pdftex,unicode=true,pdfusetitle, bookmarks=false, breaklinks=false,pdfborder=0 0 1,backref=section,colorlinks=false

ON BULDING A CONFLICT CULTURE IN ORGANIZATIONS

W. BENTLEY MACLEOD, VICTORIA VALLE LARA, AND CHRISTIAN ZEHNDER

Although conflicts are typically associated with negative emotions and waste of resources, organizations may still benefit from a corporate culture that tolerates or even encourages conflicts. The reason is that coordinated conflicts can help to enforce social norms 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.

Date: 2019-08-09.

"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

1. Introduction

A corporate culture that tolerates or even encourages conflicts in case of perceived injustice entails both opportunities and dangers. On one hand, a healthy conflict culture may establish credible threats that motivate otherwise non-complying parties to respect social norms and to contribute to the greater good. On the other hand, a conflict-friendly culture may also create escalations, negative emotions and waste of resources. In this paper we study the role of conflicts for employment relationships in which compensation depends on subjective performance evaluation. We report evidence from a series of laboratory experiments illustrating that the formation of an efficiency-enhancing 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. A functioning conflict culture requires a careful management of norms. In our experiment we find that conflicts have positive net effects only if an explicit code of conduct is established and conflicts are institutionalized through a grievance process.

There is ample evidence illustrating that perceived unfairness is an important trigger of conflicts. Many people are willing to engage in costly retaliation in response to observed violations of fairness norms. Famous examples of such behavior include rejections of small offers in ultimatum games (Güth et al., 1982; Kahneman et al., 1986), punishment of free-riders in public goods experiments (Ostrom et al., 1992; Gächter and Fehr, 2000), or third-party interventions in distribution games (Fehr and Fischbacher, 2004b; lei, 2012). Although such conflicts 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 conflicts may induce norm breakers to comply and cooperate (Gächter et al., 2008).

Existing studies reporting a strong disciplining and cooperation-enhancing effect of conflicts have predominantly focused on deterministic set-ups with symmetric information (see, Chaudhuri, 2011, for a survey).[1] Those findings may not generalize to the employment context where relationships are typically characterized by hidden actions, stochastic input-output links, and noisy information. The presence of probabilistic outcomes and information asymmetries may considerably complicate the role of conflicts, because there may be disagreement about what constitute fair behavior and violations of norms become harder to detect. Our study presents a novel experimental design that sheds new light on the empirical relevance of those important issues.

The game underlying our experiments builds on a two-person, principal-worker setting 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 signal that gives imperfect information about the output created by the worker. These signals are noisy and imperfectly correlated so that the principal and the worker sometimes have contradicting information about output. Given that signals are private information, it is not possible to contract on them. After observing the subjective signal, principals can therefore simply decide whether or not to pay a bonus to the worker. To study the impact of conflicts on efficiency, we compare two different treatments. In the no

[1]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.

conflict condition, the game ends after the worker observes the bonus choice of the principal. In the conflict condition, the worker can start a conflict with the principal after observing the bonus choice. Conflicts are implemented in the form of costly sanctions. This reflects the view that the immediate consequence of conflict is typically negative for all involved parties, but mostly so for the party under attack.

So what should the players optimally do in our setup? Efficiency requires that workers exert high effort. Principals, in turn, can use bonus payment to reward workers and establish a fair allocation of payoffs. Unfortunately, such a simple reciprocal exchange is unlikely to occur. It is a well-established fact in the social preference literature that there is large heterogeneity in the degree to which people care about social motives such as fairness. In particular, whereas some people go out of their way and make considerable sacrifices to enforce social norms, there is also a considerable part of the population whose behavior is best described as approximately selfish (see, e.g., Cooper and Kagel, 2016, for a recent survey of the literature). In the absence of conflicts, the presence of selfish principals who have no motive to pay bonuses undermines a well-functioning exchange of gifts. Intuitively, conflicts can help to mitigate this problem: if sufficiently many workers sanction principals who violate fairness norms, even selfish principals may have an incentive to reward workers appropriately (Fehr and Fischbacher, 2003, 2004a).

A complicating issue is that it is not totally obvious what a fair compensation entails in our setup. One view is that workers should be compensated for exerting high effort (the pay-for-input norm). Alternatively, one might argue that workers should be rewarded if output turns out to be high (the pay-for-output norm).[2] For a functioning conflict culture it is crucial to coordinate on a fairness norm. If trading partners (unknowingly) disagree about the definition of fair behavior, misunderstandings may lead to purely wasteful conflicts. Moreover, as it turns out, it not only matters that parties agree on some fairness norm, but the efficiency effects may also depend on the particular norm they coordinate on.

Under the pay-for-input norm workers feel entitled to a bonus payment whenever they exert high effort. Accordingly, enforcers of that norm ignore their subjective output signal and simply sanction principals who fail to pay the bonus. If there are sufficiently many norm enforcers, the threat of conflict will induce principals to pay the bonus irrespective of their signal. The advantage of the pay-for-input norm is that it doesn't require any conflicts on the equilibrium path, because principals are disciplined by the threat of conflict alone. The downside, however, is that the incentive structure established by the pay-for-input norm fails to motivate selfish workers to exert high effort. Because bonus payments are unconditional selfish workers can invade the system and free-ride on the norm enforcement of others. If the population fraction of selfish workers is significant, the pay-for-input norm will therefore not lead to a large increase in performance.

The pay-for-output norm, in contrast, implies that principals should pay the bonus only if their subjective signal indicates a high return. Moreover, workers feel entitled to a bonus payment only if their own subjective signal suggests that output is high. If a sufficiently large population of workers enforce the pay-for-output norm, principals have an incentive to make signal-contingent bonus payments and all workers (including those who are selfish) have an incentive to provide high effort. The pay-for-output norm therefore establishes an incentive structure that is robust to the invasion of (a reasonably sized population of) selfish workers. The only disadvantage of the pay-for-output norm is that it requires conflicts on the equilibrium path. Specifically, if the principal's signal is bad, but the worker's signal is good, a conflict will emerge. However, as long as

[2]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.

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.

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 or clarification. We compare two treatments in a between-subjects design: in one condition conflicts are absent by design, in the other condition workers can endogenously engage in costly conflicts after having observed the principal's bonus decision. In both treatments 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. 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 rather than the pay-for-output 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 opportunities are available and organizations can encourage conversations on conflict to foster the development of a broadly accepted fairness norms. Our second experiment (subsequently termed "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 select a message from a pre-defined set of messages and send it to their trading partner. The set of available messages consisted of strategy announcements (principals: bonus payment strategy / workers: effort choice and conflict initiation strategy) and included the strategies underlying the two fairness norms plus some other alternatives.

The results of our second experiment reveal that the 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 seems difficult in our setting. As a next step we therefore study the impact of a direct external appeal to coordinate on the pay-for-output norm. Our third experiment (henceforth called "the agreement experiment") explores the effectiveness of a code of conduct in our set-up. In this experiment we begin each period with a stage in which both the principal and the worker are asked 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 seem to 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.

In our fourth and final experiment (subsequently called "the grievance experiment") we extend the agreement experiment by adding one decisive feature: institutionalized conflicts. In this experiment the trading parties sign the same code of conduct as in the agreement experiment. The only difference is that in this experiment 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 same action space as the agreement experiment. In particular, there is no verification process after an worker files a complaint and there is no material consequence to lying. Nevertheless, we observe that the outcomes in the grievance experiment are substantially different from those in our other three experiment. In this experiment, the trading parties follow more frequently the pay-for-output norm and conflicts emerge in a much more coordinated manner. 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. The treatment comparison in this experiment reveals 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 (De Dreu et al., 2008; Coleman et al., 2014; Tjosvold et al., 2014). Our findings suggest that even conflicts that appear to be purely wasteful may be ultimately efficiency-enhancing if they occur in a coordinated manner. The reason is that the presence of these conflicts establishes an implicit incentive system that increases cooperation. The challenging part for management is the establishment of broadly shared fairness norms. In our view this is part of what creating a corporate culture means. Our point is closely related to the approach of identity economics which argues that individuals' preferences should no longer be modeled as fixed but rather as as a function of social context (Akerlof and Kranton, 2000, 2005, 2010). The assumption is that people not only care about outcomes, but are also concerned about the extent to which behavior is compatible with social norms they identify with. Our paper is a first attempt to shed some light on strategies that managers might choose to install such norms.

The remainder of our paper is organized as follows ...

Complete overview

2. Framework

Framework se needs intro

2.1. **The Subjective Evaluation Game.** As the basis for our analysis we use a simplified version of the subjective evaluation model introduced in Macleod (2003, Section III). In particular, we consider the following contracting problem: The principal hires an 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 $c(e) = ce$, 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 the principal and the worker receive private and subjective performance signals. The principal's subjective signal $s_P \in \{0, 1\}$ is directly

dependent on the return and is determined by the following conditional probabilities:

$$Pr\{s_P = 1 | r = r_H\} = p,$$
$$Pr\{s_P = 1 | r = r_L\} = 1 - p,$$

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:

$$Pr\{s_A = 1 | s_P = 1\} = q.$$
$$Pr\{s_A = 1 | s_P = 0\} = 1 - q,$$

where $q \in [0.5, 1]$.

After having observed her private signal, the principal decides whether or not she pays the bonus ($b \in \{0, B\}$) to the worker on top of the contractually agreed upon base wage w. The worker observes the bonus decision of the principal and 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 workers K, but workers may experience a subjective, psychological utility v if they sanction a principal that they believe has violated a fairness norm. The total cost of initiating a conflict for the worker is therefore $k(d, v) = dK - v$, where $0 < K \leq F$ and $v \geq 0$.

Thus, the principal and the worker realize the following payoffs:

$$\pi_P = r_L + \gamma_e \delta - w - b - f(d),$$
$$\pi_A = w + b - c(e) - k(d, v).$$

The decisive parameters of this game are γ_1, γ_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-worker model). Finally, if $\gamma_1 = 1$ and $\gamma_0 = 0$, input and output are perfectly correlated so that the contracting problem becomes trivial and first best effort can be implemented by a simple output-contingent contract.

2.2. **Norm Guided Behavior.** The economic approach to behavior is such an environment typically relies upon 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 has shown that individual do respond to material incentives much has predicted by the theory. However, there is also decades of research showing that individuals respond imperfectly. Them make mistakes, do no consider all the options, get upset and process information imperfectly. One of the jobs of management is to help organizations perform better, given the empirical fact that they must work with imperfect humans.

A fundamental reason that individual do not always choose the optimal choice is that the strategy space they face is huge. Even though our experiment is a simplified version of employment, we never the less observe all sorts of behavior. One of the ways this problem is solved in practice is to provide guidance on appropriate behavior. It is natural for humans to aspire to work hard and adopt to accepted norms. Technically, we can think about strategies as a bundle of behaviors that "make sense" and fit together naturally. It is not at all obvious what is "natural" in a particular context, but in our experiment we find that we can organize our

results around three norms. We can then ask the question, is there a set of norms that forms an efficient, stable equilibrium?

As we discussed in the introduction, Lazear's has identified two norms that are natural in a work on context. The first we call "pay for input". When the Principal and worker meet, the worker agrees to provide effort in exchange for pay. This is a norm is a version of reciprocity behavior that has been widely documented in behavioral economics. The point is that if the cost of effort is X then the worker feels that she should be compensated for X.

Such a system is unstable when effort supplied is not observed, and hence the principal cannot be sure that effort is in fact supplied. Classical agency theory predicts that if effort cannot be observed, then the optimal contract should be one in which the worker is rewarded for observed performance measures. One can immediately see the behavioral problem. In general pay for input is efficient, and one can see workers viewing firms that do not "trust" them as suspect. In fact, in our experiment we do observe situations where workers appear to expect "pay for input". The fact that we observe this in the data implies (yet again) agency theory cannot explain observed behavior!

The problem is that in a large population of workers not all workers can be trusted, and hence there will be defections. The solution proposed by agency theory is to make pay vary with observed measures of output. In our experiment the principal can observe a signal of performance. We dub "pay for output" as a social norm that entails having the principal pay a bonus if and only if the principal observes a high signal. The twist, relative to agency theory, is that this signal is *subjective* and not observed by the worker. Formally, let σ_W^{PI} and σ_P^{PI} be the strategies in the game described above corresponding to "Pay for Input" or PI.

Thus, even if the principal uses his signal to determine the bonus, the worker will on her side not be sure that the principal has performed. This is solved by the fact that the worker has her own performance signal. This can be used to ensure that the principal on his side performs as promised. In that case, if the worker observes a low signal, then she knows she does not deserve a bonus. However, if she observes a high signal, and the principal does not pay a bonus, then she is aggrieved and feels that the principal has not treated her fairly. In our experiment we allow the worker to impose a cost upon the principal. There is a large behavioral literature that shows that individuals are willing to punish others who do not behave fairly. In our experiment the "pay for output" is now in which the principal pays a bonus if and only if performance is high, and the worker punishes the principal whenever she fells unfairly treated. Formally, let σ_W^{PO} and σ_P^{PI} be the strategies in the game described above corresponding to "Pay for Output" or PO.

Finally, we consider the completely "cynical" players. In this experiment it is cost to initiate conflict, and hence a cynical/rational player would never punish the Principal for being unfair. Given this, the cynical/rational Principal would face no consequences to not paying a bonus, and hence would never pay. The worker will not be rewarded for high effort, and hence will shirk. This is the unique sub-game perfect Nash equilibrium for this game. While some players behave in this way, most do not, and hence the hypothesis that play can be predicted by the sub-game perfect Nash equilibrium concept is soundly rejected by the data.

In terms of organizing the results we can think in terms of a "norm" equilibrium, motivated in part by Axelrod (1981). Rather than assume parties can follow *any* strategy, let us suppose that cognitive/behavior effects imply that they focus upon one norm, though imperfect. Let the strategy space for each player be

given by:

$$\Sigma_P = \left\{\sigma_P^{PI}, \sigma_P^{PO}, \sigma_P^R\right\},$$
$$\Sigma_W = \left\{\sigma_W^{PI}, \sigma_W^{PO}, \sigma_W^R\right\}.$$

Let $U_P(\sigma_P, \sigma_W)$ and $U_P(\sigma_P, \sigma_W)$ be the corresponding payoffs. Given this game, the notion of a Nash equilibrium is well defined in norm space. For the payoffs of this game we shall show that both "Pay for Output" and "Rational" are equilibria for the game, while "Pay for Input" is not.

In practice there is lots of noise in the play of individuals. What we can do is measure how close a play is to a particular strategy. Notice that player faces a sequence of binary choices, that in turn can be viewed as a probability vector in a simplex. For the Principle, he has a single choice - how much bonus to pay when the signal is H or L. Hence, the strategy of the Principle is given by:

$$\sigma_P = (p_1, p_2) \in \Delta^2 = \{(p_1, p_2) \,|p_1, p_2 \geq 0, p_1 + p_2 = 1\},$$

where $p_1$is the probability of paying a bonus with the high signal, and p_2 is the probability of paying a bonus with the low signal. We this notation is follows:

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

The workers has a more complex decision. She needs to chose high or low effort, and then choose to initiate conflict depending upon her bonus and her signal. We suppose that the worker with imposes a cost or not upon the Principal. Suppose that p_E is the probability of high effort. Let her information set be given by I=$\{BH, BL, 0H, 0L\}$, where BH means a bonus was paid and she observed a high 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:

$$\sigma_W = (p_E, p_{BH}, p_{BL}, p_{0H}, p_{0L}) \in \Delta^5.$$

The three types are:

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

This defines some archetype strategies. This each choice is a binary variable, we will be able to estimate choice using a logit, and then ask how different treatments result in parties changing their strategies in response. The next section details the material payoffs. That is followed by a table showing that for this employment problem there are two Nash equilibria - either all rational or all pay for output. In the case of Pay for Input this is not an equilibrium since the worker faces an incentive to shirk on effort.

3. Experiments

Our interest in this paper is to investigate the endogenous formation of an efficiency-enhancing conflict culture. In our 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 will illustrate in detail below, 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 ended up running three additional experiments that explore the impact of structured communication (the communication experiment), codes of conduct (the agreement experiment), and an 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. 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 first present the results of each experiment in isolation and do not pool our data. We establish our main results using two-sided, non-parametric tests that use session-level averages from the same experiment as independent observations. In the final part of this section we also present additional results using a pooled data set that includes observations from all experiments together. Doing so allows us to apply more complex statistical models including regression analysis with standard errors clustered at the session level.

We next describe the subjective evaluation game underlying all our experiment 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. We wrap up the section with additional results based on the pooled data set including all experiments.

3.1. **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 rematched 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 he 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: $Prob(s_P = 1|s_P = 1) = Prob(s_P = 0|s_P = 0) = 0.75$.[3]

(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 to initiate conflict. When initiating a conflict the worker determines the intensity of the conflict $d \in [0, 100]$. Conflict harms the principal by reducing his payoff by $f(d) = dF = d$ and creates a cost $k(d) = dK$ for the worker, where $K \in \{0.1, \infty\}$ depending on the treatment.

Payoffs. For the moment suppose that when the worker decides to punish she sets d = 1-maximal punishment. We can then compute the social surplus for each combination of the three strategies, rational, pay for input and pay for output. This are shown in Table (1).

TABLE 1. Social Surplus

		Worker		
		Rational (E=0)	Pay for Input (E=1)	Pay for Output (E=1)
Principal	Rational	180	200	236
	Pay for Input	180	310	290
	Pay for Output	180	274	290

Notice that the pay for input has the highest surplus, followed by the pay for output. This is because it entail high effort, and no conflict costs. If workers and firms were trustworthy, then this would be the outcome all would agree upon. However, it is not stable. We can see this in the next figure which normalized output to the equilibrium utility, and the gain from deviation. Thus 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.

Observe that when strategies are restrict to one of the three norms of behavior, then both "rational" and "pay for output" are stable norms. Neither the Principal, nor the Agent gain from deviated from the norm. This is not true of the "pay for input" norm. The worker has a net gain of 10 by moving to the rational norm. However, as we shall see there is evidence that some players would like to coordinate upon the "pay for input" norm. The purpose of the experimental manipulations is to see if we it is possible to have players coordinate upon a single "organizational norm".

[3]Obviously, it would be possible to write down an equivalent structure of signals in which both signals depend on output. However, we chose the way of presenting the signals to the particpants, because it makes the imperfect correlation of the two signals very transparent to the particpants.

TABLE 2. Gain to Deviation from Norm

		Worker					
		Rational (E=0)		Pay for Input (E=1)		Pay for Output (E=1)	
		Principal	Worker	Principal	Worker	Principal	Worker
Principal	Rational	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

3.2. **Experiment 1: Baseline. Design**

In our first experiment the participants played the game exactly as described in section 3.1. We implemented the following two treatments:

- **No Conflict Treatment**: In this treatments the interaction of 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 efficiency gains from higher effort, so that total surplus in the conflict treatment is lower than in the no conflict treatment.*

The top row of Figure 3.1 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 the increase in effort is not statistically significant (RS: $p = 0.420$).[4] 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 3.2 illustrates the reason for the lower surplus in the conflict treatment. The figure contrasts the total cost of conflict (i.e., the worker's cost of initiating the conflict plus the damage imposed on the principal)

[4]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.

with the additional gains from trade that result from the increased effort. 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

	Effort			Surplus		
Periods	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

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 emerged 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.*

The top row of Figure 3.3 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: the conflict rate is highest if workers who decide to exert high effort are not rewarded with a bonus. Second, the pay-for-output 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. We explain these two points in detail in the subsequent paragraph.

Figure 3.3 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. 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: $p = 0.063$).[5]

Figure 3.4 provides a different perspective on 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 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 predominantly 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 (pfo), pay-for-input norm enforcers (pfi), rational non-enforcers (r) and motivated non-enforcers (mne). Pay-for-output norm 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. Pay-for-input 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. Motivated non-enforcers 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. Rational non-enforcers, finally, pick low effort and do not engage in conflict. They appear as small dots in the bottom-left quadrant.

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 (pfi or pfo). Moreover, there are more pay-for-input norm enforcers (25.5%) than pay-for-output norm enforcers (17.0%). The majority of non-enforcers (46.8%) are rational and exert mostly low effort (36.2%) and motivated non-enforcers 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 10 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 ($\tilde{4}5$-52%). This

[5]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: $p = 0.125$).

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)

	High Effort			Low Effort		
Periods	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 3.3 and 3.4 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 3.5 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

	High Signal			Low Signal		
Periods	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 3.6 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 3.4, 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 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 having observed 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 worker's inability to coordinate on the enforcement of the pay-for-output norm implies that also the principals 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 a monetary 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 rational non-enforcers observed in Figure 3.4 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.

3.3. **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; 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.[6]

[6]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.

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.
(5) I prefer not to send a message.

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

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 3.1 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 3.2 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 from 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 also 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

	Effort			Surplus		
Periods	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))**.** *Participants choose most frequently 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 condition and the no conflict condition chooses to send the message: "I will exert high effort" (79.0% and 85.6%, respectively) and only a small proportion in both treatments chooses the message "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 condition 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 or abstain from engaging in conflicts. Conflict patterns consistent with the pay-for-output norm are rarely observed.*

The second row of Figure 3.3 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 in line with the pay-for-input norm 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 3.4 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. The cluster in the top-right shows both large orange dots, but also small red dots. Whereas the large orange dots correspond to individuals who subscribe to the pay-for-input norm, the red dots represent workers who systematically initiate conflicts without providing high effort themselves. This indicates that the conflicts observed in Figure 3.3 after not obtaining a bonus payment, are not only a consequence of imperfect pay-for-input norm 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 in the pay-for-input norm 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-enforcer (pay-for-input and pay-for-output norms together) amounts to 44.4%, which is almost identical to the corresponding rate in the baseline experiment: 42.5%. However, the fraction of pay-for-input norm followers 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 remains roughly constant (11.1% (communication) vs. 10.6 (baseline)), the share of rational non-enforcers

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 10 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 a 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 pay-for-input norm in the communication experiment.

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

	High Effort			Low Effort		
Periods	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 subject signal, 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, moreover, the large majority of principals pay the bonus more frequently after having observed a positive signal. The second row of Figure 3.5 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 systematically 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 3.6 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

	High Signal			Low Signal		
Periods	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.

3.4. **Experiment 3: Agreement (Code of Conduct). Design**

The first two experiments have fully relied on self-coordination. The results indicate that people seem to 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 both the principal and the worker are asked 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 ($u_r = 40$) 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 was 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 This section presents the results of our third experiment. 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 3.1 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 3.2 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 overall 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 the no conflict treatment, nearly all participants accept the agreement; in the conflict treatment all participants accept the agreement.*

Overall, a total of 97.5% of the participants accept the agreement (95.1% in the no conflict treatment and 100% in the conflict treatment) and 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 week 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 3.3 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 observes 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.[7]

The punishment strategies of individual workers displayed in Figure 3.4 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 rational 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-input 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 3.5 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 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 3.6 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.

[7]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% ($p = 0.844$).

3.5. **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 in 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 one of 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 worker 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 3.1 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 $p = 0.844$).[8]

[8]See the bottom right panel of Figure 3.2 for an illustration of the total cost of conflict (20.2) in comparison with the benefits of additional effort (26.2).

An analysis of the dynamics reveals that institutionalized conflicts help sustain a higher effort level in all phases of the experiment. Table 3 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.

TABLE 9. Formal agreement with grievance - Development of effort and surplus over time

	Effort			Surplus		
Periods	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

We next explore how the grievance procedure affects behavior in the different stages of the subjective evaluation game.

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

Similar to the agreement treatment, most of the participants (98.8%) accept the agreement (98.1% in the no conflict treatment and 99.5% in the conflict treatment). Acceptance rates do not change substantially in either treatment over time.

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 3.3 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: $p = 0.032$).[9]

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 3.4 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

[9]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% ($p = 0.094$).

of the worker's 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)

	High Effort			Low Effort		
Periods	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

In the next result we explore the principals' bonus payments.

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 3.5 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 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.

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

	Positive Signal			Negative Signal		
Periods	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

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 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. The Effect of Treatments on Norms

In this section we report the effects of treatment upon the incidence of different norms.

Section needs be completed

5. Concluding Discussion

References

(2012). An exploration of third and second party punishment in ten simple games. *Journal of Economic Behavior and Organization*, 84(3):753–766.

Akerlof, G. A. and Kranton, R. (2010). Identity economics. *The Economists' Voice*, 7(2).

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753.

Akerlof, G. A. and Kranton, R. E. (2005). Identity and the economics of organizations. *Journal of Economic perspectives*, 19(1):9–32.

Ambrus, A. and Greiner, B. (2012). Imperfect public monitoring with costly punishment: An experimental study. *American Economic Review*, 102(7):3317–32.

Axelrod, R. (1981). The emergence of cooperation among egoists. *American Journal of Political Science*, 75(2):306–318.

Brandts, J. and Cooper, D. J. (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.

Brandts, J., Ellman, M., and Charness, G. (2015). LetâĂŹs talk: How communication affects contract design. *Journal of the European Economic Association*, 14(4):943–974.

Cason, T. N. and Mui, V.-L. (2007). Communication and coordination in the laboratory collective resistance game. *Experimental Economics*, 10(3):251–267.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, 14(1):47–83.

Coleman, P. T., Deutsch, M., and Marcus, E. C. (2014). *The handbook of conflict resolution: Theory and practice*. John Wiley Sons.

Cooper, D. and Kagel, J. (2016). Other regarding preferences: A selective survey of experimental results. *The handbook of experimental economics*, 2.

Cooper, R., DeJong, D. V., Forsythe, R., and Ross, T. W. (1992). Communication in coordination games. *The Quarterly Journal of Economics*, 107(2):739–771.

De Dreu, C. K., Gelfand, M. J., et al. (2008). *The psychology of conflict and conflict management in organizations*. Lawrence Erlbaum Associates New York.

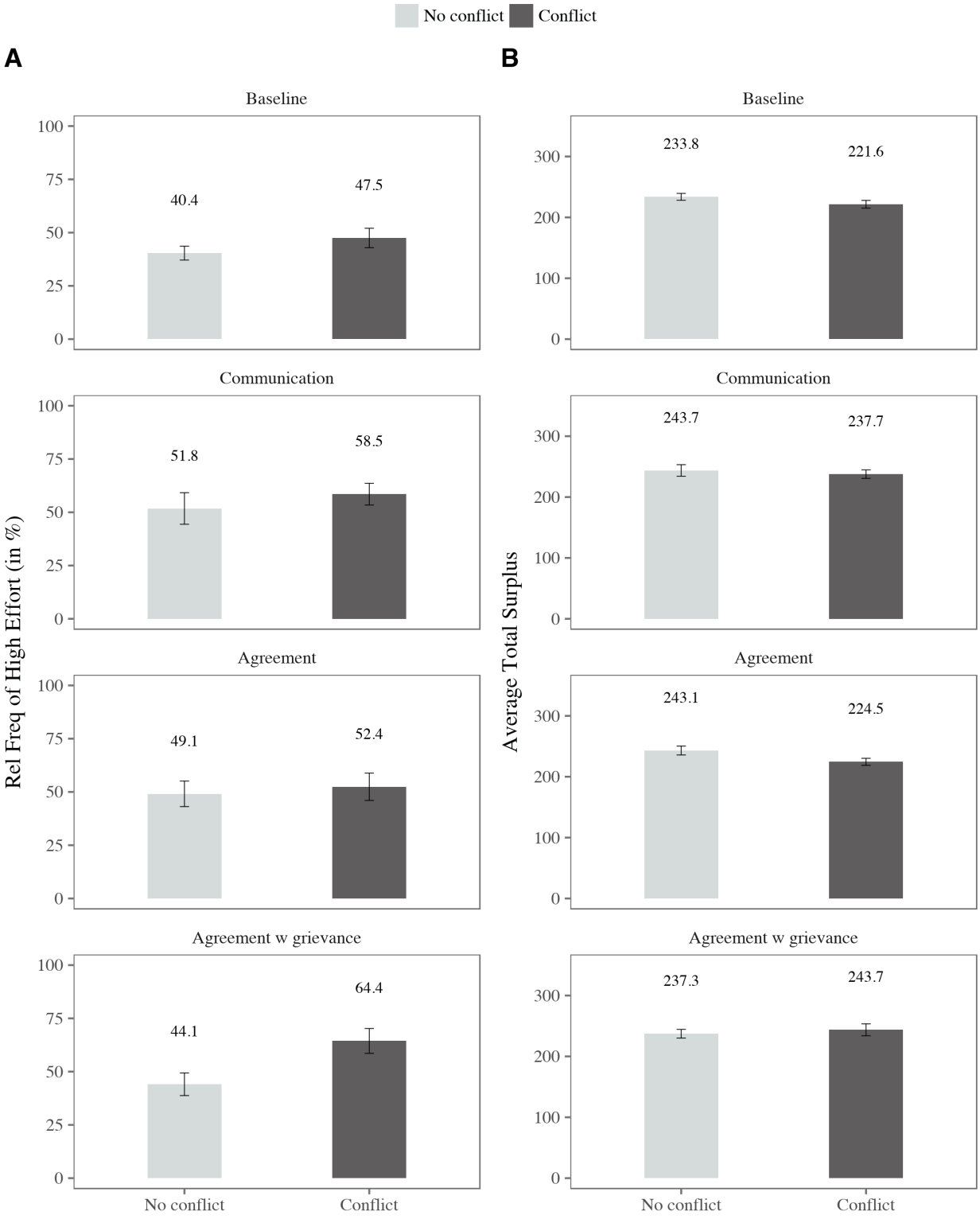
Dreber, A., Rand, D. G., Fudenberg, D., and Nowak, M. A. (2008). Winners donâĂŹt punish. *Nature*, 452(7185):348.

Egas, M. and Riedl, A. (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society B: Biological Sciences*, 275(1637):871–878.

Fehr, E. and Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425(6960):785.

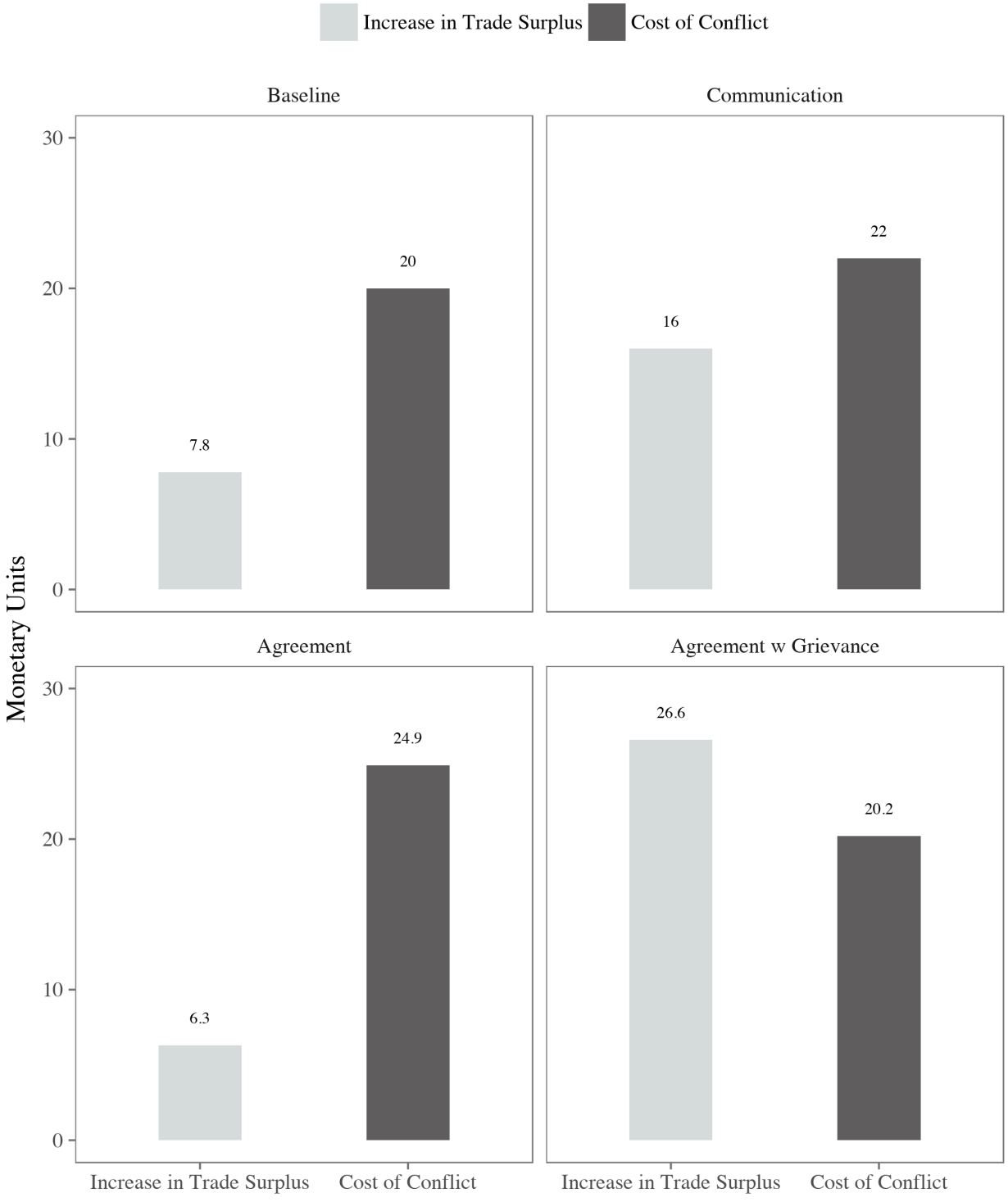
Fehr, E. and Fischbacher, U. (2004a). Social norms and human cooperation. *Trends in cognitive sciences*, 8(4):185–190.

FIGURE 3.1. Effort and surplus



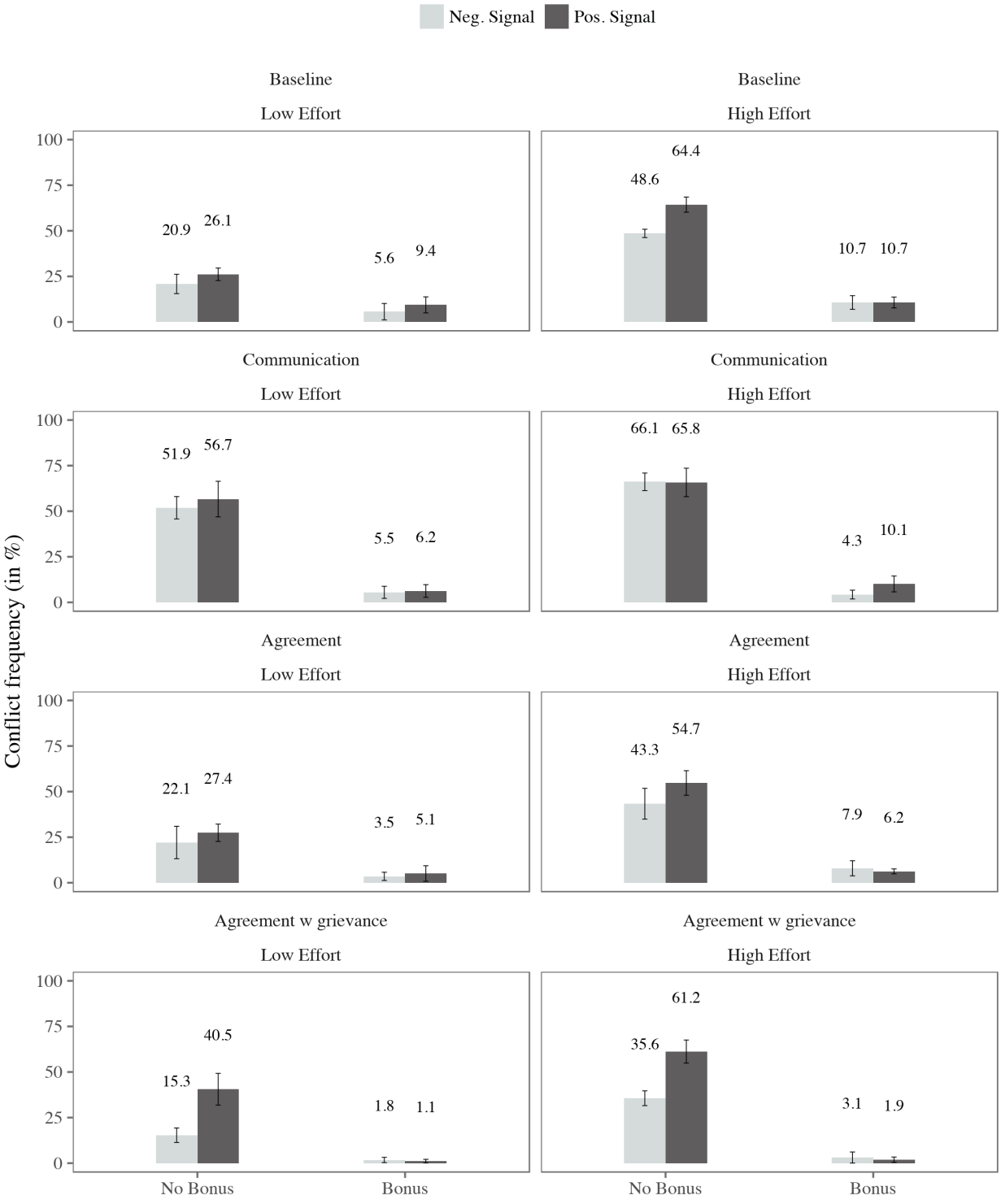
Note: Error bars represent plus/minus one standard error of the mean, clustered at the session level.

FIGURE 3.2. Gains and costs of conflict



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

FIGURE 3.3. Determinants of conflict



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

Fehr, E. and Fischbacher, U. (2004b). Third-party punishment and social norms. *Evolution and human behavior*, 25(2):63–87.

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

Gächter, S. and Fehr, E. (2000). Cooperation and Punishment in Public Goods Experiments. *American Economic Review*, 90(4):980–994.

Gächter, S., Renner, E., and Sefton, M. (2008). The long-run benefits of punishment. *Science*, 322(5907):1510–1510.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95(1):384–394.

Gneezy, U., Rockenbach, B., and Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior and Organization*, 93:293–300.

Grechenig, K., Nicklisch, A., and Thöni, C. (2010). Punishment despite reasonable doubtâĂŤa public goods experiment with sanctions under uncertainty. *Journal of Empirical Legal Studies*, 7(4):847–867.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125.

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, 3(4):367–388.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of business*, pages S285–S300.

Lazear, E. (1986). Salaries and piece rates. *The Journal of Business*, 59(3):405–31.

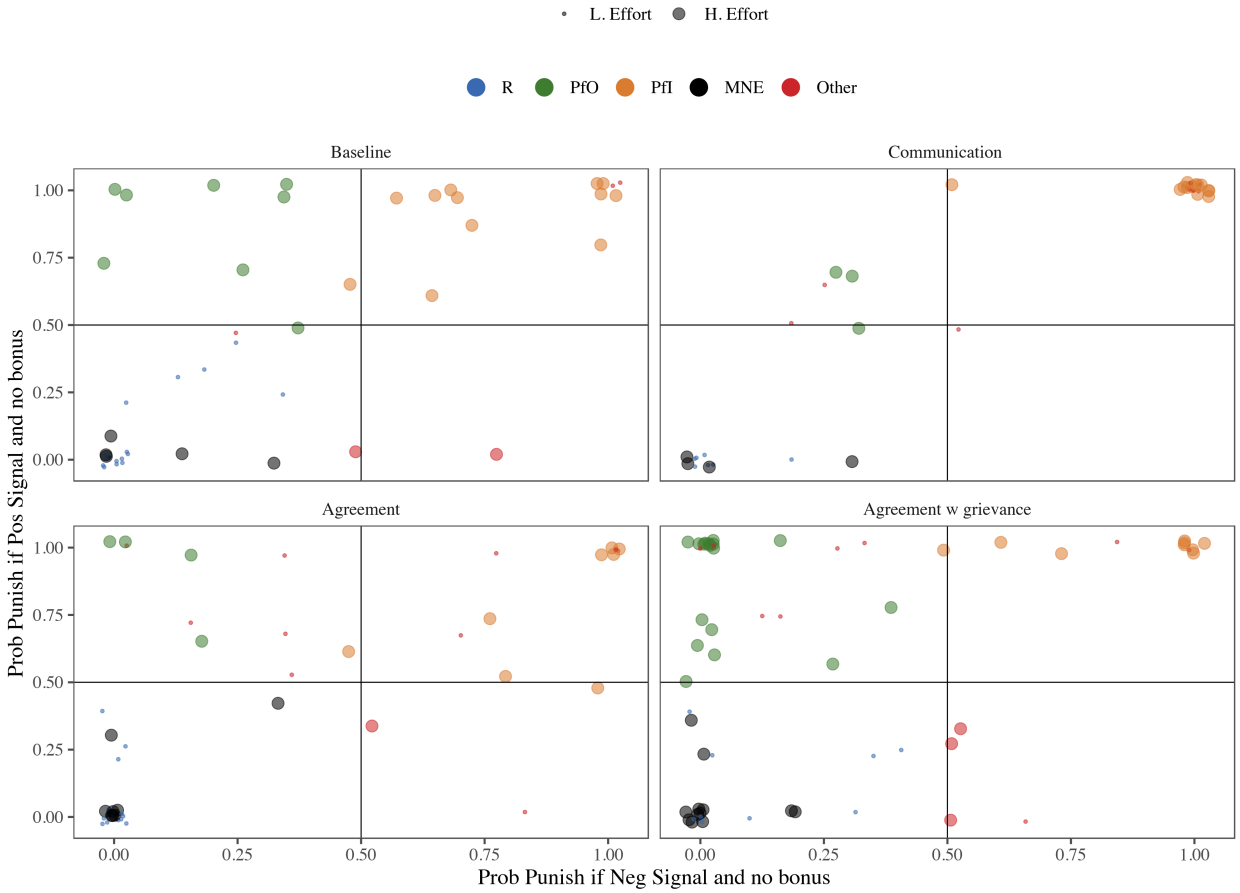
Lundquist, T., Ellingsen, T., Gribbe, E., and Johannesson, M. (2009). The aversion to lying. *Journal of Economic Behavior and Organization*, 70(1-2):81–92.

Macleod, W. B. (2003). Optimal contracting with subjective evaluation. *American Economic Review*, 93(1):216–240.

Ostrom, E., Walker, J., and Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *American political science Review*, 86(2):404–417.

Tjosvold, D., Wong, A. S., and Feng Chen, N. Y. (2014). Constructively managing conflicts in organizations. *Annu. Rev. Organ. Psychol. Organ. Behav.*, 1(1):545–568.

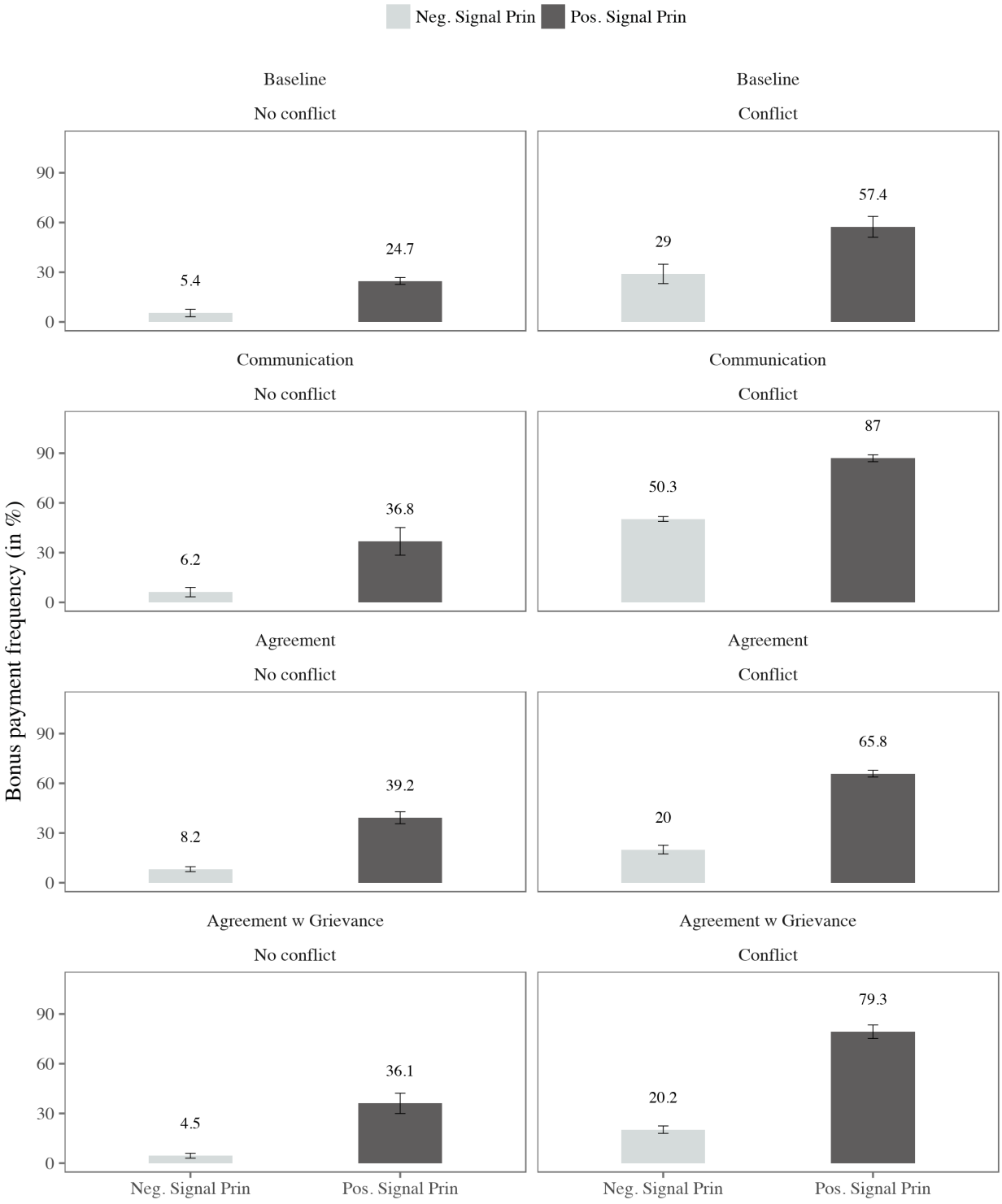
FIGURE 3.4. worker's effort and conflict decisions averaged at the individual level



	R	PfO	PfI	MNE	Other
Baseline	36.2	17	25.5	10.6	10.6
Communication	19.4	8.3	36.1	11.1	25
Agreement	32.6	8.7	17.4	17.4	23.9
Agreement w grievance	14.8	27.9	14.8	19.7	23

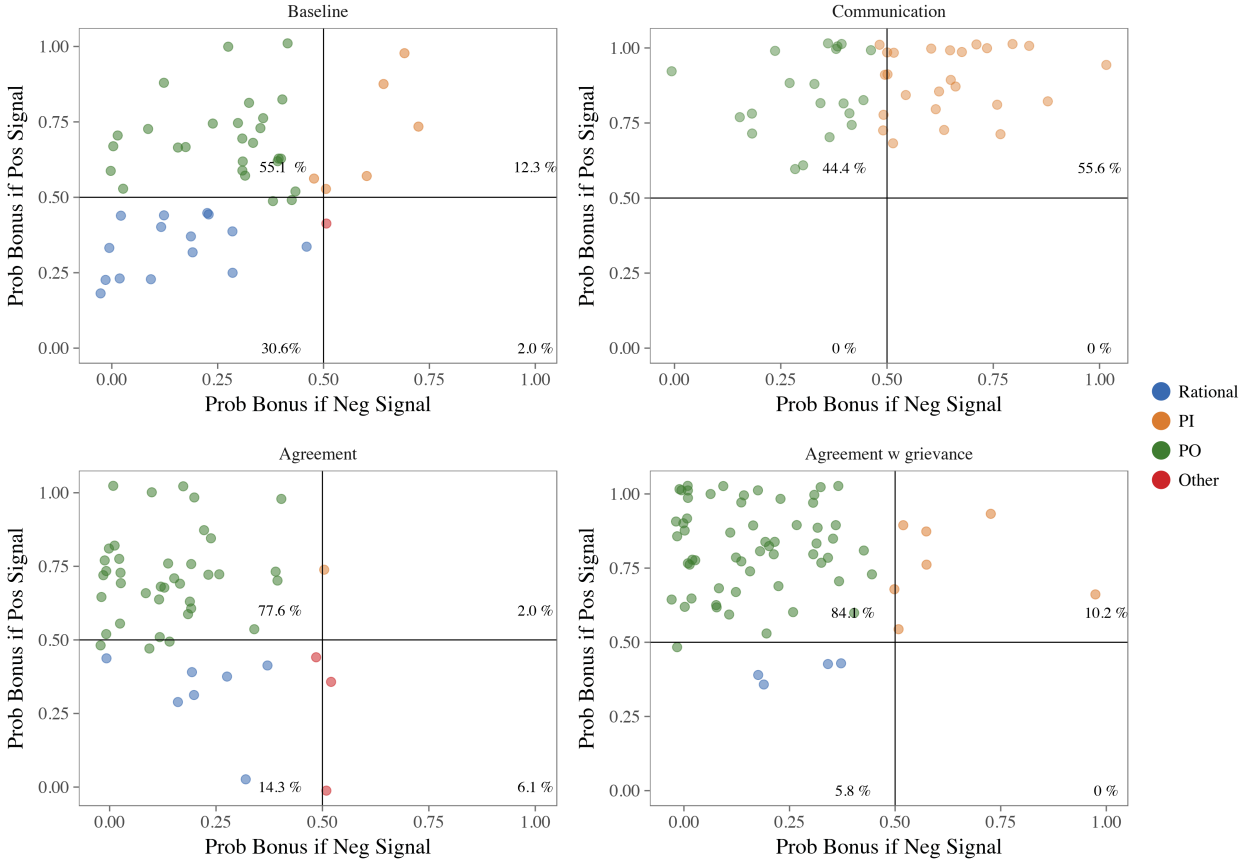
Note: In this graph each dot represents one worker in each treatment. The dots give information on (1) the level of effort, represented by size of the dot; and (2) the conflict behavior, represented by the position of the dot in the graph. The space of the graph is determined by the probability of conflict 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.

FIGURE 3.5. Determinants of bonus



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

FIGURE 3.6. 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.