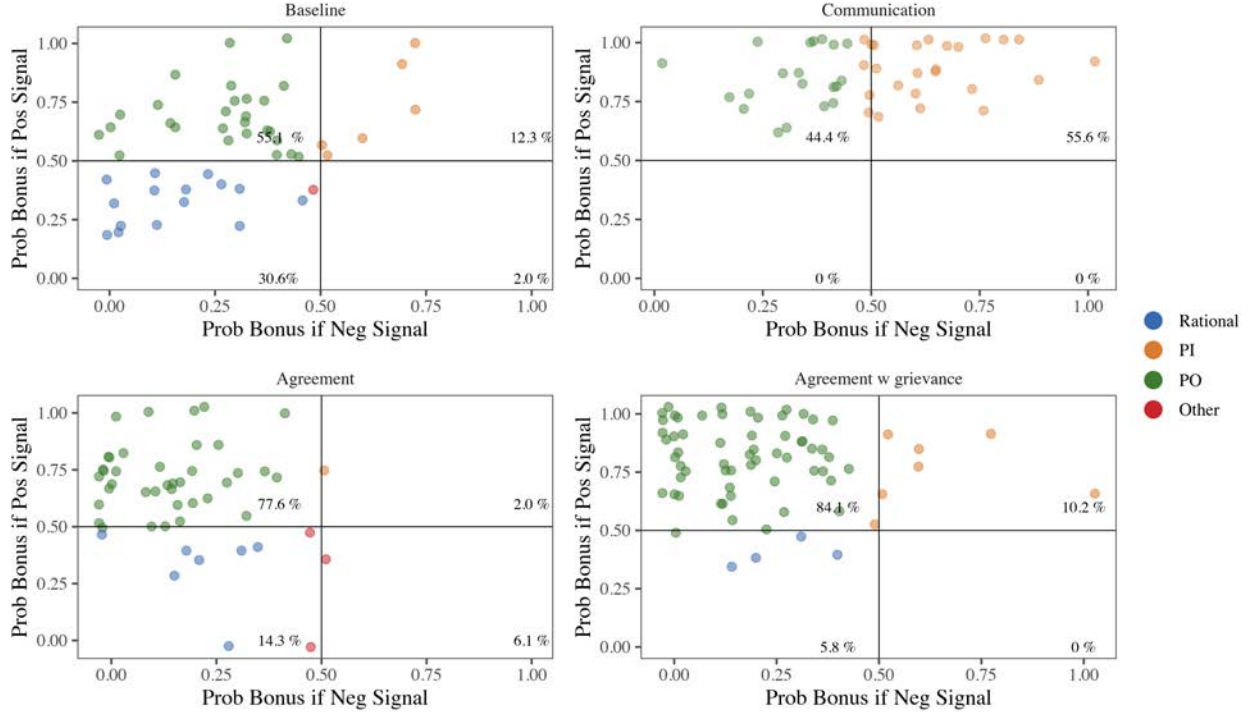
Note: Error bars represent plus/minus one standard error of the mean, clustered at the session level.

after having observed a positive signal are located near the top-left corner. Those who subscribe to the pay-for-input norm and pay bonuses irrespective of their signal are located near the top-right corner. Principals who follow neither norm and never pay bonuses appear near the bottom-left corner.

The Figure reveals that no principal strictly sticks to a particular fairness norm. The large majority of principals pay the bonus more frequently after observing a positive signal (55.1%), but there is considerable heterogeneity in both the bonus frequency and the weight that principals assign to their subjective signal. This suggests that workers' inability to coordinate on the enforcement of the pay-for-output norm is consistent with the observation that many principals also fail to coordinate on the norm. The lack of coordination on the side of principals, in turn, implies that the bonus payments fail to create the optimal monetary



FIGURE 6. Principal’s Bonus Payment Decisions Averaged at the Individual level



Note: In this graph each dot represents one individual in each treatment. The dots give information on the bonus payment behavior of the Principal. The space of the graph is determined by the probability of paying a bonus after a positive signal (y-axis) and the probability of paying a bonus after a negative signal (x-axis). Furthermore, the colors of the dots help to emphasize the norm behavior that each principal most closely follows.

incentive for workers to exert high effort. When comparing the profits of workers before the conflict stage, we find that the expected gains from exerting low effort in the conflict treatment amount to 19.1, whereas the expected gains from exerting high effort are only 14.1. The fact that the opportunity to engage in conflicts does not prevent self-interested workers from being better off by choosing low effort provides an explanation for the high population fraction of self-interest non-enforcers observed in Figure 4 and the limited positive impact that the conflict treatment has on overall effort (see Result 1).

Our first experiment illustrates that it is not obvious that conflicts have an efficiency-enhancing effect in a work environment with subjective performance evaluation. Our results confirm that the presence of multiple fairness norms can lead to coordination failure which, in turn, implies that conflicts fail to have the desirable motivating effects. As a consequence, the cost of conflicts dominate and total surplus is lower in the presence of conflict opportunities than in their absence.

## Experiment 2: Communication.

**Design.** The lack of communication is arguably one factor that renders coordination difficult in our baseline experiment. There is ample evidence in the literature that pre-play communication helps to overcome coordination failure (see e.g., Cooper et al., 1992; Brandts and MacLeod, 1995; Cason and Mui, 2007; Brandts and Cooper, 2007) and communication opportunities are available in most real-life settings in organizations. Our second experiment (subsequently termed “communication experiment”) therefore investigates the extent to which communication helps the trading parties to coordinate on an efficiency-enhancing use of conflicts. The design of the experiment is identical to the one of the baseline experiment, but we now add a communication stage at the beginning of every period. In the communication stage both the Principal and the Worker could pick a message to their trading partner from pre-defined sets of messages.<sup>9</sup>

The message set of the Principal was the same in both the no conflict treatment and the conflict treatment. It contained the following announcements of different bonus strategies:

- (1) I will pay the bonus with certainty.
- (2) I will pay the bonus with high probability.
- (3) I will pay the bonus if I have the impression that you exerted high effort.
- (4) I pay the bonus if my private information indicates a high return.
- (5) I will not pay the bonus.
- (6) I prefer not to send a message.

In the no conflict treatment workers’ messages contained only announcements of their effort strategies. The following options were available:

- (1) I will exert high effort.
- (2) I will exert low effort.
- (3) I prefer not to send a message.

In the conflict treatment workers could add a second part to their message in which they announced their conflict strategies:

- (1) I will always reduce the return.
- (2) I will reduce the return if I do not get the bonus.
- (3) I will reduce the return if I do not get the bonus although I got a good signal.
- (4) I will never reduce the return.

---

<sup>9</sup>There is an emerging consensus in the experimental literature that in many environments free-form communication is more effective than structured communication (Brandts and Cooper, 2007; Brandts et al., 2015). Nevertheless, we consciously decided to use structured communication in this experiment, because including the different fairness norms in the set of pre-specified messages allowed us to make these norms very salient to the participants. This point is particularly relevant in light of the fact that the pay-for-output norm is somewhat complicated. Including it in the set of messages made this choice available even to participants who did not think about this possible norm themselves.

(5) I prefer not to send a message.

This set of pre-specified messages includes the fully self-interested strategy, the pay-for-input norm, the pay-for-output norm and the possibility not to communicate. Messages were selected and sent simultaneously by the Worker and the Principal.

**Data Collection and Procedural Details.** The laboratory, the subject pool, the recruitment process and the software used to program and run the experiment were the same as in baseline experiment. We ran 10 sessions with a total of 176 subjects. We conducted them in March, April, May and December 2013. We aimed at 24 subjects per session but some sessions were smaller due to no-shows. Five sessions had 16 subjects only and all other sessions had at least 18 subjects. Average total earnings were 35.8 CHF (36.6 CHF for subjects in the role of worker, and 35 CHF for subjects in the role of employer).

**Results.** We begin by analyzing how conflicts affect effort and surplus in the presence of communication before we turn to a detailed investigation of communication strategies, conflict patterns and bonus payments.

**Result 4** (Communication - Effort and Surplus). *The impact of conflicts on efficiency in the presence of pre-play communication is similar to the one observed in the baseline experiment. The availability of conflicts leads to a slight increase in workers' effort, but the effect remains insignificant. Total surplus is slightly lower in the conflict treatment than in the no conflict treatment.*

The second row of Figure 1 displays the relative frequency of high effort and average total surplus in the communication experiment. The rate with which workers choose high effort rises from 51.8% in the no conflict treatment to 58.5% in the conflict treatment. This corresponds to an increase of 13%, but the effect is not statistically significant (RS:  $p = 0.690$ ). Total surplus decreases insignificantly from 243.7 in the no conflict treatment to 237.7 in the conflict treatment (RS:  $p = 1.000$ ). The top, right-hand panel of Figure 2 contrasts the total cost of conflict with the additional gains from trade that result from the increased effort in the conflict treatment. Average costs (22) are larger than average benefits (16) resulting in the above reported negative effect of conflict on surplus.

Table 6 presents effort and total surplus over time (in bins of 5 periods). The table reveals that the difference in effort arises exclusively in the final periods of the experiment. In periods 11-15 workers choose high effort in 57.8% of the cases in the conflict treatment, but only in 34.0% of the cases in the no conflict treatment. However, considerable variance across sessions implies that this difference remains statistically insignificant (SR:  $p = 0.420$ ). Total surplus in the conflict treatment never surpasses total surplus in the no conflict treatment in any phase of the experiment.

TABLE 6. Communication - Development of effort and surplus over time

Periods	High Effort (in %)			Average Surplus		
	1-5	6-10	11-15	1-5	6-10	11-15
Conflict	60.89	45.58	57.78	243.28	234.71	235.12
No conflict	63.72	46.05	33.96	256.19	238.42	236.61
p-values (RS)	1.00	0.55	0.42	0.70	1.00	1.00

Result 4 reveals that the availability of structured pre-play communication does not trigger the efficiency-enhancing effects of conflicts. In the following we shed more light on the reasons for this result. We begin with an analysis of communication choices.

**Result 5** (Communication - Message Choices). *Workers and principals most frequently choose messages that are in line with the pay-for-input norm. First, a large majority of workers announce that they will exert high effort and threaten to engage in conflict if they do not get a bonus. Second, the majority of principals state that they will pay the bonus if they have the impression that the Worker exerted high effort, but only a minority communicates explicitly that they plan to make their bonus payment contingent on their subjective signal.*

We start with the communication strategies of workers. The vast majority of workers in both the conflict and the no conflict condition choose to send the message: “I will exert high effort” (79.0% and 85.6%, respectively) and only a small proportion in both treatments chooses to send “I will exert low effort” (15.1% and 12.1%, respectively). Regarding conflict strategies, the most frequent message chosen is “I will reduce the return if I do not get the bonus” (55.1%). This observation hints at the fact that many workers seem to subscribe to the pay-for-input norm. Instead, the message corresponding to the pay-for-output norm “I will reduce the return if I do not get the bonus although I got a good signal” was picked less than half as often (21.8%). Other messages played only a minor role (“I will never reduce the return”, 16.7%) or barely any role at all (“I will always reduce the return”, 0.3%).

Principals remain rather non-committal in their bonus payments announcements. In both the conflict and the no conflict condition, nearly half of the principals choose to send “I will pay the bonus if I have the impression that you exerted high effort” (49.9% and 49.6%, respectively). To interpret the choice of this message it is important to keep in mind that principals also had the possibility to choose the message “I pay the bonus if my private information indicates a high return” (which was picked by 25.6% and 20.0%, respectively). Thus, principals who choose the former message decide explicitly not to signal that they will make their bonus payment strictly contingent on their subjective information because, otherwise, they should have selected the latter message. From this point of view, it seems inconsistent to interpret the most frequently chosen message as being fully in line with the

pay-for-output norm. In fact, the message can also be consistent with the pay-for-input norm, if the Principal plans to base her impression on the worker’s effort message. The other messages available to principals were chosen with low frequency: “I will pay the bonus with certainty” (12.6% and 15.3%, respectively), “I will pay the bonus with high probability” (8.4% and 9.0%, respectively), and “I will not pay the bonus” (0.6% and 1.1%, respectively).

The analysis of the communication strategy suggests that most trading parties signal to each other that they intend to follow the pay-for-input norm. The most frequent message combination (26.2%) is the one in which the Worker announces to exert high effort and to engage in conflict if no bonus is paid and the Principal announces to pay the bonus as long as she has the impression that the Worker exerts high effort. Next we analyze the extent to which actual behavior corresponds with these messages.

**Result 6** (Communication - Conflict Pattern). *Most workers in the conflict treatment either follow conflict patterns that are in line with the pay-for-input norm (PI) or abstain from engaging in conflicts. Conflict patterns consistent with the pay-for-output norm (PO) are rarely observed.*

The second row of Figure 3 presents the relative frequency of conflict initiation in the communication experiment contingent on the worker’s effort level, the worker’s subjective performance signal and the Principal’s bonus payment. The panel on the right-hand side covers observations in which workers exert high effort. The conflict pattern observed in these situations is in line with the PI behavior: the conflict rate is high if the workers receive no bonus and low otherwise, moreover, the subjective signal of the workers plays no role. In the absence of a bonus a high signal triggers conflicts in 65.8% of the cases and a low signal in 66.1% of the cases (SR:  $p = 0.625$ ). Somewhat surprisingly, we observe a similar pattern for observations with low effort levels: in the absence of a bonus a high signal triggers conflicts in 56.7% of the cases and a low signal in 51.9% of the cases (SR:  $p = 0.625$ ).

To better understand the conflict strategies of workers it is instructive to take a look at Figure 4 which shows distributions of punishment patterns based on observations at the individual level (see Result 2 for a detailed explanation of this figure). The figure reveals a strong clustering of observations near the top-right corner showing both large orange dots and small red dots. Whereas the orange dots correspond to individuals who subscribe to the PI, the red dots represent workers who systematically initiate conflicts without providing high effort themselves. This indicates that the conflicts observed in Figure 3 after not obtaining a bonus payment, are not only a consequence of imperfect PI enforcers (who shirk on effort from time to time), but are also caused by workers who do not provide high effort and systematically initiate conflicts.

In this treatment workers tend to coordinate on the PI more frequently than in the baseline. Overall, the population size of norm enforcers remains the same as in the baseline, however, the concentration of norm enforcers who follow the pay-for-input norm is even more pronounced. The overall population share of norm-enforcers - PI and PO together - amounts to 44.4%, which is almost identical to the corresponding rate in the baseline experiment: 42.5%. However, the fraction of PI corresponds to 80% of the norm enforcers in the communication experiment compared to 60% in the baseline experiment. Moreover, the population share of motivated non-enforcers (MNE) remains roughly constant - from 11.1% in the communication treatment to 10.6 in the baseline. The share of SI decreases from 36.2% (baseline) to 19.4% (communication). However, this decrease is misleading because it is almost fully compensated by an increase in the share of unclassified workers who mostly consist of the above discussed systematic conflict initiators who do not provide high effort.

In Table 7 we show how the conflict rate after not receiving a bonus develops over time. The table distinguishes between observations with positive and negative subjective signals of the Worker and high and low effort provision. The Table reveals the absence of response of the workers to their private signal when deciding to initiate conflict. This finding reinforces the interpretation that workers predominantly focus on the enforcement of the PI norm in the communication experiment.

TABLE 7. Communication - Development of conflict rate over time (only if no bonus received)

Effort Choice Periods	Conflict Rate (in %)					
	High Effort			Low Effort		
	1-5	6-10	11-15	1-5	6-10	11-15
Positive Signal	63.64	62.50	72.73	71.43	50.00	37.50
Negative Signal	58.33	68.42	73.68	40.00	61.54	53.57
p-values (SR)	0.89	0.85	NA	0.10	NA	0.63

The analysis of the communication strategies and the conflict patterns suggests that workers have a strong focus on the pay-for-input norm in the communication experiment. From the analysis of the communication choices we already know that most principals do not explicitly announce that they will make their bonus payment contingent on their subjective signal (as in compliance with the pay-for-output norm would require), but seem to prefer a non-committed message that promises a bonus payment as long as the Principal has the impression that the Worker worked hard. We now explore how all this translates into the principals' decision to pay bonuses.

**Result 7** (Communication - Bonus Payments). *As in the baseline experiment, principals are more likely to make bonus payments in the conflict treatment than in the no conflict treatment.*

However, in line with their non-committal announcements, principals do not seem to follow either norm very systematically. Most principals condition their bonus payments to some extent on their signal of Worker performance, but many principals also pay bonuses with high frequency after a negative signal.

At the average level, we observe that conflict increases the fraction of bonus payments and that the large majority of principals pay the bonus more frequently after having observed a positive signal. The second row of Figure 5 shows the frequency of bonus payments as a function of the Principal’s private signal. In the conflict treatment the bonus payment rate is 87% if the Principal’s subjective signal is positive and 50.3% if the signal is negative ( $p = 0.063$ ). The corresponding numbers in the no conflict treatment are 36.8% and 6.2% ( $p = 0.063$ ). These bonus patterns suggest that principals are not always following a norm.

At the individual level, we observe that most principals play a mix of the pay-for-input and the pay-for-output norm. The top-right panel of Figure 6 displays bonus payment patterns in the conflict treatment at the individual level. If most principals were to follow the pay-for-input norm (pay-for-output norm), dots should be clustered in the top-right (top-left) corner. Instead we observe that most dots lie between the top-right and top-left quadrant, indicating that the principals are not systematically following either norm.

Table 8 shows bonus payments contingent on the signal observed by the Principal over time. In the conflict treatment, the bonus payment frequency remains very stable over time. In the no conflict treatment, in contrast, the frequency of bonus payments decreases quite drastically towards the end of the experiment.

TABLE 8. Communication - Development of bonus payments over time

Principal’s Signal Periods	Bonus Payments (in %)					
	High Signal			Low Signal		
	1-5	6-10	11-15	1-5	6-10	11-15
Conflict	86.21	85.60	89.42	46.79	49.00	54.55
No Conflict	41.46	41.96	7.77	10.87	25.49	0.89
p-values (RS)	0.02	0.01	0.01	0.01	0.01	0.01

Our second experiment shows that communication does not solve the coordination problem of principals and workers. If anything, workers seem to have an even stronger focus on the pay-for-input norm than in the baseline experiment (this is reflected in both their communication strategies and their actual choices). In the conflict treatment, principals do not systematically follow a norm and, accordingly, their bonus payments fail to provide financial incentives for the Worker to work hard. The expected gains for the Worker after exerting high effort in the conflict treatment are only 27.5, whereas the expected gains after exerting

low effort are 31.1. As a consequence, the results resemble those of the baseline experiment; conflicts fail to enhance cooperation and efficiency because the costs of conflict outweigh the benefits of additional effort.

### **Experiment 3: Agreement (Code of Conduct).**

**Design.** The first two experiments rely on self-coordination. The results indicate that people have a tendency to focus on the simpler and arguably more intuitive pay-for-input norm. Unfortunately, this norm is not well-suited to increase cooperation in the subjective evaluation framework we are interested in. The fact that spontaneously emerging conflict cultures may be ineffective or even counterproductive suggests that there may be an important role for active management. Code of conducts are one way through which many organizations transmit values, guidelines and proper practices to their workforce. Our next experiment (henceforth called “the agreement experiment”) therefore aims at studying the impact of an external appeal to coordinate on the pay-for-output norm. In this experiment we begin each period with a stage in which we ask both the Principal and the Worker to electronically sign a non-binding agreement in which they confirm their intention to follow the pay-for-output norm. Failure to sign this agreement by at least one party implies that trade does not occur and an unattractive outside option ( $u_r = 40$ ) is implemented. The agreement specifies a simple code of conduct for the relationship. The suggested choices correspond to the strategies prescribed by the pay-for-output norm. In the no conflict treatment the Worker agrees to exert high effort, whereas the Principal agrees to pay the bonus if and only if her subjective signal is positive. In the conflict treatment, the code of conduct has an additional paragraph in which the Worker agrees to engage in conflict only if he does not receive a bonus although his subjective performance signal is positive. In the instructions all participants are made aware of the fact that the agreement is not binding, i.e., signing the code of conduct does not alter the strategy set available to the players. An example of the agreement in English can be found in the Appendix.

**Data Collection and Procedural Details.** The laboratory, the subject pool, the recruitment process and the software used to program and run the experiment were the same as in the previous experiments. We ran 10 sessions with a total of 200 subjects and we conducted them in November and December 2017. We aimed at 24 subjects per session but some sessions were smaller due to no-shows. Two sessions had 18 subjects only and the other session had either 20 or 22 subjects. Sessions lasted for 50 to 80 minutes including the reading of the instructions and the final cash payments. Subjects received a show-up fee of 10 CHF and experimental points were converted at a rate of 70 points per CHF. Average total earnings



were 34.7 CHF (33.3 CHF for subjects in the role of worker, and 36.1 CHF for subjects in the role of employer).

**Results.** We first analyze how the need to sign a code of conduct before interacting with a trading partner affects the main outcomes of the experiment:

**Result 8** (Agreement - Effort and Surplus). *The introduction of a code of conduct does not substantially change the impact of conflicts on effort and surplus. As in the baseline experiment, workers' effort is not significantly higher and total surplus is lower in the conflict treatment than in the no conflict treatment.*

The third row of Figure 1 shows the relative frequency of high effort and average total surplus in the agreement experiment. The figure shows that the rate with which workers choose high effort rises from 49.1% in the no conflict treatment to 52.4% in the conflict treatment. This corresponds to an increase of 6.7%, but this effect is far from significant (RS:  $p = 0.842$ ). Moreover, total surplus decreases from 243.1 in the no conflict treatment to 224.5 in the conflict treatment (RS:  $p = 0.151$ ). Figure 2 shows that the average cost of conflict (24.9) is higher than the average benefit from higher higher effort (6.3) resulting in a negative effect of conflict on efficiency.

Result 8 indicates that the code of conduct fails to have a positive impact on the outcomes. To better understand why the code of conduct doesn't have an effect on effort provision, we analyze the participants' decisions in the different stages of the game in more detail.

**Result 9** (Agreement - Acceptance of agreement). *In both the conflict and no conflict treatment, nearly all participants accept the agreement.*

A total of 95.1% of the participants accept the agreement in the no conflict treatment and 100% in the conflict treatment. There is no time trend in the acceptance rate in either treatment.

**Result 10** (Agreement - Conflict Pattern). *Signing a code of conduct does not coordinate workers on the pay-for-output norm. Norm enforcement is rather weak in the agreement experiment in general, but among the workers who engage in norm enforcement, the clear majority follows the pay-for-input norm.*

Despite the fact that all workers who interact with a principal sign a code of conduct that instructs them to follow the pay-for-output norm, the conflict pattern looks almost identical to the one observed in the baseline experiment: conflict initiation is mostly determined by whether the Principal pays a bonus or not. The third row of Figure 3 presents the relative frequency of conflict initiation in the agreement experiment contingent on the worker's effort

level, the worker’s subjective performance signal and the Principal’s bonus payment decision. The Worker’s subjective signal has only a very small impact. If the Worker exerts high effort, observes a positive signal, but gets no bonus, conflicts are initiated in 54.7% of the cases. If everything remains constant, but the Principal decides to pay a bonus, the conflict rate decreases to 6.2% (a decrease of 88.7%, SR:  $p = 0.063$ ). However, if we compare the initial situation to the same situation except that the workers now observe a negative signal the conflict rate goes down only to 43.3% (a decrease of 21%, SR:  $p = 0.188$ ). This pattern is consistent with the pay-for-input norm.<sup>10</sup>

The punishment strategies of individual workers displayed in Figure 4 further reinforces that the code of conduct doesn’t affect the decisions of the parties. The population is divided into non enforcers (50.0%) and norm enforcers (26.1%). The share of norm enforcers in this experiment is nearly half the size of what we observe in the baseline experiment (42.5%) or in the communication experiment (44.4%). Among the non enforcers, roughly two thirds exert low effort (blue dots - 32.6%) and the rest exert high effort (black dots - 17.4%). The share of self-interested non enforcers is the same as the share of pay-for-input norm followers (orange dots - 17.4%), while only 8.7% of the workers subscribe to the pay-for-output norm (green dots). Moreover, 23.9% of the workers cannot be defined by any of our pre-defined norms (red dots) giving further support to the claim that the code of conduct fails to coordinate workers.

Despite its failure to coordinate workers (and the resulting erratic punishment pattern), we observe that the code of conduct helps to coordinate bonus payments of principals.

**Result 11** (Agreement - Bonus Payments). *As in the baseline and the communication experiments, principals pay bonuses more often in the conflict treatment than in the no conflict treatment. Moreover, signing the code of conduct induces more principal to follow the pay-for-output norm, that is, to condition their bonus payments on their subjective performance signal.*

At the average level, principals pay bonuses more often in the conflict treatment than in the no conflict treatment. Overall, the frequency with which principals pay the bonus nearly doubles, from 23.5% in the no conflict treatment to 42.8% in the conflict treatment (RS:  $p = 0.008$ ). Furthermore, in this experiment, the signal becomes relevant in the decision to pay a bonus. The third row of Figure 5 presents the bonus payments in the agreement experiment contingent on the Principal’s private signal in both the conflict and no conflict treatments. The bonus patterns shown in this figure indicate that the principals pay attention

---

<sup>10</sup>The pattern is similar for low effort choices. Low effort, no bonus and a positive signal yield a conflict rate of 27.4%. A bonus payment lowers the conflict rate to 5.1% ( $p = 0.126$ ), but a positive signal only reduces it to 22.1% ( $p = 0.844$ ).

to their own subjective signal when deciding about the bonus payment. In the conflict treatment, the bonus rate is 20.0% if the Principal’s subjective signal is negative and 65.8% if the signal is positive (an increase of 69.6% SR:  $p = 0.126$ ). The corresponding numbers in the no conflict treatment are 8.2% and 39.2% (an increase of 79.1% SR:  $p = 0.062$ ).

At the individual level, signing the code of conduct induces more principals to follow the pay-for-output norm. The bottom left panel in figure 6 shows the individual patterns. The figure reveals that 77.6% (green dots) of the principals in the agreement experiment subscribe to the pay-for-output norm, compared to 55.8% in the baseline experiment and 44.4% in the communication experiment. However, a careful examination of the figure, shows that the green dots representing the pay-for-output principals are concentrated between the 0.50-0.75 range of the vertical axis (probability of bonus payment after a high signal). Hence, although the tendency is to pay the bonus after a positive signal, principals do not do this often enough to create financial incentives for workers to exert high effort: the worker’s expected payoff from exerting low effort is still higher (17) than the one from exerting high effort (15.5).

Our third experiment shows that a code of conduct alone does not successfully coordinate the trading parties on the pay-for-output norm. Although nearly all participants accept the agreement, only the principals follow the pay-for-output norm more frequently. The workers, in contrast, mostly choose to either abstain from conflict initiation or to enforce the pay-for-input norm.

#### **Experiment 4: Grievance.**

**Design.** Our third experiment explored the idea that organizations might channel the disciplining effect of conflicts by employing a code of conduct to explicitly define the situations for which conflicts are appropriate. In our final experiment we take this idea one step further and study the impact of an institutionalization of conflicts. Instead of simply providing behavioral guidelines, organizations can also design explicit procedures that allow their members to file formal complaints against each other if the organization’s code of conduct has been violated. The design of the experiment is identical to the agreement experiment, except for one important detail: instead of simply initiating conflicts directly, workers need to follow a formal grievance procedure if they want to sanction their principal. The grievance procedures requires the filing of a standardized report in which the Worker needs to explicitly confirm that no bonus has been paid although the Worker has received a positive subjective signal. Once the report is submitted the punishment of the Principal is initiated with the exact same consequences as in the previous experiments. From an experimental point of view the institutionalization of the conflict corresponds to a pure re-framing of the sanctioning decision. However, the fact that the Worker needs to file a formal complaint

implies that the he needs to lie in order to trigger a sanction in a situation in which the code of conduct does not deem a conflict appropriate. If workers are averse to lying, such a grievance procedure might help to coordinate workers on the pay-for-output norm (see, e.g., Gneezy, 2005; Lundquist et al., 2009; Gneezy et al., 2013).

**Data Collection and Procedural Details.** The laboratory, the subject pool, the recruitment process and the software used to program and run the experiment were the same as in the previous experiments. We ran 12 sessions with a total of 272 subjects and we conducted them in May 2017. We aimed at 24 subjects per session but some sessions were smaller due to no-shows. Two sessions had 18 subjects only and the other session had either 20 or 22 subjects. Sessions lasted for 50 to 80 minutes including the reading of the instructions and the final cash payments. Subjects received a show-up fee of 10 CHF and experimental points were converted at a rate of 70 points per CHF. Average total earnings were 35.6 CHF (33.8 CHF for subjects in the role of worker, and 37.3 CHF for subjects in the role of employer).

**Results.** We first explore how the introduction of a grievance procedure in addition to the code of conduct shapes the impact of conflicts on effort and surplus.

**Result 12** (Grievance - Effort and Surplus). *In the presence of a grievance procedure, the opportunity to engage in conflicts has a significantly positive effect on workers' effort and leads to higher total surplus.*

The fourth row of Figure 1 displays the relative frequency of high effort and average total surplus in the grievance experiment. The figure shows that the rate with which workers choose high effort rises from 44.1% in the no conflict treatment to 64.4% in the conflict treatment. This corresponds to a statistically significant increase of 46% (RS:  $p = 0.026$ ). The increase in effort in the conflict treatment also induces an increase in total surplus from 237.3 in the no conflict treatment to 243.7 in the conflict treatment. However, the increase in surplus effect is not statistically significant (RS: one-sided  $p = 0.844$ ).<sup>11</sup>

An analysis of the dynamics reveals that institutionalized conflicts help sustain a higher effort level in all phases of the experiment. Table 9 presents the relative frequency of high effort and average total surplus over time. In the conflict condition, effort decreases only slightly over time, from 71.1% in the first five periods to 61.3% in the last five periods (a decrease of 13.8%, SR:  $p = 0.063$ ). In the no conflict condition, in contrast, effort clearly declines over time, from 57.1% in periods 1-5 to 35.6% in periods 11-15 in the no conflict treatment (a decrease of 35.9%, SR:  $p = 0.031$ ). Average surplus shows similar time patterns.

---

<sup>11</sup>See the bottom right panel of Figure 2 for an illustration of the total cost of conflict (20.2) in comparison with the benefits of additional effort (26.2).

TABLE 9. Grievance - Development of effort and surplus over time

Periods	High Effort (in %)			Average Surplus		
	1-5	6-10	11-15	1-5	6-10	11-15
Conflict	71.10	60.86	61.32	253.50	237.26	240.54
No Conflict	57.10	39.51	35.60	252.93	237.41	221.36
p-values (RS)	0.180	0.040	0.020	1.00	0.820	0.180

**Result 13** (Grievance - Acceptance of agreement). *In the no conflict and conflict treatments nearly all participants accept the agreement.*

Again, most of the participants accept the agreement - 98.1% in the no conflict treatment and 99.5% in the conflict treatment. In both treatments, acceptance rates remain constant overtime.

**Result 14** (Grievance - Conflict Pattern). *workers in the conflict treatment of the grievance experiment are more likely to follow the pay-for-output norm than the pay-for-input norm. As a consequence, the worker’s subjective performance signal has an impact on the conflict rate. However, also in the grievance experiment there is a substantial share of workers who refrain from initiating conflict and do not enforce any norm.*

The fourth row of Figure 3 presents the relative frequency of conflict initiation in the grievance experiment contingent on worker’s effort level, the worker’s subjective performance signal and the Principal’s bonus payment. We observe that, as in all previous experiments, conflicts are most often initiated when effort is high, the worker’s signal is positive and no bonus is paid. In this case the conflict rate amounts to 61.2%. If all remains equal except that the Principal pays the bonus, the conflict rate decreases substantially to 1.9% (a decrease of 98%, SR:  $p = 0.032$ ). However, whereas the impact of the bonus payment is comparable to what we observed in previous experiments, we now also observe that the worker’s subjective signal has a decisive impact. If we replace the positive signal in the initial situation with a negative one, the conflict rate drops to 35.6% (a decrease of 42%, SR:  $p = 0.032$ ).<sup>12</sup>

At the individual level, we observe that while institutionalized conflicts in the grievance experiment help to coordinate workers on the pay-for-output norm, it is not the case that all workers respect the code of conduct. The individual conflict patterns illustrated in the bottom-right panel of Figure 4 shows that overall 42.7% of the workers engage in norm enforcement. Two thirds of the norm enforcers follow the pay-for-output norm (27.9%), whereas the rest (14.8%) enforces the pay-for-input norm. Moreover, about a third of the worker’s

<sup>12</sup>The pattern is similar for low effort choices. Low effort, no bonus and a positive signal yield a conflict rate of 40.5%. A bonus payment lowers the conflict rate to 1.1% ( $p = 0.062$ ) and a positive signal reduces it to 15.3% ( $p = 0.094$ ).

population (34.5%) abstains from initiating conflicts. Table 10 presents the development of the conflict rate of workers who do not receive a bonus over time.

TABLE 10. Grievance - Development of conflict rate over time (only if no bonus received)

Effort Choice Periods	Conflict Rate (in %)					
	High Effort			Low Effort		
	1-5	6-10	11-15	1-5	6-10	11-15
Positive Signal	66.67	60.00	57.50	30.30	32.08	48.78
Negative Signal	40.00	44.44	37.04	22.50	13.73	11.86
p-values (SR)	0.040	0.040	0.460	0.860	0.160	0.320

**Result 15** (Grievance - Bonus Payments). *The structure of principals' bonus payments in the grievance treatment is similar to the one in the agreement treatment. Principals pay bonuses more often in the conflict treatment than in the no conflict treatment. Moreover, many principals follow the pay-for-output norm and condition their bonus payments on their subjective performance signal.*

The fourth row of Figure 5 shows the frequency of bonus payments in the grievance experiment as a function of the private signal of the Principal in both treatments. In this experiment bonus payments in the conflict treatment are highly contingent on the signal observed. The bonus rate is 79.3% if the Principal's subjective signal is positive and 20.2% if the signal is negative (SR:  $p = 0.032$ ). A similar pattern (although at a lower level) is observed in the no conflict treatment where the corresponding numbers are 36.1% and 4.5% (SR:  $p = 0.032$ ).

At the individual level, the bottom-right panel of Figure 6 adds supportive evidence to the claim that most principals in the grievance treatment condition their bonus payments on their subjective signal: 84.4% of the principals are classified as pay-for-output norm followers (compared to 55.1% in the baseline) and only 10.2% of principals follow the pay-for-input norm. These bonus patterns finally create (weak) monetary incentives for the Worker to exert high effort as the expected payoffs of exerting high effort are 20.1 and the gains from exerting low effort are 19.7.

In terms of bonus payment dynamics, table 5 show bonus payments over time contingent on the signal observed by the Principal and conflict. The bonus payment patterns remain very stable over time.

Our final experiment shows that conflicts can have beneficial effects on motivation (and efficiency) even in a challenging environment like our subjective evaluation game. At the same time, the experiment also demonstrates that the establishment of a conflict culture is a delicate and somewhat risky endeavor. Despite the establishment of a clear code of conduct

TABLE 11. Formal agreement with grievance - Development of bonus payments over time

Principal's Signal Periods	Bonus Payments (in %)					
	Positive Signal			Negative Signal		
	1-5	6-10	11-15	1-5	6-10	11-15
Conflict	81.58	78.38	77.90	18.59	20.61	21.38
No Conflict	42.26	37.95	27.03	7.05	4.43	2.29
p-values (RS)	0.00	0.00	0.00	0.01	0.03	0.00

and grievance procedure that forces workers to lie if they wish to deviate from the code of conduct, not all workers follow the code of conduct and conflicts still emerge in situations in which they have no beneficial effects. These findings suggests that conflict cultures need to be carefully managed in order to be successful.

#### 4. MEASURING NORMS AND CULTURES

The relationship between action and payoffs is unchanged across our different experiments, but we find that behavior is sensitive to variations in contextual elements. Such a pattern is incompatible with models that assume fixed preferences over distributions of payoffs, but is consistent with the idea that people care about being behaviorally consistent with context-driven and endogenously selected norms. Our results illustrate that—if a group (or an organization) manages to successfully set up a set of interlocking norms—conflicts may have a disciplining effect on principals and may ultimately also lead to an increase in workers' *effort*. We call such a set of interlocking norms a “culture”.<sup>13</sup>

From an empirical point of view, our approach raises the question of how to best measure the extent to which behavior is consistent with norm-abidance. In this section we provide a framework—using off-the-shelf econometric packages—that allows us to measure norm and cultural compliance in a simple manner. We build upon the fact that any strategic situation can be represented as an extensive form. We know that the end nodes of the decision tree provide a complete representation of the game. More precisely, let  $Z$  be the end nodes of an interaction between a Principal, a Worker and Nature (in line with the tradition in game theory, Nature is represented as a player with fixed mixed strategies). The outcome of an interaction induces a probability distribution over  $Z$ . A set of norms or culture effectively says that certain behaviors imply that some actions should never be taken, which in turn implies that there is a  $Z^{0K} \subset Z$ , where  $K$  refers to culture  $K \in \{SI, PI, PO\}$ , such that  $Pr [z \in Z^{0K}] = 0$ . For example, under culture *SI* (*self-interest choice*), effort should be zero

<sup>13</sup>This definition is consistent with a literature that views culture through the lens of game theory. Kreps (1990) and Greif (1994) view culture as a Nash equilibrium to an employment game. MacLeod and Malcolmson (1989, 1998) show that in market contexts multiple self-enforcing “social norm” equilibria may exist.

( $e = 0$ ). Thus, if we observe many pairs with positive effort ( $e = 1$ ), we get  $Pr [Z^{0SI}] > 0$ , which implies that behavior is not well represented by self-interest.

However, our aim in this paper is obviously not to reject the self-interest model. Such findings are neither new, nor particularly useful. Rather, the question we want to ask is: which culture best fits our data? It might be the case that even though mistakes are frequent, rational choice still provides the best description of behavior (a common claim). The goal of this section is to provide a method to measure the descriptive quality of any behavioral norm. Second, we use the term “culture” to mean a set of *interlocking norms*, so that we can also ask whether the population as a whole can be characterized by a common culture.

Our approach is inspired by Selten (1975)’s notion of a perfect equilibrium. He envisions players who at each node may make mistakes (trembling hand). In our experiment the game can be represented as a binary tree, and hence at each decision node choice can be viewed as a binomial distribution which can be represented using the logit model. For each context we can measure the distance between observed choices and a norm. Norm compliance is less general than a “culture”, which we see as a collection of norms that work together.<sup>14</sup> After discussing norm compliance by the Principal and Worker, we measure the extent to which the population (organization) as a whole can be characterized by a single culture.

**The Principal’s Norms.** Let us begin by measuring the Principal’s adherence to a norm. Figure 7 illustrates the Principal’s choices parameterized for the case of the pay-for-output norm (PO). The Principal cannot observe whether or not the Worker has produced effort, and hence the Principal does not know the probability of a high signal. The choice of a high signal is an action chosen by Nature that we can estimate as a binomial distribution characterized by  $\gamma_H$  that determines the probability of a high signal via:

$$(4.1) \quad Prob [s_P = 1] = \frac{\exp(\gamma_1)}{\exp(\gamma_1) + \exp(-\gamma_1)}.$$

In the figure we use the terminology  $P \sim \exp(\gamma_1)$  to indicate that the probability varies with  $\gamma_1$  modulo a factor to normalize to a probability as we have done in (4.1) (see DeGroot, 1972, for details on this approach). Given the state, the Principal then chooses to pay a bonus or not. Notice that as  $\gamma_{PO} \rightarrow \infty$ , then

$$Prob [b = B | s_P = 1] = Prob [b = 0 | s_P = 0] \rightarrow 1.$$

These choices are exactly the ones indicated by the PO norm. The outcomes  $z \in Z_P = \{Z_1, \dots, Z_4\}$  are observed in the data. Since the realization of the signal is independent of

---

<sup>14</sup>On the notion that culture is based upon constituent parts see Homans (1961).



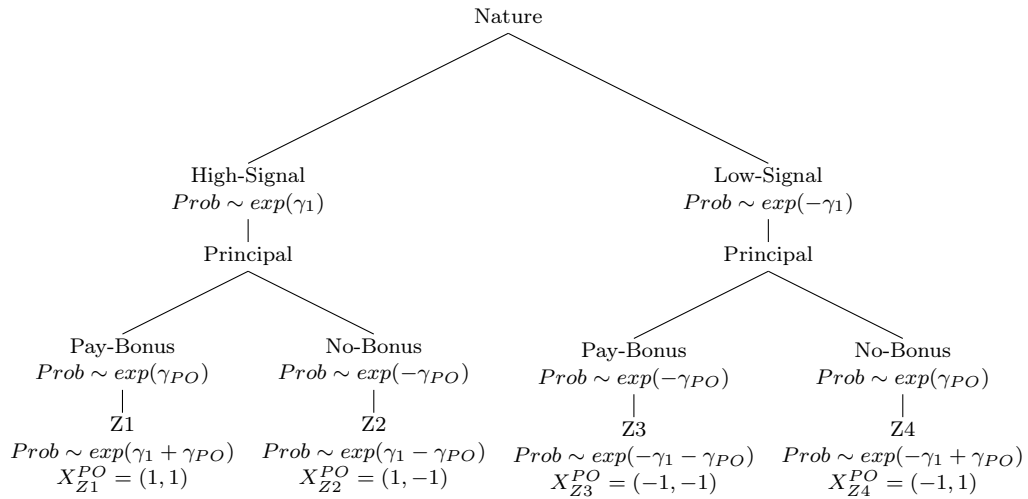
the Principal's choice, then the PO norm described in Figure (7) satisfies:

$$(4.2) \quad \text{Prob}[z] = \text{logit}(X_z^{PO} \vec{\gamma}_{PO}),$$

where  $X_z^{PO}$  is defined in the figure and:

$$\vec{\gamma}_{PO} = \begin{bmatrix} \gamma_1 \\ \gamma_{PO} \end{bmatrix}.$$

FIGURE 7. Parameterizing the Principal's Choices under *Pay-for-Output*



This model can be estimated from the data as a multinomial logit from which we obtain parameters  $\vec{\gamma}_{PO}^* = (\gamma_1^*, \gamma_{PO}^*)$ . If the Principal followed the PO norm perfectly, then  $\gamma_{PO}^* \rightarrow \infty$ . However, as we discussed above, we never get perfect compliance with any norm in any treatment, and hence these models are always well identified. A larger  $\gamma_{PO}^*$  indicates a behavior that is more closely approximated by the PO norm. In a similar way we can estimate  $\gamma_{PI}^*$  and  $\gamma_R^*$  for the *pay-for-input norm (PI)* and *self-interest choice (SI)* respectively.

Different norms correspond to different values for  $X_z$ , whose coefficients reflect the strength of a norm. The values of  $X_z$  for different norms are illustrated in Table (12).

TABLE 12. Empirical Norm Model for Principals

$Z_{Principal}$	Principal Signal	Bonus Paid?	$X_Z^{PO}$	$X_Z^{PI}$	$X_Z^{SI}$
Z1	High	Yes	(1, 1)	(1, 1)	(1, -1)
Z2	High	No	(1, -1)	(1, -1)	(1, 1)
Z3	Low	Yes	(-1, -1)	(-1, 1)	(-1, -1)
Z4	Low	No	(-1, 1)	(-1, -1)	(-1, 1)

Notice that SI calls for never paying a bonus, and hence is simply the opposite of PI, which implies  $\gamma_R^* = -\gamma_{PI}^*$ . Thus we do not need to run separate regressions for SI choice. The results from estimating model (4.2) for each treatment are given in Table 13.

TABLE 13. Estimated Adherence of Principal’s Behavior to a Norm

	No Conflict Treatments		Conflict Treatments	
	Pay-for-output ( $\gamma_{PO}^*$ )	Pay-for-input ( $\gamma_{PI}^*$ )	Pay-for-output ( $\gamma_{PO}^*$ )	Pay-for-input ( $\gamma_{PI}^*$ )
Baseline	0.250*** (0.037)	-0.894*** (0.051)	0.298*** (0.039)	-0.149*** (0.037)
Communication	0.288*** (0.041)	-0.628*** (0.047)	0.394*** (0.042)	0.401*** (0.042)
Code of Conduct	0.330*** (0.039)	-0.590*** (0.043)	0.496*** (0.042)	-0.145*** (0.038)
Grievance	0.332*** (0.034)	-0.687*** (0.040)	0.678*** (0.038)	0.043 (0.031)

Notes: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01, errors clustered by session.

The results confirm what we see visually in Figure 7. In the no conflict treatments the coefficient for the SI norm—as measured by  $\gamma_R^* = -\gamma_{PI}^*$ —is larger in value than the PO coefficient  $\gamma_{PO}^*$ . Negative coefficients mean that the Principal is more likely not to pay the bonus regardless of the signal. The fact that  $\gamma_{PO}^* > 0$  for the no conflict treatments reflects that effort is lower and hence the low signal is observed more often, so that there is a greater than 50% chance of the Principal observing the low signal, and then paying no bonus.

Things are very different in the conflict treatments. In the communication treatment the PI coefficient is indistinguishable from the PO coefficient. Since the PI coefficient is significantly positive, this implies that this norm dominates the SI, in contrast to the no conflict treatment where SI is a better fit. Hence the threat of conflict greatly increases the probability that the Principal pays the bonus.

In the communication treatment PI fits better than PO, but this result changes in the code of conduct and grievance cases. There the PO norm better describes the behavior of principals than either the PI norm or SI choice. Finally notice that the values of  $\gamma_{PO}^*$  for the code of conduct and grievance treatments in the no conflict treatment are comparable. Since these treatments are otherwise identical (same code of conduct), this finding shows that the “framing” of the conflict decision (grievance procedure) does not impact principal behavior a lot.

In addition to comparing coefficients, we can also compare the overall fit of the models. The logit reports the value of the likelihood function, which is a measure of the overall goodness of fit, and provides us with a metric to rank models. Since each model corresponds to a different set of cultural norms, this approach provides a way to measure how closely a population of individuals comply with cultural norms. We use the non-nested likelihood ratio test developed by Vuong (1989) to see if the overfits of the models differ significantly.

TABLE 14. Comparing Norms in Conflict Treatments

	Log Likelihood		Is a Norm Dominant at 1% significance levels?		
	PO	PI/SI	PO	PI	SI
Baseline (p-value)	<b>-867.7</b>	-1,010.16	<b>Yes</b> (0.004)	No (0.906)	No (0.906)
Communication (p-value)	-887	<b>-885.4</b>	No (0.544)	No (0.456)	No (0.456)
Code of Conduct (p-value)	<b>-916.6</b>	-986.5	<b>Yes</b> (0.000)	No (1.000)	No (1.000)
Grievance (p-value)	<b>-1,250.66</b>	-1,444.25	<b>Yes</b> (0.000)	No (1.000)	No (1.000)

*Notes:* P-values based upon non-nested Vuong where the null hypothesis is that the two norms, PO and PI are indistinguishable. The SI norm is the same model as the PI norm, but with a change in the sign of the coefficient.

In the no conflict treatments the  $R$  norm fits better than the  $PO$  norm in all treatments. These observations are confirmed by the Vuong test (not reported). However, matters are reversed with conflict. The log-likelihood values in Table 14 come from the logit estimates. These provide a measure of the quality of the fit (bigger is better).<sup>15</sup>

The PI norm and the SI norm are the same model but with  $\gamma_P^R = -\gamma_P^{PI}$ . For the treatments baseline, Code of Conduct and Grievance the null is rejected in favor of the PO norm. In the case of communication the null is not rejected, which implies that neither norm provides a superior fit to the data.

**The Worker’s Norms.** In the case of the Worker there are more information sets—the Worker observes both his signal and the bonus choice. Moreover, the probability of a bonus is likely to be correlated with the signal, though the signal can be assumed to be exogenous as long as we condition on worker effort. Hence, in this case we have  $\vec{\gamma}_{norm}^* = [\gamma_1^*, \gamma_{B0}^*, \gamma_{B1}^*, \gamma_{norm}^*]^T$ , where  $\gamma_1^*$  measures the probability of a high signal,  $\gamma_{B0}^*$  measures the probability of a bonus given a low signal,  $\gamma_{B1}^*$  measures the probability of a bonus given a high signal, and, finally,  $\gamma_{norm}^*$  measures the probability of the worker following the  $norm \in \{PO, PI, R\}$ . We run a multinomial logit model with standard errors clustered at the session level:

$$Prob[z] = \text{logit}(X_z \vec{\gamma}),$$

where  $z \in Z_{Worker}$  and  $X_z$  are defined in Table 15.

Workers know the effort level they have chosen, and so the question we ask is what norm does a worker who has high or low effort adopt. The results of the estimation are shown in Table 16. In the low effort case the workers are shirking, and hence presumably do not expect a bonus for their efforts. In this case, except for the Communication treatment, SI

<sup>15</sup>The Vuong test provides a way to compare two non-nested models. The null hypothesis is that the models are indistinguishable. The null is rejected when one model fits better than the other.

TABLE 15. Empirical Norm Model for Workers

$Z_{Worker}$	Worker Signal	Bonus Paid?	Principal Punished?	$X_Z^{PO}$	$X_Z^{PI}$	$X_Z^{SI}$
Z1	High	Yes	No	(1, 0, 1, 1)	(1, 0, 1, 1)	(1, 0, 1, 1)
Z2	High	No	No	(1, 0, -1, -1)	(1, 0, -1, -1)	(1, 0, -1, 1)
Z3	Low	Yes	No	(-1, 1, 0, 1)	(-1, 1, 0, 1)	(-1, 1, 0, 1)
Z4	Low	No	No	(-1, -1, 0, 1)	(-1, -1, 0, -1)	(-1, -1, 0, 1)
Z5	High	Yes	Yes	(1, 0, 1, -1)	(1, 0, 1, -1)	(1, 0, 1, -1)
Z6	High	No	Yes	(1, 0, -1, 1)	(1, 0, -1, 1)	(1, 0, -1, -1)
Z7	Low	Yes	Yes	(-1, 1, 0, -1)	(-1, 1, 0, -1)	(-1, 1, 0, -1)
Z8	Low	No	Yes	(-1, -1, 0, -1)	(-1, -1, 0, 1)	(-1, -1, 0, -1)

choice does fit better. However, it is not perfect, implying that there are also individuals who choose low effort and pay the cost of punishing the Principal who does not pay a bonus. Since effort is zero, this does not fit with a theory of reciprocity, but does suggest that there are individuals who shirk and get utility from punishing others who do not pay them. This effect is particularly prominent in the communication treatment.

In the case of high effort in the baseline experiment all three norms have similar coefficients. With communication we see that the PI norm has the largest weight. The PO norm provides the best fit for both the code of conduct treatment and the grievance treatment, with the grievance treatment having a particularly large effect, going from 0.542 to 0.811. This result highlights the fact that Principals and Workers have different concerns. Above we saw that the code of conduct treatment did push the principals towards the PO norm, but here we see that the effect on workers is more muted. Adding a formal grievance procedure provided for many individuals a mechanism that appears to legitimize imposing a cost upon a principal.

For the communication treatment the PI norm provides the best fit to the data, suggesting that in the absence of other coordinating information individuals appear to focus upon reciprocal norms, even though they are not self-enforcing.

TABLE 16. Estimated Adherence of Behavior to a Worker’s Norm

	Low Effort			High Effort		
	SI	PI	PO	SI	PI	PO
Baseline	0.780*** (0.067)	-0.057 (0.051)	0.466*** (0.057)	0.336*** (0.057)	0.521*** (0.061)	0.536*** (0.061)
Communication	0.569*** (0.070)	0.639*** (0.072)	0.608*** (0.071)	0.617*** (0.060)	0.890*** (0.072)	0.709*** (0.064)
Code of Conduct	0.793*** (0.072)	-0.038 (0.054)	0.543*** (0.062)	0.481*** (0.058)	0.461*** (0.057)	0.542*** (0.059)
Grievance	0.876*** (0.073)	0.070 (0.052)	0.740*** (0.067)	0.682*** (0.048)	0.595*** (0.046)	0.811*** (0.052)
Observations	1,588	1,588	1,588	2,319	2,319	2,319
Log Likelihood	-3,864.113	-4,150.913	-3,971.681	-5,837.546	-5,790.553	-5,691.499

Notes: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01, clustered by session

Next we assess norm compliance using the Vuong test. In this case we have three models, rather than two as in the Principal case. The Vuong test allows for only two models. We

deal with this by comparing models to the top ranked model (and hence the top ranked model gets compared to the second ranked, and not the third). In the zero effort case SI choice is the dominant norm, and so again we do not report those results here. The effects of the treatments are reported in Table 17.

Under baseline the best fit is PO, while the second best is PI. The next column reports the test of the null hypothesis that norms SI and PO are indistinguishable relative to norm SI. Given the p-value of 0.998 we cannot reject the null. Next norm PI is compared to the null hypothesis that PO and PI are indistinguishable and again we cannot reject the null. Finally, PO, the best fitting norm is compare to the null that SI and PO are indistinguishable and again the null is not rejected. Thus, in the absence of any coordination there does not seem to be convergence upon any norm.

TABLE 17. Comparing Norms in Conflict Treatments

	Log Likelihood			Is a Norm Dominant at 1% significance levels?		
	SI	PI	PO	SI	PI	PO
Baseline (p-value)	-949.0	-925.9	<b>-923.8</b>	No (0.998)	No (0.595)	No (0.405)
Communication (p-value)	-971.3	<b>-923.5</b>	-955.4	No (0.999)	<b>Yes</b> (0.005)	No (0.995)
Code of Conduct (p-value)	-998.7	-1001.6	<b>-989.5</b>	No (0.822)	No (0.897)	No (0.178)
Grievance (p-value)	-1,703.5	-1,729.1	<b>-1,664.8</b>	No (0.992)	No (1.000)	<b>Yes</b> (0.008)

*Notes:* P-values based upon non-nested Vuong test where the null hypothesis is that two norms are indistinguishable compared to a single norm. The comparison is always made to the best alternative model as measured by the likelihood value.

In the case of the communication treatment the *PI* norm provides the best fit (as measured by the log likelihood), followed by the PO norm. We reject the null hypothesis that PO and PI are indistinguishable relative to the PI norm at the 0.5% level. Both the PO norm and SI norms are rejected as potential best fits. However, even though there is communication, the results for the Principal show that they do not adopt PI. In this case there is a clear incentive for the workers to cheat, with the consequence that Principals often deviate from paying a bonus when they receive the low signal.

The interesting result is that under the code of conduct, in contrast to the results for the principal, no norm dominates. This reinforces the earlier observation that the code of conduct alone failed to coordinate workers. Finally, we find that in the case of the grievance treatment we can reject the hypothesis that PO and the second best norm, SI, are indistinguishable relative to PO alone. Though the fit is better the likelihood function is quite small relative to the other cases indicating that there is still quite a bit of noise in workers' choice.

**Culture.** In this section we apply the same methodology, except now to the full population. The number of end nodes for the full game, denoted by  $Z$ , is  $32 = 2^5$ . These are found by combining all the possible binary choices: effort ( $e \in \{0, 1\}$ ), Principal’s signal ( $s_P \in \{0, 1\}$ ), Agent’s signal ( $s_A \in \{0, 1\}$ ), bonus pay ( $b \in \{0, B\}$ ) and conflict ( $d \in \{0, 1\}$ ). We do not need to consider the effect of productivity given by  $r$  since it is not observed by parties. The effect it has upon the signals is estimated endogenously. Thus, we run a model of the form:

$$(4.3) \quad Prob[z] = \text{logit} \left( \gamma^{norm} X_z^{norm} + \sum_{E \in \{0,1\}} \gamma_{es_P} X_z^{es_P} + \sum_{E \in \{0,1\}, b \in \{0,B\}} \gamma_{es_A} X_z^{es_A} \right).$$

The first term estimates the weight,  $\gamma^{norm}$ , for the cultural norm. If all members of the population follow the norm perfectly then  $\gamma^{norm} \rightarrow \infty$ . Since perfection is not attainable, this value is always finite. Its size can be used to see how closely the population follows a particular norm. The remaining terms estimate the behavior of Nature in this relationship. Since the random draws are independent, these terms are unbiased estimates of the probabilities of the signals workers observe. Table 18 reports the estimate of  $\gamma^{norm}$  for each conflict treatment. In the absence of communication, the SI culture provides the best description. When there is communication the PI culture dominates, followed by the PO culture. For the next two treatments the PO culture provides the best description.

TABLE 18. Estimated Adherence of Behavior to a Work Culture

	Cultural Fits with Conflict		
	(Self-Interest)	(Pay-for-input)	(Pay-for-output)
Baseline	0.462*** (0.059)	0.391*** (0.047)	0.427*** (0.047)
Communication	-0.265** (0.074)	0.845** (0.044)	0.698*** (0.044)
Code of Conduct	0.438*** (0.047)	0.458*** (0.046)	0.578*** (0.045)
Grievance	0.240*** (0.044)	0.711*** (0.034)	0.889*** (0.034)

Notes: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

We can get a sense of the quality of the fit by formally testing one culture against the other using a non-nested hypothesis test, as we have done above. These results are reported in Table 19. In contrast to the case of norms, we have a clear winner for each treatment. It is worth highlighting the point that we are estimated the same model for each treatment. Notice that in the Grievance case the PO culture best describes the behavior relative to the other cultures, however the likelihood values are smaller than in the other cases, which indicates that there is a great deal of noise in individual behavior.

TABLE 19. Comparing Cultures in Conflict Treatments

	Log Likelihood			Is a Culture Dominant at 1% significance levels?		
	SI	PI	PO	SI	PI	PO
Baseline	<b>-2,346.5</b>	-2,364.8	-2,358.2	Yes	No	No
(p-value)				(0.1)	(0.90)	(0.78)
Communication	-2,186.4	<b>-2,023.7</b>	-2,082.4	No	Yes	No
(p-value)				(1.00)	(0.00)	(1.00)
Code of Conduct	-2,264.0	-2,259.4	<b>-2,229.2</b>	No	No	Yes
				(0.98)	(1.00)	(0.02)
Grievance	-3,384.3	-3,208.1	<b>-3,086.5</b>	No	No	Yes
				(1.00)	(1.00)	(0.00)

Notes: P-values based upon non-nested Vuong test where the null hypothesis is that two norms are indistinguishable compared to a single norm. The comparison is always made to the best alternative model as measured by the likelihood value or p-value.

## 5. DISCUSSION

Our experiment shows that in addition to affecting the allocation of resources, potential conflict can also increase overall performance. Potential conflict provides a self-help mechanism to implement agreements. This general point has been made many times within the context of the repeated prisoner’s dilemma where conflict is represented by playing “defect” for several periods. In those games defection is perfectly observable, so that it is clear when breach of agreement has occurred. In our experiment, in contrast, a Worker who receives a good self evaluation of performance is not sure whether or not the Principal has received a similar signal. This creates ambiguity regarding what is “fair” behavior, and what is a “breach” of an agreement for which the Worker should impose a cost upon the Principal.

To analyze the stability and effectiveness of different plausible fairness norms in our setting, we introduce the notion of a “norm equilibrium”. We show that in our principal-worker relationship with potential conflict there are two norm equilibria, self-interested play and pay-for-output. The pay-for-output equilibrium provides higher payoffs than the rational play equilibrium, and hence it is a reasonable hypothesis that either party would play this equilibrium, and—if permitted to communicate—would recommend this equilibrium.<sup>16</sup>

It turns out that this hypothesis receives little support. In the baseline experiment *self-interest* provides the best (but imperfect) fit to the data, whereas in the communication experiment parties tend to coordinate upon pay-for-input. However, norm compliance is very limited, and workers in the communication treatment still choose low effort in more than 40% of the pairs. We do know that context matters. For example, Cohn et al. (2014) find that bank employees are honest or not depending upon the work culture indicating that honesty is not a person specific immutable trait. After having observed the surprising results of our first two experiments, we therefore reasoned that the parties needed more nudging

<sup>16</sup>The idea that parties would play the most efficient Nash equilibrium goes back to Schelling (1980).

towards the pay-for-output equilibrium. In our third experiment parties enter into an explicit code of conduct agreement. This additional element does move the culture towards pay-for-output to some extent (mostly on the side of principals), but average effort does not rise. The final experiment formalizes employment expectations and makes violations of the pay-for-output norm a “breach” event that gives workers’ the right to use a grievance procedure. This results in a significant increase in effort, and suggests that contracts, and formal procedures in an organization have a role to play in improving performance.

Our analysis highlights important challenges in the application of the principal-agent approach in practice.<sup>17</sup> Whereas principal-agent theory focuses upon the impact of information quality on contract design, the aspect of the theory that has been emphasized most in practice is the need to link compensation to performance.<sup>18</sup> We explore the arguably realistic case in which subjectivity in performance evaluations complicates using performance pay. We show that another important ingredient to high performance entails providing formal power and voice to employees. Firms are often reluctant to do that because of the fear that it will result in rent seeking and lower performance. Our results do not dispute that observation, but suggest that with appropriate design, providing worker with more power in an organization can lead to enhanced performance. However, it is a delicate balance. Our framework provides a way to explore this question experimentally, and provides a way to quantify “culture” in a simple manner.

The term “culture” has many different meanings (Alesina and Giuliano, 2015). Our experimental setting allows for a crisp definition of “culture” as a collection of interacting norms of behavior broadly consistent with the notions of culture as used by Kreps (1990) and Greif (1994). One of the implications of our analysis is that even though the material payoffs remain fixed across our experiments, we nevertheless observe a great deal of variation in behavior and performance. This may help explain some of the heterogeneity in organizational performance, even for firms in the same industry (Syverson, 2011; Bloom et al., 2012). Obviously, this paper is only the beginning of a research agenda. We need much more work exploring different parameter values and different organizational reforms to eventually better understand the full role of culture in organizational performance.

---

<sup>17</sup>See Gibbons (1997) and Eisenhardt (1989) for thought assessments of the strengths and weaknesses of agency theory for understanding organizations.

<sup>18</sup>This oversimplification has created many examples of dysfunctional employment relationship (Kerr, 1975; Hall, 2000).



## REFERENCES

- Akerlof, G. A. (1980). A theory of social custom of which unemployment May be one consequence. *Quarterly Journal of Economics* 94, 749–775. 1, 2
- Akerlof, G. A. and R. Kranton (2010). Identity economics. *The Economists' Voice* 7. 1, 2
- Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *Quarterly Journal of Economics* 115(3), 715–753. 1, 2
- Alesina, A. and P. Giuliano (2015). Culture and institutions. *Journal of Economic Literature* 53(4), 898–944. 5
- Ambrus, A. and B. Greiner (2012). Imperfect public monitoring with costly punishment: An experimental study. *American Economic Review* 102(7), 3317–32. 1
- Axelrod, R. (1981). The emergence of cooperation among egoists. *American Journal of Political Science* 75(2), 306–318. 2
- Binmore, K., J. Swierzbinski, S. Hsu, and C. Proulx (1993). Focal points and bargaining. *International Journal of Game Theory* 22(4), 381–409. 1
- Bloom, N., C. Genakos, R. Sadun, and J. V. Reenen (2012). Management practices across firms and countries. NBER Working Papers 17850, NBER. 5
- Brandts, J. and D. J. Cooper (2007). It's what you say, not what you pay: An experimental study of manager-employee relationships in overcoming coordination failure. *Journal of the European Economic Association* 5(6), 1223–1268. 3, 9
- Brandts, J., M. Ellman, and G. Charness (2015). Let's talk: How communication affects contract design. *Journal of the European Economic Association* 14(4), 943–974. 9
- Brandts, J. and W. B. MacLeod (1995). On the strategic stability of equilibria in experimental games. *Games and Economic Behavior* 11, 36–63. 3
- Cason, T. N. and V.-L. Mui (2007). Communication and coordination in the laboratory collective resistance game. *Experimental Economics* 10(3), 251–267. 3
- Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics* 14(1), 47–83. 1
- Cohn, A., E. Fehr, and M. A. Marechal (2014). Business culture and dishonesty in the banking industry. *Nature* 516(7529), 86–U190. 5
- Coleman, P. T., M. Deutsch, and E. C. Marcus (2014). *The handbook of conflict resolution: Theory and practice*. John Wiley Sons. 1
- Cooper, D. and J. Kagel (2016). Other regarding preferences: A selective survey of experimental results. *The handbook of experimental economics* 2, 217–289. 4, 2
- Cooper, R., D. V. DeJong, R. Forsythe, and T. W. Ross (1992). Communication in coordination games. *The Quarterly Journal of Economics* 107(2), 739–771. 3

- De Dreu, C. K., M. J. Gelfand, et al. (2008). *The psychology of conflict and conflict management in organizations*. Lawrence Erlbaum Associates New York. 1
- DeGroot, M. H. (1972). *Optimal Statistical Decisions*. NY: McGraw-Hill Book C. 4
- Dreber, A., D. G. Rand, D. Fudenberg, and M. A. Nowak (2008). Winners don't punish. *Nature* 452(7185), 348. 1
- Egas, M. and A. Riedl (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society* 275(1637), 871–878. 1
- Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *The Academy of Management Review* 14(1), 57–74. 17
- Falk, A., E. Fehr, and U. Fischbacher (2003). On the nature of fair behavior. *Economic Inquiry* 41(1), 20–26. 2
- Fehr, E. and U. Fischbacher (2003). The nature of human altruism. *Nature* 425(6960). 1
- Fehr, E. and U. Fischbacher (2004). Social norms and human cooperation. *Trends in cognitive sciences* 8(4), 185–190. 1
- Fehr, E., O. Hart, and C. Zehnder (2011). Contracts as reference points: Experimental evidence. *American Economic Review* 101, 493–525. 1
- Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114(3), 817–68. 1, 2
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10(2), 171–178. 3
- Gächter, S. and E. Fehr (2000). Cooperation and Punishment in Public Goods Experiments. *American Economic Review* 90(4), 980–994. 1
- Gächter, S., E. Renner, and M. Sefton (2008). The long-run benefits of punishment. *Science* 322(5907), 1510–1510. 1
- Gibbons, R. (1997). Incentives and careers in organizations. In D. M. Kreps and K. F. Wallis (Eds.), *Advances in Economics and Econometrics: Theory and Applications*, pp. 1–37. Cambridge, UK: Cambridge University Press. 17
- Gneezy, U. (2005). Deception: The role of consequences. *The American Economic Review* 95(1), 384–394. 3
- Gneezy, U., B. Rockenbach, and M. Serra-Garcia (2013). Measuring lying aversion. *Journal of Economic Behavior and Organization* 93, 293–300. 3
- Grechenig, K., A. Nicklisch, and C. Thöni (2010). Punishment despite reasonable doubt—a public goods experiment with sanctions under uncertainty. *Journal of Empirical Legal Studies* 7(4), 847–867. 1
- Greif, A. (1994). Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies. *Journal of Political Economy* 102(5),

- 912–950. 13, 5
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association* 1(1), 114–125. 3
- Güth, W., R. Schmittberger, and B. Schwarze (1982). An experimental-analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization* 3(4), 367–388. 1
- Hall, B. (2000). Compensation and performance evaluation at arrow electronics. Technical Report 9-800-290, Harvard Business School. 18
- Homans, G. C. (1961). *Social Behavior Its Elementary Forms*. New York, NY: Harcourt, Brace & World, Inc. 14
- Kagel, J. H., C. Kim, and D. Moser (1996). Fairness in ultimatum games with asymmetric information and asymmetric payoffs. *Games and Economic Behavior* 13(1), 100–110. 1
- Kahneman, D., J. L. Knetsch, and R. H. Thaler (1986). Fairness and the assumptions of economics. *Journal of business* 59(4), S285–S300. 1
- Kerr, S. (1975). On the folly of rewarding A, while hoping for B. *Academy of Management Journal* 18(4), 769–783. 18
- Koszegi, B. (2014). Behavioral contract theory. *Journal of Economic Literature* 52(4), 1075–1118. 1
- Kreps, D. M. (1990). Corporate culture and economic theory. In J. E. Alt and K. A. Shepsle (Eds.), *Perspectives on Positive Political Economy*, pp. 90–143. Cambridge, U.K.: Cambridge University Press. 13, 5
- Krupka, E. L., S. Leider, and M. Jiang (2017). A meeting of the minds: informal agreements and social norms. *Management Science* 63(6), 1708–1729. 1
- Lazear, E. P. (1986). Salaries and piece rates. *Journal of Business* 59, 405–431. 3, 2
- Leibbrandt, A. and R. López-Pérez (2012). An exploration of third and second party punishment in ten simple games. *Journal of Economic Behavior & Organization* 84(3), 753 – 766. 1
- Lundquist, T., T. Ellingsen, E. Gribbe, and M. Johannesson (2009). The aversion to lying. *Journal of Economic Behavior & Organization* 70(1–2), 81–92. 3
- MacLeod, W. B. (2003). Optimal contracting with subjective evaluation. *American Economic Review* 93(1), 216–240. 2
- MacLeod, W. B. and J. M. Malcomson (1989). Implicit contracts, incentive compatibility, and involuntary unemployment. *Econometrica* 57(2), 447–480. 13
- MacLeod, W. B. and J. M. Malcomson (1998). Motivation and markets. *American Economic Review* 88(3), 388–411. 2, 13
- McAdams, R. H. (2017). *The Expressive Powers of Law*. Cambridge, MA: Harvard University Press. 1

- Ostrom, E., J. Walker, and R. Gardner (1992). Covenants with and without a sword: Self-governance is possible. *American political science Review* 86(2), 404–417. 1
- Prasnikar, V. and A. E. Roth (1992). Considerations of fairness and strategy: Experimental data from sequential games. *The Quarterly Journal of Economics* 107(3), 865–888. 1
- Rubinstein, A. (1998). *Modeling bounded rationality*. MIT press. 2
- Schelling, T. C. (1980). *The Strategy of Conflict*. Harvard University Press. 16
- Selten, R. (1975). Re-Examination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory* 4, 25–55. 4
- Simon, H. A. (1982). *Models of Bounded Rationality*. MIT Press. 5
- Simon, H. A. (1997). *Models of bounded rationality: Empirically grounded economic reason*, Volume 3. MIT press. 2
- Syverson, C. (2011). What determines productivity?. *Journal of Economic Literature* 49(2), 326–365. 5
- Tjosvold, D., A. S. Wong, and N. Y. Feng Chen (2014). Constructively managing conflicts in organizations. *Annu. Rev. Organ. Psychol. Organ. Behav.* 1(1), 545–568. 1
- Townsend, R. (2007). *Up the Organization*. San Francisco, CA: John Wiley & Sons. (document)
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57(2), 307–333. 4
- Weber, M. (1947). *The Theory of Social and Economic Organization*. New York, NY: The Free Press. 1

## ONLINE APPENDIX

### INSTRUCTIONS FOR PARTICIPANTS

In this section we present an English translation of the originally French instructions that the participants received before the experiment. There were separate instructions for employers and workers in each experiment. We first provide the version for employers and then the version for workers.

The black text corresponds to the Baseline instructions. The colored paragraphs include text that we added in the other versions of the experiment. The **blue** paragraphs were only present in the communication experiment, the **red** paragraphs were only present in the agreement experiment, and the **teal** paragraphs were only present in the grievance experiment. In some cases the added paragraphs changed the numeration in a list of paragraphs. In those cases the numeration of the colored paragraphs shows where the paragraphs were placed.

## Instructions for Employers

### Introduction

You are about to participate in an experiment of the University of Lausanne. During this experiment you have the opportunity to earn a sum of money that will be paid to you at the end of the experiment. The amount of money you earn may be more significant if

- you read the instructions carefully.
- you think carefully about the decisions you make.

If you have any questions while reading the instructions or while the experiment is in progress, feel free to call us by raising your hand. By contrast, any communication between participants—except through the channels offered as part of the experiment—is prohibited. In the event of non-compliance with these instructions, we will be obliged to exclude you from the experience without any payment.

In today's experiment, you will interact with other participants via your computer. The decisions you make will have an impact on your profit. Your decisions will also influence the profit of other participants, just as the decisions of other participants may influence your profit.

Your profit is calculated in points. At the end of the experiment your points will be converted into Swiss Francs according to the following exchange rate:

$$\mathbf{70 \text{ points} = 1 \text{ Swiss Franc}}$$

Regardless of your decisions in the experiment, you will also receive a fixed amount of CHF 10 for your participation.

The experiment consists of several identical rounds. At the end of the session, your remuneration will be calculated as the sum of your income obtained in all these rounds.

## SUMMARY OF THE EXPERIMENT

There are two types of participants in today's experiment: employers and workers. Each participant is randomly assigned to one of the two roles.

You will be an **EMPLOYER** for the whole duration of the experiment.

The experiment consists of a total of 15 periods. At the beginning of each period you are randomly matched with a new worker. In each period you have the possibility to interact with the worker you are paired with. These interactions will be governed by the same rules in each period. However, since you are randomly assigned a new worker in each period, you will interact with many different trading partners over the duration of the experiment.

### Interaction between employers and workers

In what follows we summarize how the interaction between an employer and a worker takes place. In each period the employer and the worker go through the same six stages. At this point we only provide a very brief and general description of each stage. The purpose of this first part of the instructions is to give you an overview of what will happen in the experiment. In the second part of the instructions we provide a much more detailed description of each stage and we also illustrate how you can enter your decisions into the computer.

#### **(1) Assignment of a new worker**

At the beginning of every period you are randomly assigned a new worker. The compensation of the worker is regulated in the following way:

- You pay the worker a fixed wage that he receives with certainty. The fixed wage is the same in each period and you cannot change it.
- In addition, you have the possibility to pay the worker a bonus. You decide whether or not to pay the bonus after you have received some information about the return that the worker has produced for you in this period.

#### **(1-2) Communication between employer and worker**

At the beginning of each period you have the opportunity to send a message to your worker. In this message you can indicate your strategy regarding the bonus payment.

At the same time, your worker has the opportunity to send you a message in which he can indicate his strategies regarding the effort level and the potential decision to reduce the return.

### **(1-2) Message exchange between employer and worker**

At the beginning of each period you must make the decision to accept a proposed working agreement or not. At the same time, your worker has to make the same decision - accepting the proposed working agreement or not.

It is necessary for both parties – your worker and you – to accept the agreement for it to be in place during the interactions.

Two situations can arise:

- (1) If your worker AND you both decide to accept the agreement, the subsequent interactions will follow with the agreement in place.
- (2) If your worker OR you decide not to accept the agreement, the subsequent interactions will follow without the agreement in place.

### **(1-2) Message exchange between employer and worker**

At the beginning of each period you must make the decision to accept a proposed working agreement or not. At the same time, your worker has to make the same decision - accepting the proposed working agreement or not.

It is necessary for both parties – your worker and you – to accept the agreement. Otherwise, the interactions will cease and you will receive a fixed amount of **40 points** for this period. In the case that both your worker and you accept the agreement, the interactions will proceed as stated in the agreement.

### **(2) Worker's effort choice**

Your worker produces a return for you. The return that the worker produces can be either large or small. The worker can influence the probability with which the return is high by choosing his effort level.

- If the worker chooses to exert high effort, it is very likely that he produces a large return and not very likely that he produces a low return.
- If the worker chooses to provide low effort, in contrast, it is very likely that the return is low and not very likely that the return is high.

For the worker providing high effort is more costly than providing low effort. Once the worker has determined his effort level a computerized random device determines whether the return is large or small.

### **(3) Information about the quality of the worker's output**

Neither you as the employer nor your worker can directly observe the return. However, you and your worker both receive separate and subjective pieces of information about the return.

It is important to understand two facts about these subjective pieces of information:



- The subjective information is not always correct. Your subjective information may indicate a large return although the true return is small (or vice versa).
- The same is also true for your worker's information.
- In addition, it is also possible that your information differs from the information of the worker, i.e., you may get the subjective information that the return is small, while your worker gets the information that the return is large (or vice versa).

#### **(4) Employer's bonus payment**

After you have received your subjective information about the return produced by the worker, you need to decide whether you would like to pay a bonus to the worker or not.

- If you decide to pay the bonus, the worker receives his fixed wage plus the bonus as compensation.
- If you decide not to pay the bonus, the worker receives only his fixed wage as compensation.

#### **(5) Worker's decision to reduce the return**

The worker observes your decision regarding the bonus payment and can then decide whether or not he would like to reduce the return he has produced for you. If the worker would like to reduce the return this is also costly to him.

*Important:* The worker does not know with certainty whether the return is large or small. He makes his decision solely based on his knowledge about your bonus payment and his subjective information about the return.

#### **(5-6) Decision to file a complaint**

The worker observes your bonus payment decision and can then decide whether to file a complaint or not. Filing a complaint imposes a cost on you and it is also costly for the worker.

*Important:* The worker does not know with certainty whether the return is large or small. He makes his decision solely based on his knowledge about your bonus payment and his subjective information about the return.

#### **(6) Profits**

After the worker's return reduction decision, the period is over. At this point you are informed about the true return that the worker has produced and your profit for this period is displayed on your screen.

The worker also receives information about his profit on the screen. However, the worker is never informed about the true return which he has produced and therefore does not know how much profit you made as employer.

You as the employer do not receive any information about the worker's effort choice and therefore you do not know the profit of the worker.

Subsequently, the next period begins.

**Remember:** At the beginning of each new period each employer is randomly assigned a new worker.

## DETAILED DESCRIPTION OF THE EXPERIMENT

You are an **EMPLOYER** for the whole duration of the experiment.

The experiment is computerized. All decisions that you make during the experiment need be entered in the computer in front of you.

In this second part of the instructions we explain in detail which decision you and the other participants can make, how you can enter these decisions in the computer, and how these decisions affect your own payoff and the payoffs of other participants.

If you have questions while reading the instructions, please raise your hand. An experimenter will come to your workplace and answer your question in private.

### (1) Assignment of a new worker

At the beginning of each period you are randomly assigned a new worker who works for you in the current period. The employment relationship adheres to the following rules:

- You pay the worker a fixed wage of 100 points that he gets with certainty. This wage is the same in every period and you cannot change it.
- In addition, you have the possibility to pay the worker a bonus of 50 points after you have received subjective information about the the return that the worker has generated for you. Whether you you pay the bonus or not is completely up to you.

### (1-2) Communication between employer and worker

At the beginning of each period you have the opportunity to send a message to your worker. In this message you can indicate your strategy regarding the bonus payment. You can choose any of the following messages:

- (1) I will pay the bonus with certainty.
- (2) I will pay the bonus with high probability.
- (3) I will pay the bonus if I have the impression that you exerted high effort.

- (4) I pay the bonus if my private information indicates a high return.
- (5) I will not pay the bonus.
- (6) I prefer not to send a message.

*Important:* The message you send is not binding. always full discretion to decide whether to pay the bonus to your worker or not. If you rather not send a message to your worker, you also have the option to do so.

Likewise, your worker also has the opportunity to send you a message indicating his strategy regarding effort level and the potential reduction of return. The workers messaging options regarding the effort level are:

- (1) I will exert high effort.
- (2) I will exert low effort.
- (3) I prefer not to send a message.

The workers messaging options regarding the decision to reduce the return are:

- (1) I will always reduce the return.
- (2) I will reduce the return if I do not get the bonus.
- (3) I will reduce the return if I do not get the bonus although I got a good signal.
- (4) I will never reduce the return.
- (5) I prefer not to send a message.

The message of the worker is not binding. The worker is always free to decide the level of effort and whether to reduce the return or not. If the worker prefers not to send you a message, he also has the option to do so.

### **(1-2) Message exchange between employer and worker**

At the beginning of each period you are given the option to accept a working agreement. This agreement specifies the optimal strategy. You can find an example of this agreement below.

In order to begin the interaction, your worker and you need to take a decision whether to accept the agreement or not. Furthermore, It is necessary for both parties – your worker and you – to accept the agreement for it to be in place during the interactions. Two situations can occur:

- (1) If your worker AND you both decide to accept the agreement, the subsequent interactions will follow with the agreement in place.
- (2) If your worker OR you decide not to accept the agreement, the subsequent interactions will follow without the agreement in place.

*Important:* The agreement is not binding. You have always full discretion to pay the bonus. Likewise, the worker is not obliged to comply with the terms of the agreement. The worker has full discretion on the level of effort exerted and he can decide freely whether to initiate conflict or not.

## (1-2) Message exchange between employer and worker

At the beginning of each period you are given the option to accept a working agreement. This agreement specifies the optimal strategy. You can find an example of this agreement below.

In order to begin the interaction, your worker and you need to take a decision whether to accept the agreement or not. Furthermore, it is necessary for both parties – your worker and you – to accept the agreement for it to be in place during the interactions. Otherwise, the interactions will cease and you will receive a fixed amount of **40 points** for this period.

*Important:* The agreement is not binding. You have always full discretion to pay the bonus. Likewise, the worker is not obliged to comply with the terms of the agreement. The worker has full discretion on the level of effort exerted and he can decide freely whether to initiate conflict or not.

**Working Agreement**

The worker and the employer accept:

- 1. Activity**  
The worker must exert a high level of effort to increase the probability that the employer receives a high return.
- 2. Duration**  
The working relationship begins when both worker and employer accept the agreement until the end of the period.
- 3. Wage**
  - a) The employer must pay the worker a fix salary of 100 points
  - b) The employer must pay a bonus of 50 points if his subjective information indicates that the return is high.
- 4. Complaint report**
  - a) If the subjective information of the workers indicates a high return, and he did not receive a bonus, the worker must file a report that imposes a fine of 100 points to the employer.
  - b) In every other case, i.e., if the bonus is paid or the subjective information of the worker indicates a low return, the worker must

## (2) Worker's effort choice

Your worker produces a return for you. There are two possibilities:

- The return can be large: 350 Points
- The return can be small: 150 Points

Your worker cannot directly choose the return. However, he can influence the probability with which the return will be large by choosing his level of work effort on the job. The worker can either provide high effort or low effort. Choosing high effort creates a cost for the worker:

- If the worker decides to provide high effort his effort costs are equal to 10 points.
- If the worker decides to provide low effort his effort costs are equal to 0 points.

The worker's effort choice affects the probability that the return is large as follows:

- If the worker provides high effort, the return is large with a probability of 85% and the return is low with a probability of 15%
- If the worker provides low effort, in contrast, the return is large with a probability of only 15% and the return is small with a probability of 85%.

Once the worker has made his effort choice a computerized random device determines the return.

### **(3) Information about the quality of the worker's output**

Neither you nor your worker can directly observe whether the return is large or small at this point in time. However, you will get a subjective piece of information about the return. Likewise your worker also gets a subjective piece of information. The subjective information is private, i.e., you only get to know your information and you never observe the worker's information. At the same time your worker only observes his information and never learns about your information.

***Important:*** The subjective information does not always reflect the truth:

- It is possible that your subjective information indicates a high value although the true value is low (and vice versa). The same is also true for the information of the worker.
- In addition, you and the worker may receive different information, i.e., it is possible that you get the information that the value is high, while the worker gets the information that the value is low (or vice versa).

#### ***Details on subjective information:***

Your subjective information as the employer depends on the true return:

- If the true return is large, your subjective information indicates a large return with a probability of 75% and a small return with a probability of 25%.

- If the true return is small, your subjective information indicates a small return with a probability of 75% and a large return with a probability of 25%.

Your subjective information as the employer depends on the true return:

- If your subjective information indicates a large return, the probability that the worker receives the same subjective information is 75%. However, with a probability of 25% the worker's subjective information indicates a small return.
- If your subjective information indicates a small return, the probability that the worker receives the same subjective information is 75%. However, with a probability of 25% the worker's subjective information indicates a large return.

#### **(4) Employer's bonus payment**

After you have received your subjective information about the return, you need to decide whether you would like to pay a bonus of 50 points to the worker.

- If you decide to pay the bonus the worker not only receives his fixed wage of 100 points, but also the bonus of 50 points. Thus, in this case the worker's total compensation amounts to 150 points.
- If you decide not to pay the bonus, the worker receives only the fixed wage of 100 points as compensation in this period.

#### **(5) Worker's return reduction decision**

The worker is informed about your decision regarding the bonus payment. Then the worker can decide whether or not he would like to reduce the return which he has produced for you as the employer. The worker can reduce your return by at most 80 points. However, reducing the return is also costly to the worker: for each point that he deducts from the return the worker has to pay 0.1 points out of his own profit. For example, if the worker decides to reduce the return by 50 points, he needs to give up 5 points of his own profit.

If the worker decides not to reduce the return, nothing happens and your profit and the worker's profit are unaffected.

The worker's return reduction decision concludes the period. In the third part of the instructions we explain in detail how your choices and the choices of the worker influence your profit as an employer and the profit of the worker.

#### **(5-6) Decision to file a complaint**

The worker is informed about your decision regarding the bonus payment. Then the worker can decide whether or not he would like to file a complaint that will impose a cost on you of 100 points. Filing the complaint is also costly for the worker: he needs to pay 10 points.

In order to file a complaint correctly, the worker needs to accept the reasons behind filing the complaint. If the worker decides not to file a complaint, nothing happens and your profit and the worker's profit are unaffected.

The worker's return reduction decision concludes the period. In the third part of the instructions we explain in detail how your choices and the choices of the worker influence your profit as an employer and the profit of the worker.

## CALCULATION OF INCOMES AT THE END OF THE PERIOD

### **Your profit as an employer**

Your profit as employer is determined as follows:

$$\begin{aligned} \text{Your profit} = & \text{Return produced by the worker} \\ & - \text{fixed wage to the worker} \\ & - \text{bonus payment to the worker if you decide to pay the bonus} \\ & - \text{return reduction if the worker decides to reduce the return} \end{aligned}$$

Thus, your profit increases if the return is high instead of low. Paying a bonus to the worker reduces your profit. In addition, your profit is also reduced if the worker decides to reduce the return at the end of the period.

*Warning:* It is possible that your profit becomes negative in certain periods. For example, assume that you have decided to pay the bonus to the worker, the worker has decided to not reduce the return, but it turns out that the return produced by the worker is small. In this case your profit is  $-50$  ( $\text{Profit} = \text{Return} - \text{Wage} - \text{Bonus} = 150 - 100 - 50 = -50$ ). You have to pay for these negative profits with accumulated profits from previous periods, and, if necessary, with your show-up fee.

### **The worker's profit**

Your profit as worker is determined as follows:

$$\begin{aligned} \text{Worker's profit} = & \text{Fixed wage from employer} \\ & + \text{bonus from employer if the employer decides to pay the bonus} \\ & - \text{cost of effort, if the worker decides to provide high effort} \end{aligned}$$

Thus the worker's profit is higher if you decide to pay him the bonus. Choosing a high effort reduces the profit of the worker.

After the payoff screen disappears a new period begins in which you are randomly rematched with a new worker.

In the fourth part of these instructions we will ask you to answer a series of control questions. Please fill in the necessary calculations in the empty spaces provided on the exercise sheets. Once you have found the final result for a question, you can enter the result on your computer screen and the computer will tell you whether your answer is correct or not.

If you have questions, please raise your hand and wait until an experimenter comes to your workplace to answer your question in private.

The following list summarizes all the important parameters of the experiment. You need these numbers to answer the control questions.

*Return*

- Large return = 350 Points
- Small return = 150 Points

*Cost of worker's effort*

- Cost of high effort = 10 Points
- Cost of low effort = 0 Points

*Worker's compensation*

- Fixed wage = 100 Points
- Bonus = 50 Points

*Probabilities that a small or large return is realized:*

Worker's effort	Large return (300 Points)	Small return (100 Points)
High	85%	15%
Low	15%	85%

*Probabilities that your subjective information indicates a small or large return:*

True return	Subjective information: large	Subjective information: small
Large	75%	25%
Small	25%	75%

*Probabilities that the worker's subjective information indicates a small or large return:*

Your subjective Information	Worker's subj. info: large	Worker's subj. info: small
Large	75%	25%
Small	25%	75%

*Cost of return reduction:* The worker can reduce your return by 100 points which as a cost of 10 points.



## CONTROL QUESTIONS

To check if you comprehended the consequences of your decisions during this experiment, we ask you to answer the following question. You can use the piece of paper to solve the questions. To check whether you've reached the right solution, please, enter your responses in the computer.

### Exercise 1

Assume that the true return is large:

What is the probability that your subjective information indicates a large return: \_\_\_

What is the probability that your subjective information indicates a small return: \_\_\_

Assume that the true return is small:

What is the probability that your subjective information indicates a large return: \_\_\_

What is the probability that your subjective information indicates a small return: \_\_\_

Assume that your subjective information indicates a large return:

What is the probability that your worker has the same subjective information: \_\_\_

What is the probability that your worker's subjective information is different: \_\_\_

Assume that your subjective information indicates a small return:

What is the probability that your worker has the same subjective information: \_\_\_

What is the probability that your worker's subjective information is different: \_\_\_

### Exercise 2

For this exercise assume that your worker chooses to provide **high effort**.

- (1) What is the probability that a *large* (350 points) return is realized?
- (2) What is the probability that a *small* return (150 points) is realized?
- (3) Assume that you do NOT pay the bonus and the worker does NOT reduce the return

	Large return (350 points)	Small return (150 points)
Your profit	___	___
Profit of your worker	___	___

- (4) Assume that you pay the bonus and the worker does NOT reduce the return

	Large return (350 points)	Small return (150 points)
Your profit	___	___
Profit of your worker	___	___

- (5) Assume that you do NOT pay the bonus and the worker reduces the return by 100 points

	Large return (350 points)	Small return (150 points)
Your profit	___	___
Profit of your worker	___	___

- (6) Assume that you pay the bonus and the worker reduces the return by 100 points

	Large return (350 points)	Small return (150 points)
Your profit	___	___
Profit of your worker	___	___

### Exercise 3

For this exercise assume that your worker chooses to provide **low effort**.

- What is the probability that a *large* (350 points) return is realized?
- What is the probability that a *small* return (150 points) is realized?
- Assume that you do NOT pay the bonus and the worker does NOT reduce the return

	Large return (350 points)	Small return (150 points)
Your profit	___	___
Profit of your worker	___	___

- (4) Assume that you pay the bonus and the worker does NOT reduce the return

	Large return (350 points)	Small return (150 points)
Your profit	___	___
Profit of your worker	___	___

- (5) Assume that you do NOT pay the bonus and the worker reduces the return by 100 points

	Large return (350 points)	Small return (150 points)
Your profit	___	___
Profit of your worker	___	___

(6) Assume that you pay the bonus and the worker reduces the return by 100 points

	Large return (350 points)	Small return (150 points)
Your profit	___	___
Profit of your worker	___	___

#### Exercise 4

For this exercise assume that both your worker and you have accepted the working agreement.

(1) According to the agreement, what level of effort should the worker exert?

High or Low

(2) If your subjective information indicates a low return, should you pay the bonus according to the agreement?

Yes or No

(3) If your subjective information indicates a high return, should you pay the bonus according to the agreement?

Yes or No

(4) If you haven't paid the bonus and the subjective information of the worker indicates a high return, should the worker reduce your return according to the agreement?

Yes or No

(5) If you haven't paid the bonus and the subjective information of the worker indicates a low return, should the worker reduce your return according to the agreement?

Yes or No

## Instructions for Workers

### Introduction

You are about to participate in an experiment of the University of Lausanne. During this experiment you have the opportunity to earn a sum of money that will be paid to you at the end of the experiment. The amount of money you earn may be more significant if

- you read the instructions carefully.
- you think carefully about the decisions you make.

If you have any questions while reading the instructions or while the experiment is in progress, feel free to call us by raising your hand. By contrast, any communication between participants—except through the channels offered as part of the experiment—is prohibited. In the event of non-compliance with these instructions, we will be obliged to exclude you from the experience without any payment.

In today's experiment, you will interact with other participants via your computer. The decisions you make will have an impact on your profit. Your decisions will also influence the profit of other participants, just as the decisions of other participants may influence your profit.

Your profit is calculated in points. At the end of the experiment your points will be converted into Swiss Francs according to the following exchange rate:

$$\mathbf{70 \text{ points} = 1 \text{ Swiss Franc}}$$

Regardless of your decisions in the experiment, you will also receive a fixed amount of CHF 10 for your participation.

The experiment consists of several identical rounds. At the end of the session, your remuneration will be calculated as the sum of your income obtained in all these rounds.

## SUMMARY OF THE EXPERIMENT

There are two types of participants in today's experiment: employers and workers. Each participant is randomly assigned to one of the two roles.

You will be an **WORKER** for the whole duration of the experiment.

The experiment consists of a total of 15 periods. At the beginning of each period you are randomly matched with a new employer. In each period you have the possibility to interact with the employer you are paired with. These interactions will be governed by the same rules in each period. However, since you are randomly assigned to a new employer in each period, you will interact with many different trading partners over the duration of the experiment.

### Interaction between employers and workers

In what follows we summarize how the interaction between an employer and a worker takes place. In each period the employer and the worker go through the same six stages. At this point we only provide a very brief and general description of each stage. The purpose of this first part of the instructions is to give you an overview of what will happen in the experiment. In the second part of the instructions we provide a much more detailed description of each stage and we also illustrate how you can enter your decisions in the computer.

#### **(1) Assignment to a new employer**

At the beginning of every period you are randomly assigned to a new employer. Your compensation as worker is regulated in the following way:

- You receive a fixed wage that the employer has to pay with certainty. The fixed wage is the same in each period and the employer cannot change it.
- In addition, the employer has the possibility to pay you a bonus. The employer decides whether or not to pay the bonus after he has received some information about the return that you (as the worker) have produced for him in this period.

#### **(1-2) Communication between employer and worker**

At the beginning of each period you have the opportunity to send a message to your employer. In this message you can indicate your strategy regarding the effort level and the potential decision to reduce the return. At the same time, your employer has the opportunity to send you a message in which he can indicate his strategies regarding the bonus payment.

#### **(1-2) Message exchange between employer and worker**

At the beginning of each period you must make the decision to accept a proposed working agreement or not. At the same time, your employer has to make the same decision - accepting the proposed working agreement or not.

It is necessary for both parties – your employer and you – to accept the agreement for it to be in place during the interactions.

Two situations can arise:

- (1) If your employer AND you both decide to accept the agreement, the subsequent interactions will follow with the agreement in place.
- (2) If your employer OR you decide not to accept the agreement, the subsequent interactions will follow without the agreement in place.

### **(1-2) Message exchange between employer and worker**

At the beginning of each period you must make the decision to accept a proposed working agreement or not. At the same time, your worker has to make the same decision - accepting the proposed working agreement or not.

It is necessary for both parties – your employer and you – to accept the agreement. Otherwise, the interactions will cease and you will receive a fixed amount of **40 points** for this period. In the case that both your employer and you accept the agreement, the interactions will proceed as stated in the agreement.

### **(2) Worker's effort choice**

As a worker you produce a return for your employer. The return that you produce can be either large or small. As a worker you can influence the probability with which the return is high by choosing your effort level.

- If you choose to exert high effort, it is very likely that you produce a large return and not very likely that you produce a low return.
- If you choose to provide low effort, in contrast, it is very likely that the return is low and not very likely that the return is high.

For you providing high effort is more costly than providing low effort. Once you have determined your effort level a computerized random device determines whether the return is large or small.

### **(3) Information about the quality of the worker's output**

Neither you as the worker nor your employer can directly observe the return. However, you and your employer both receive separate and subjective pieces of information about the return.

It is important to understand two facts about these subjective pieces of information:

- The subjective information is not always correct. Your subjective information may indicate a large return although the true return is small (or vice versa).
- The same is also true for your employer's information.
- In addition, it is also possible that your information differs from the information of the employer, i.e., you may get the subjective information that the return is small, while your employer gets the information that the return is large (or vice versa).

#### **(4) Employer's bonus payment**

After your employer has received his subjective information about the return that you have produced, he needs to decide whether he would like to pay a bonus to you or not.

- If the employer decides to pay the bonus, you receive the fixed wage plus the bonus as compensation.
- If the employer decides not to pay the bonus, you receives only the fixed wage as compensation.

#### **(5) Worker's return reduction decision**

As a worker you observe the employer's decision regarding the bonus payment and you can then decide whether or not you would like to reduce the return that you have produced for the employer. If you would like to reduce the return this is slightly costly to you.

*Important:* You do not know with certainty whether the return is large or small. You make your decision solely based on your knowledge about the bonus payment of the employer and your subjective information about the return.

#### **(5-6) Decision to file a complaint**

As a worker you observe the employer's decision regarding the bonus payment and you can then decide whether you would like or not to file a complaint. Filing a complaint imposes a cost on your employer of 100 points and it also costs you 10 points. Instead, if you decide not to file a complaint, neither your employer or you would need to pay the associated cost.

#### **(6) Profits**

After your return reduction decision, the period is over. At this point you receive information about your profit on the screen. However, you are never informed about the true return which you have produced and therefore you do also not know how much profit your employer made. The employer is informed about the true return that you have produced as a worker and his profit for this period is displayed on his screen. The employer does not receive any information about your true effort choice and therefore the employer does not know how much profit you made as a worker.

## DETAILED DESCRIPTION OF THE EXPERIMENT

You are a **WORKER** for the whole duration of the experiment.

The experiment is computerized. All decisions that you make during the experiment need be entered in the computer in front of you.

In this second part of the instructions we explain in detail which decision you and the other participants can make, how you can enter these decisions in the computer, and how these decisions affect your own payoff and the payoffs of other participants.

If you have questions while reading the instructions, please raise your hand. An experimenter will come to your workplace and answer your question in private.

### (1) Assignment to a new employer

At the beginning of each period you are randomly assigned to a new employer who works for you in the current period. The employment relationship adheres to the following rules:

- The employer pays you a fixed wage of 100 points with certainty. This wage is the same in every period and the employer cannot change it.
- In addition, the employer has the possibility to pay you a bonus of 50 points after he has received subjective information about the return that you have generated for him. Whether the employer pays the bonus or not is completely up to him.

### (1-2) Communication between employer and worker

At the beginning of each period you have the opportunity to send you a message to your employer indicating your strategy regarding effort level and the potential reduction of return. At the beginning of each period you have the opportunity to send a message to your worker. In this message you can indicate your strategy regarding the bonus payment. You can choose any of the following messages:

- (1) I will pay the bonus with certainty.
- (2) I will pay the bonus with high probability.
- (3) I will pay the bonus if I have the impression that you exerted high effort.
- (4) I pay the bonus if my private information indicates a high return.
- (5) I will not pay the bonus.
- (6) I prefer not to send a message.

**Important:** The message you send is not binding. You are always free to decide whether to pay the bonus to your worker or not. If you rather not send a message to your worker,



you also have the option to do so. You can choose any of the following messages regarding effort:

- (1) I will exert high effort.
- (2) I will exert low effort.
- (3) I prefer not to send a message.

You can choose any of the following messages regarding the decision to reduce the employer's return:

- (1) I will always reduce the return.
- (2) I will reduce the return if I do not get the bonus.
- (3) I will reduce the return if I do not get the bonus although I got a good signal.
- (4) I will never reduce the return.
- (5) I prefer not to send a message.

**Important:** The messages you choose are not binding. You have full discretion to decide the level of effort and whether to reduce the return or not. If you prefer to not send a message to your employer, you also have the option to do so .

Simultaneously, your employer also has the opportunity to send you a message indicating his strategy regarding bonus payment:

- (1) I will pay the bonus with certainty.
- (2) I will pay the bonus with high probability.
- (3) I will pay the bonus if I have the impression that you exerted high effort.
- (4) I pay the bonus if my private information indicates a high return.
- (5) I will not pay the bonus.
- (6) I prefer not to send a message.

The message of the employer is not binding. The employer is always free to decide the final bonus payment decision. If the employer prefers not to send you a message, he also has the option to do so.

### **(1-2) Message exchange between employer and worker**

At the beginning of each period you are given the option to accept a working agreement. This agreement specifies the optimal strategy. You can find an example of this agreement one Figure 8.

In order to begin the interaction, your employer and you need to take the decision of whether to accept the agreement or not. Furthermore, it is necessary for both parties – your employer and you – to accept the agreement for it to be in place during the interactions. Two situations can occur:

- (1) If your employer AND you both decide to accept the agreement, the subsequent interactions will follow with the agreement in place.
- (2) If your employer OR you decide not to accept the agreement, the subsequent interactions will follow without the agreement in place.

**Important:** The agreement is not binding. You have always full discretion to choose the level of effort and the reduce the return of your employer. Likewise, the employer is not obliged to comply with the terms of the agreement. The employer has full discretion on decision to pay the bonus.

FIGURE 8. Agreement

<b>Working Agreement</b>
<p>The worker and the employer accept:</p> <p><b>1. Activity</b></p> <p>The worker must exert a high level of effort to increase the probability that the employer receives a high return.</p> <p><b>2. Duration</b></p> <p>The working relationship begins when both worker and employer accept the agreement until the end of the period.</p> <p><b>3. Wage</b></p> <p>a) The employer must pay the worker a fix salary of 100 points</p> <p>b) The employer must pay a bonus of 50 points if his subjective information indicates that the return is high.</p> <p><b>4. Complaint report</b></p> <p>a) If the subjective information of the workers indicates a high return, and he did not receive a bonus, the worker must file a report that imposes a fine of 100 points to the employer.</p> <p>b) In every other case, i.e., if the bonus is paid or the subjective information of the worker indicates a low return, the worker must</p>

(2) Worker's effort choice +

As a worker you produce a return for your employer. There are two possibilities:

- The return can be large: 350 Points
- The return can be small: 150 Points

You cannot directly choose the return. However, you can influence the probability with which the return will be large by choosing your level of work effort on the job. You can either provide high effort or low effort. Choosing high effort creates a cost for you:

- If you decide to provide high effort your effort costs are equal to 10 points.

- If you decide to provide low effort your effort costs are equal to 0 points.

Your effort choice affects the probability that the return is large as follows:

- If you provide high effort, the return is large with a probability of 85% and the return is low with a probability of 15%.
- If you provide low effort, in contrast, the return is large with a probability of only 15% and the return is small with a probability of 85%.

### **(3) Information about the quality of the worker's output**

Neither you nor your employer can directly observe whether the return is large or small at this point in time. However, you will get a subjective piece of information about the return.

Likewise your employer also gets a subjective piece of information. The subjective information is private, i.e., you only get to know your information and you never observe the employer's information. At the same time your employer only observes his information and never learns about your information.

#### ***Important:***

The subjective information does not always reflect the truth:

- It is possible that your subjective information indicates a high value although the true value is low (and vice versa). The same is also true for the information of the employer.
- In addition, you and the employer may receive different information, i.e., it is possible that you get the information that the value is high, while the employer gets the information that the value is low (or vice versa).

#### ***Details on subjective information:***

The employer's subjective information depends directly on the true return:

- If the true return is large, your subjective information indicates a large return with a probability of 75% and a small return with a probability of 25%.
- If the true return is small, your subjective information indicates a small return with a probability of 75% and a large return with a probability of 25%.

Your own subjective information as the worker depends on the subjective information of the employer:

- If the employer's subjective information indicates a large return, the probability that you receive the same subjective information is 75%. However, with a probability of 25% your subjective information indicates a small return.

- If the employer's subjective information indicates a small return, the probability that you receive the same subjective information is 75%. However, with a probability of 25% your subjective information indicates a large return.

#### **(4) Employer's bonus payment**

After the employer has received his subjective information about the return, he needs to decide whether he would like to pay you a bonus of 50 points.

- If the employer decides to pay the bonus you not only receive the fixed wage of 100 points, but also the bonus of 50 points. Thus, in this case your total compensation amounts to 125 points.
- If the employer decides not to pay the bonus you receive only the fixed wage of 100 points as compensation in this period.

After the employer has made his bonus payment decision, you are informed about his decision.

#### **(5) Worker's return reduction decision**

After you have been informed about the employer's bonus decision, you can decide whether or not you would like to reduce the return which you have produced for the employer. You can reduce the return by at most 100 points. However, reducing the return is also costly for you: for each point that you deduct from the return you have to pay 0.1 points out of your own profit. For example, if you decide to reduce the return by 50 points, you need to pay 5 points out of your own profit. Instead, if you decide not to reduce the return, nothing happens and the profits of the employer and your own profit are unaffected.

Your return reduction decision concludes the period. In the third part of the instructions we explain in detail how your choices and the choices of the employer influence your profit as a worker and the profit of the employer.

#### **(5-6) Decision to file a complaint**

After you have been informed about the employer's bonus decision, you can decide whether or not you would like to file a complaint. Filing a complaint imposes a cost on your employer of 100 points and it also costs you 10 points.

Instead, if you decide not to file a complaint, nothing happens and the profits of the employer and your own profit are unaffected.

Your return reduction decision concludes the period. In the third part of the instructions we explain in detail how your choices and the choices of the employer influence your profit as a worker and the profit of the employer.

## CALCULATION OF INCOMES AT THE END OF THE PERIOD

### Your profit as an employer

Your profit as employer is determined as follows:

$$\begin{aligned} \text{Your profit} = & \text{Fixed wage from employer} \\ & + \text{bonus from employer if the employer decides to pay the bonus} \\ & - \text{cost of effort, if you decide to provide high effort} \end{aligned}$$

Thus your profit is higher if your employer decides to pay you the bonus. Choosing a high effort reduces your profit.

### The employer's profit

Your profit as worker is determined as follows:

$$\begin{aligned} \text{Employer's profit} = & \text{Return produced by the worker} \\ & - \text{fixed wage to the worker} \\ & - \text{bonus payment to the worker if you decide to pay the bonus} \\ & - \text{return reduction if the worker decides to reduce the return} \end{aligned}$$

Thus, the employer's profit increases if the return is high instead of low. Paying a bonus reduces the employer's profit. In addition, the employer's profit is also reduced if you decide to reduce the return at the end of the period.

**Warning:** It is possible that the employer's profit becomes negative in certain periods. For example, assume that the employer has decided to pay the bonus, you have decided not to reduce the return, but it turns out that the return is small. In this case the employer's profit is  $-50$  (Profit = Return - Wage - Bonus =  $100 - 100 - 50 = -50$ ). The employer has to pay for these negative profits with accumulated profits from previous periods, and, if necessary, with his show-up fee.

You are never informed about the true return which you have produced and therefore you do also not know how much profit your employer made.

The employer is informed about the true return that you have produced as a worker and his profit for this period is displayed on his screen, but the employer does not receive any information about your true effort choice and therefore the employer does not know how much profit you made as a worker.

After the payoff screen disappears a new period begins in which you are randomly rematched with a new worker.

In the fourth part of these instructions we will ask you to answer a series of control questions. Please fill in the necessary calculations in the empty spaces provided on the exercise sheets. Once you have found the final result for a question, you can enter the result on your computer screen and the computer will tell you whether your answer is correct or not.

If you have questions, please raise your hand and wait until an experimenter comes to your workplace to answer your question in private.

The following list summarizes all the important parameters of the experiment. You need these numbers to answer the control questions.

*Return*

- Large return = 350 Points
- Small return = 150 Points

*Your effort costs*

- Cost of high effort = 10 Points
- Cost of low effort = 0 Points

*Your compensation*

- Fixed wage = 100 Points
- Bonus = 50 Points

*Probabilities that a small or large return is realized:*

Worker's effort	Large return (300 Points)	Small return (100 Points)
High	85%	15%
Low	15%	85%

*Probabilities that the employer's subjective information indicates a small or large return:*

True return	Subjective information: large	Subjective information: small
Large	75%	25%
Small	25%	75%

*Probabilities that your subjective information indicates a small or large return:*

Your subjective Information	Worker's subj. info: large	Worker's subj. info: small
Large	75%	25%
Small	25%	75%

*Cost of return reduction:* The worker can reduce your return by 100 points. For each point that he deducts from the return he has to pay 0.1 Points out of his own profit.

## CONTROL QUESTIONS

To check your if you comprehended the consequences of your decisions during this experiment, we ask you to answer the following question. You can use the piece of paper to solve the questions. To check whether you've reached the right solution, please, enter your responses in the computer.

### Exercise 1

Assume that the true return is large:

- (1) What is the probability that the employer's subjective information indicates a large return? \_\_\_
- (2) What is the probability that the employer's subjective information indicates a small return? \_\_\_

Assume that the true return is small:

- (1) What is the probability that the employer's subjective information indicates a large return? \_\_\_
- (2) What is the probability that the employer's subjective information indicates a small return? \_\_\_

Suppose that the employer's subjective information indicates a large return:

- (1) What is the probability that you receive the same subjective information? \_\_\_
- (2) What is the probability that your subjective information is different? \_\_\_

Suppose that the employer's subjective information indicates a small return:

- (1) What is the probability that you receive the same subjective information? \_\_\_
- (2) What is the probability that your subjective information is different? \_\_\_

### Exercise 2

For this exercise assume that your worker chooses to provide **high effort**.

- (1) What is the probability that a *large* (350 points) return is realized?
- (2) What is the probability that a *small* return (150 points) is realized?
- (3) Assume that the employer does NOT pay the bonus and you do NOT reduce the return

	Large return (350 points)	Small return (150 points)
Your profit	___	___
Profit of your employer	___	___

(4) Assume that the employer pays the bonus and you do NOT reduce the return

	Large return (350 points)	Small return (150 points)
Your profit	___	___
Profit of your employer	___	___

(5) Assume that the employer does NOT pay the bonus and you reduce the return by 100 points

	Large return (350 points)	Small return (150 points)
Your profit	___	___
Profit of your employer	___	___

(6) Assume that the employer pays the bonus and you reduce the return by 100 points

	Large return (350 points)	Small return (150 points)
Your profit	___	___
Profit of your employer	___	___

### Exercise 3

For this exercise assume that your worker chooses to provide **low effort**.

- (1) What is the probability that a *large* (350 points) return is realized?
- (2) What is the probability that a *small* return (150 points) is realized?
- (3) Assume that the employer does NOT pay the bonus and you do NOT reduce the return

	Large return (350 points)	Small return (150 points)
Your profit	___	___
Profit of your employer	___	___

(4) Assume that the employer pays the bonus and you do NOT reduce the return

	Large return (350 points)	Small return (150 points)
Your profit	___	___
Profit of your employer	___	___

(5) Assume that the employer does NOT pay the bonus and you reduce the return by 100 points



	Large return (350 points)	Small return (150 points)
Your profit	___	___
Profit of your employer	___	___

(6) Assume that the employer pays the bonus and you reduce the return by 100 points

	Large return (350 points)	Small return (150 points)
Your profit	___	___
Profit of your employer	___	___

### Exercise 4-2

For this exercise assume that both your worker and you have accepted the working agreement.

(1) According to the agreement, what level of effort should the worker exert?

High or Low

(2) If your subjective information indicates a low return, should you pay the bonus according to the agreement?

Yes or No

(3) If your subjective information indicates a high return, should you pay the bonus according to the agreement?

Yes or No

(4) If you haven't paid the bonus and the subjective information of the worker indicates a high return, should the worker reduce your return according to the agreement?

Yes or No

(5) If you haven't paid the bonus and the subjective information of the worker indicates a low return, should the worker reduce your return according to the agreement?

Yes or No

### Exercise 4-2

For this exercise assume that both your worker and you have accepted the working agreement.

(1) According to the agreement, what level of effort should the worker exert?

High or Low

(2) If your subjective information indicates a low return, should you pay the bonus according to the agreement?

Yes or No

- (3) If your subjective information indicates a high return, should you pay the bonus according to the agreement?

Yes or No

- (4) If you haven't paid the bonus and the subjective information of the worker indicates a high return, should the worker reduce your return according to the agreement?

Yes or No

- (5) If you haven't paid the bonus and the subjective information of the worker indicates a low return, should the worker reduce your return according to the agreement?

Yes or No

COLUMBIA UNIVERSITY, DEPARTMENT OF ECONOMICS 420 WEST 118TH, MAIL CODE 3308, NEW YORK, NY 10027-7296.

*Email address:* `wbmacleod@wbmacleod.net`

UNIVERSITY OF LAUSANNE, UNIL-CHAMBERONNE, INTERNEF 496, CH-1015 LAUSANNE.

*Email address:* `victoria.vallelara@unil.ch`

UNIVERSITY OF LAUSANNE, UNIL-CHAMBERONNE, INTERNEF 412, CH-1015 LAUSANNE.

*Email address:* `christian.zehnder@unil.ch`