Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Can Contracts Signal Social Norms? Experimental Evidence

Anastasia Danilov, Dirk Sliwka

To cite this article:
Anastasia Danilov, Dirk Sliwka (2017) Can Contracts Signal Social Norms? Experimental Evidence. Management Science 63(2):459-476. http://dx.doi.org/10.1287/mnsc.2015.2336

http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Vol. 63, No. 2, February 2017, pp. 459–476
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Can Contracts Signal Social Norms? Experimental Evidence

Anastasia Danilov,[a] **Dirk Sliwka** [a]

[a] Faculty of Management, Economics, and Social Sciences, University of Cologne, D-50923 Cologne, Germany
Contact: danilov@wiso.uni-koeln.de (AD); dirk.sliwka@uni-koeln.de (DS)

Received: December 16, 2013
Accepted: April 29, 2015
Published Online in Articles in Advance: February 29, 2016

https://doi.org/10.1287/mnsc.2015.2336

Abstract. We investigate whether incentive schemes signal social norms and thus affect behavior beyond their direct economic consequences. A one-shot principal–agent experiment is studied where prior to contract choice principals are informed about the past actions of other agents and thus have more information about norms of behavior. Compared with a setting in which principals are uninformed, agents exert substantially higher effort under a fixed wage contract when they are aware that an informed principal chose this contract. The informed principal's choice apparently signals a norm not to exploit trust, which leads to more trustworthy behavior. This mechanism's robustness is explored in further experiments.

History: Accepted by Uri Gneezy, behavioral economics.
Funding: The authors thank the German Research Foundation for financial support through the research unit "Design & Behavior" [FOR 1371].
Supplemental Material: Data and the online appendix are available at https://doi.org/10.1287/mnsc.2015.2336.

Keywords: social norms • shirking • contracts • incentives • signaling • experiment • trust

1. Introduction

In recent years, the notion that social norms matter for behavior has gained considerable attention in economics.[1] Indeed, there is now substantial evidence that individuals are influenced in their choices by the observed behavior of others in an identical situation.[2] Many individuals tend to avoid deviations from prevailing norms of behavior, for instance, as these deviations may cause negative emotions such as remorse or shame.

However, individuals often confront a situation in which there is uncertainty about prevailing norms. Consider, for example, an employee who has just joined an organization and may be uncertain about expected effort, working time, private Internet use in the office, or the extent to which she is expected to support colleagues. A very natural reaction for this employee would be to gather information about the behavior of colleagues, which enables her to detect a potential norm of conduct. This may be easy for observable actions (such as working time) but difficult for unobservable actions (such as productively spent working time) that are crucial for the performance of the organization. Even after several years in the same organization, this employee might be unable to assess the behavior of her colleagues with complete certainty in some situations, and she may have to rely on additional information or clues.

On the other hand, owners or managers often have means, such as active monitoring and accounting systems, key performance figures, or employee surveys, to gain a deeper understanding about existing work norms and attitudes in their organizations. Even when direct information about individual behavior is not available, they may be able to infer "average behavior" from these other sources. When designing management tools, such as incentive schemes or monitoring technologies, managers may naturally use this information about observed behavioral patterns. For instance, if a principal observes an underprovision of effort, she may choose to use higher-powered incentives or impose tighter monitoring. But this could lead to an important effect that may sometimes be overlooked: that such interventions convey information about prevailing behavioral norms in an organization—and this, in turn, can have an indirect effect on employees' actions as their perceptions about the behavior of others are altered. Indeed, Sliwka (2007), Friebel and Schnedler (2011), van der Weele (2012), and Bénabou and Tirole (2012) have shown in formal economic models that contract choices may signal information about the actions of other agents and thus create indirect effects on behavior. In a field experiment, Gneezy and Rustichini (2000a) found that introducing a fine in child-care facilities for picking up children late increased the number of parents who came late, arguing that this makes late pickups more acceptable.[3]

We explore the idea that contracts can signal social norms in a set of laboratory experiments. Our key mechanism is most closely related to the theoretical approach by Sliwka (2007). Suppose that agents have

a preference for conformity, because their behavior is influenced by their beliefs about the behavior of others. Conformists act prosocially if they believe many other agents also do so. If a principal who has more information about the distribution of types in an organization now proposes a specific compensation contract, her choice may reveal information about the behavior of others and thus the prevailing norms in the organization. In particular, when a principal proposes a pure fixed wage, she is apparently confident that most agents will not shirk—and in turn that conformists' inclination to shirk should be reduced. On the other hand, the choice of performance-contingent pay or a tight monitoring scheme may reveal the principal's pessimism about the behavior of the agents—and in turn increase conformists' willingness to act more selfishly.[4] Because of the presence of selfish individuals, signaling a strong work norm by choosing a fixed wage or not using a monitoring technology is costly to the principal, and this can indeed make the signal credible.

To study this idea and its implications in detail, we conducted several lab experiments. In our first experiment, we implemented a very simple one-shot principal–agent game. Our main treatment variation was as follows: In the *Baseline* treatment, principals could choose between a fixed wage and performance-contingent compensation. Each principal was matched to an agent, who then chose his level of effort. We elicited the agents' efforts for both forms of compensation using the strategy method.[5] In the *Norms* treatment, we replicated this Baseline treatment with one addition: we showed the principals a table containing the efforts chosen by agents in a preceding Baseline session and informed the agents that their principals had seen such a table (without showing the agents its content). Hence, the treatment intervention was rather weak on the agents' side: they did not have more specific information about the behavior of others but knew that the principals had this information prior to the contract choice. We studied the effect of this intervention in two settings, varying whether or not the choice of the performance-contingent contract was costly for the principal. In the costly contract choice setting, principals had to bear a direct fixed cost for choosing the performance-contingent contract, which may affect the signaling value of the contract choice, as the analysis of a formal signaling model shows.

It turns out that Norms treatment variation has a substantial effect on chosen efforts under the fixed wage. In the Norms treatments, average effort increases by 25% (if the performance-contingent contract comes with no costs) and by 42% (if it is costly) compared with the Baseline treatments in each setting. In other words, agents become much more trustworthy when they know that the principal who decided not to use the performance-contingent contract made this decision being well informed about the behavior of other agents in the same situation. But we do not find evidence that the exogenous variation in the relative costs of the contracts affects the signaling value of the contract choice.

The mechanism described in the above rests on three arguments. First, the principals' contract choice must be affected by the observed information about past behavior of other agents. Second, the contract choice in turn must affect the agents' beliefs about the behavior of others. Third, changes in beliefs about the behavior of others must affect the agents' own behavior. We conducted further experiments and extended our experimental design to test these elements of the proposed theoretical mechanism in more detail, showing that (i) agents' beliefs about the behavior of others are indeed substantially affected by an informed principal's contract choice, (ii) principals vary their contract choices depending on information they receive about agents' behavior in a previous experiment, and (iii) agents choose different actions when informed about the selected choices of other agents in the same situation.[6]

Our study is related to other recent contributions on the interaction of social norms and contracts. According to the model by Bénabou and Tirole (2012), norms do not arise because of preferences for conformity but because the behavior of others influences how publicly observed actions affect social esteem. In their model, agents differ with respect to their intrinsic motivation to choose a certain prosocial action and have a preference to be esteemed, i.e., that others perceive them to be intrinsically motivated. An observer's perception of a certain act depends now on the equilibrium strategies all agents in the population choose—hence, social norms arise because observed actions have different signaling values that are conditional on the strategies of other agents. Similar to Sliwka (2007), changes in extrinsic incentives may reveal a designer's private information on the distribution of types of agents and therefore affect the way in which outside observers interpret the chosen actions. In the Friebel and Schnedler (2011) and van der Weele (2012) models, there is a complementarity between efforts of different agents, and therefore, information about the behavior of others is directly valuable to improving coordination. Galbiati et al. (2013) studied behavior in a twice-repeated "weakest link" coordination game experiment with technological complementarities comparing sanctions that were exogenously imposed after the first round, unconditional on previous behavior, to sanctions that were endogenously imposed by a subject who had observed previous

behavior and benefitted from high levels of coordination. They found that players who made high contributions in the first round contributed less under endogenous sanctions in the second round.[7]

In our setup, the behavior of agents is not publicly observable; there is no interdependence in production between the agents, and the agents themselves do not observe the behavior of others. We show that contract choices reveal information on norms, and this matters for behavior even when individual choices remain unobservable and in the absence of any technological interdependence. Hence, the observed effects can neither be driven by image concerns nor by technological complementarities but are well in line with the idea that people can intrinsically prefer norm compliance.

The remainder of the paper is organized as follows: In the next section we present the design and procedures of our core experimental setting. In Section 3 we derive our hypothesis. In Section 4 we report results of our main treatments. In Section 5 we explore the underlying mechanism in more detail. In Section 6 we study the restriction game. Section 7 concludes.

2. Experimental Design and Procedures

Our baseline design is a simple, *one-shot* principal–agent game. At the beginning of the experiment, all participants receive an endowment of €6. Participants are randomly matched in pairs, where one subject is assigned the role of a principal (labeled *employer*) and the other to the role of an agent (*employee*). The principal chooses between a fixed wage (labeled *trust compensation*) and performance-contingent pay (*contingent compensation*) for her agent. The agent chooses an effort level $e \in [0, 100]$ at private costs of $c(e) = e^2/1{,}200$. The agent's effort level determines the probability that the principal will receive a high payoff, i.e., with probability e she will earn €12 and nothing otherwise. The agents know that the principals cannot observe their efforts but rather only the project's success. Under the trust compensation, the principal pays an unconditional wage of €5. Under the contingent compensation, the agent receives €5 only if the principal earns the high payoff of €12 and nothing otherwise.[8]

Efforts for both contract types are elicited using the strategy method, such that each agent has to state an effort level for each of the two compensation schemes before learning about the principal's choice. After all choices are made, the payoffs are computed based on the respective choices made by the principal and agent.

Our main treatment variation is the following: We compare Baseline treatments, in which the game is played as described above, with Norms treatments. There is only one difference between these two treatment types: on the decision screen of principals in the Norms treatments, we introduce an additional table showing the real decisions of 10 agents from a previous session of the Baseline treatment. The agents know the principals have this information but do not know its content.[9] The payoff functions for both principal and agent remain unchanged. Hence, from the agents' perspective, the treatment intervention varies only the fact that the principals are ex ante better informed about the behavior of other agents in the same population. Thus, any changes in agents' behavior must be driven by their awareness that the principals had more information prior to the contract choice.

We extend our experimental setup to a 2×2 between-subjects design and study the question of whether having an informed principal also affects agents' behavior in a setting where the performance-contingent contract implies additional costs to a principal. The difference between the Costless and Costly treatments is only that, in the latter setting, principals have to bear additional costs of €2 when choosing the performance-contingent compensation, and the agents are aware of this. All other parameters remain unchanged. The reasons for studying the Costly treatments are twofold: First, it provides an additional robustness test. In the Costless treatment, the choice of a fixed wage is rather risky and potentially inferior from a principal's perspective, whereas the contingent contract only requires a payment to the agent in case of success. The Costly treatment increases the relative advantage of the fixed wage contract. Second, it may allow testing a further implication of the signaling model: as the costs of choosing the contingent contract change the relative attractiveness of the two contracts, it can affect their signaling value and thus the efforts chosen by the agents. We explain this in more detail in Section 4.

Because we study a one-shot decision situation, it was important for us to make sure that the participants understood the instructions and the decision situation as well as possible. Before proceeding to the decision stage, subjects had to complete a short quiz on the structure of the experiment and the computation of payoffs. Only after answering correctly could they proceed to the decision stage. In the decision stage, we provided agents with an on-screen calculator (see Figure D1 in Online Appendix D). The agents could use it to compute the expected payoffs from any effort level for themselves and the principal under each compensation form. The calculator was used by 89.6% of agents, who pushed the button an average of seven times.

The experiment consisted of 25 sessions with 20–32 subjects each.[10] We observe 80–93 principal–agent dyads per treatment (see Table A.1 in Appendix A for more details). All sessions were conducted in the Laboratory for Experimental Research of the University of Cologne, using the experimental software z-Tree (Fischbacher 2007). A total of 691 participants (56% females and mostly students) were recruited

via ORSEE (Online Recruitment System for Economic Experiments; see Greiner 2004). All subjects participated only once. All decisions were anonymous and no communication was permitted during the experiment. Average earnings (including a €2.50 show-up fee) were €12.55. The sessions lasted about one hour.

3. Key Hypothesis

Our key hypothesis is that agents react differently to an identical contract when they know it has been chosen by a principal who is informed about the behavior of other agents in the same situation. In particular, we expect agents to choose a higher effort level under the fixed wage contract if an informed principal proposes it. In Appendix B, we analyze a formal signaling model supporting this claim. This mechanism is based on two assumptions: that (i) agents differ in their prosocial preferences, and (ii) a subset of the agents is influenced by social norms in the sense that their prosociality depends on their beliefs about the average prosociality of the others. If the principal is then informed about behavior of other agents in the same situation, her contract choice can indeed reveal information about the social norm of behavior and in turn affect agents' choices.

The key idea of this mechanism is the following: when the principal has observed that many agents in the relevant population are selfish and shirk under the fixed wage contract, offering this contract is very costly given the high likelihood that the agent she is matched with is also selfish and will shirk. If, however, the principal has observed that there are many prosocial agents, i.e., agents who exert high efforts under the fixed wage, in the population, shirking will be less likely. Thus, the choice of a fixed wage contract is relatively less costly when the principal has observed more prosocial behavior, and offering a fixed wage is therefore more attractive. The contract choice thus becomes a credible signal about the social norm. If conformity matters, agents adapt their behavior: the choice of a fixed wage by an informed principal should lead to higher efforts by conformists. The choice of the contingent compensation, on the other hand, may reveal that the principal apparently observed more selfish behavior, and in turn, the agents can become more selfish.

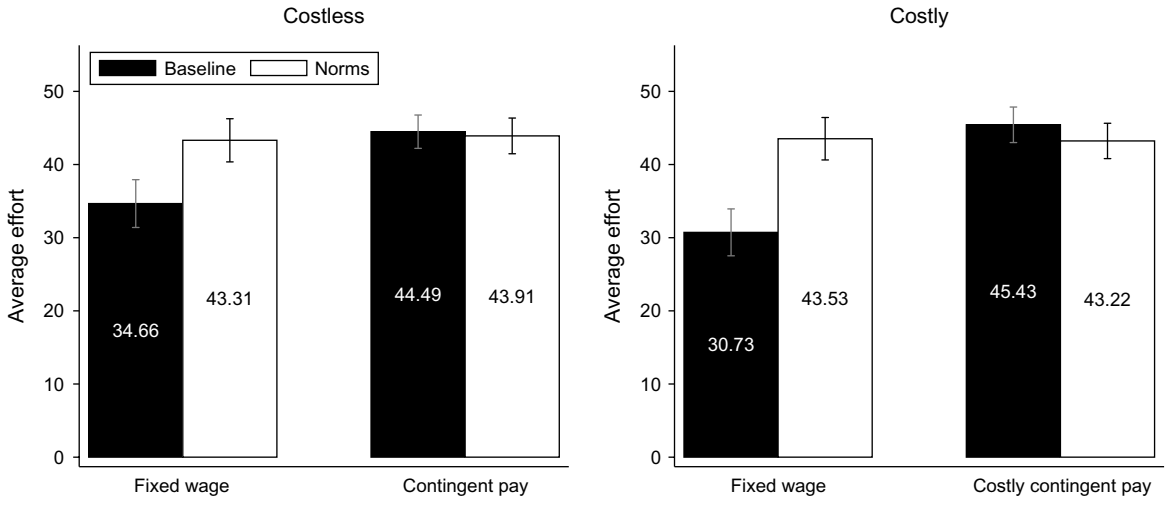
Since the contingent contract provides extrinsic incentives, even purely selfish agents have a reason to work under this contract type. Thus, the agents' efforts should be less elastic to information about the social norm, and the negative effect of the Norms intervention on efforts should be weaker under this contract.

4. Results

We start by investigating average effort levels in all four treatments for both contract types. As Figure 1 shows, the agents' effort reaction to the fixed wage contract is indeed strongly affected by the Norms intervention: if agents know that the principal is informed about the behavior of other agents prior to her contract choice, they exert substantially higher efforts under the fixed wage contract—and this is the case in both the Costless and the Costly contract choice treatments. The average effort increases by 25% in the Costless contract choice treatment ($p = 0.0575$, two-sided Mann–Whitney U (MWU) test) and by 42% in the Costly contract choice treatment ($p = 0.0034$). Hence, in line with our key hypothesis, the mere fact that the agents know that the principal had been informed about the behavior of others before the contract choice substantially alters their reaction to a fixed wage.

Under the contingent contract, efforts are hardly affected by the Norms intervention. Recall that the theory predicted a negative effect as an informed

Figure 1. Average Efforts and Standard Errors of Means



principal's choice of a contingent contract entails a negative signal about the prosociality of the agent's population. The theory also suggested that the negative effect of the Norms intervention under the contingent contract is weaker than the positive effect under the trust contract, because the incentives provided by the contingent contract do not rely on the norm-driven prosociality of the agents. Efforts are indeed slightly smaller in the Norms treatments—by 1.3% in the Costless setting and by 4.8% in the Costly setting. However, these differences are insignificant ($p = 0.8087$ in Costless and $p = 0.5031$ in Costly treatments, two-sided MWU tests). Efforts are thus inelastic to norms signaling under the contingent contract but strongly elastic under the fixed wage contract. The between treatment difference-in-difference in efforts under the two contract types ($p = 0.0552$ in Costless and $p = 0.0004$ in Costly treatments, two-sided MWU test) is thus nearly entirely driven by an increase in efforts under the fixed wage.[11]

The econometric analysis reported in Table 1 confirms these patterns. Models (1) and (2) regress the effort separately for the Costless and Costly treatments on a dummy indicating the Norms intervention and demographic controls. Model (3) is based on the pooled data across the Costly and Costless treatments. In all specifications, the difference-in-difference, as measured by the interaction term *Trust contract* × *Norms treatment*, is positive and significant. The trust contract leads to substantially higher efforts relative to the contingent contract in the settings where the agents know that the principal is informed about behavior of others.

Table 1. Regression Results: Effect of the Norms Intervention on Effort

Dependent variable: *Effort*	Model (1) Costless	Model (2) Costly	Model (3) Pooled
Fixed wage	−9.83***	−14.70***	−9.83***
	(2.98)	(3.13)	(2.98)
Norms treatment	−0.88	−2.06	−0.57
	(3.39)	(3.42)	(3.38)
Fixed wage × *Norms treatment*	9.22**	15.00***	9.22**
	(4.32)	(4.22)	(4.30)
Costly			0.80
			(3.40)
Costly × *Fixed wage*			−4.88
			(4.31)
Costly × *Norms treatment*			−1.58
			(4.85)
Costly × *Fixed wage* × *Norms treatment*			5.78
			(6.02)
Female	1.58	4.06	2.27
	(3.28)	(3.41)	(2.33)
Age	−0.15	0.63	0.10
	(0.42)	(0.69)	(0.42)
Constant	47.51***	28.42*	41.04***
	(10.37)	(16.45)	(10.43)
Observations	370	322	692
R-squared	0.02	0.06	0.04

Notes. Ordinary least squares (OLS) regressions. The data consist of two effort decisions per subject: under the fixed wage and under the performance-contingent pay. Robust standard errors clustered on the subject level are reported in parentheses.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

Comparing the Costly and Costless contract choice settings, we do not find significant differences in the average effort reaction. All p-values for nonparametric effort comparisons between the Costly and Costless settings are above 0.50 (two-sided MWU tests; see also Model (3) in Table 1 for the regression analysis). Average effort under the fixed wage is lower in the Costly Baseline setting than in the Costless Baseline setting, but this difference is also not significant ($p = 0.5087$, two-sided MWU test).

One potential explanation for the absence of a treatment difference between the Costly and Costless settings is that there are countervailing effects in our formal model reported in Appendix B. We show that when the contingent contract is costly, separating equilibria (in which the choice of the fixed wage is informative about the norm) exist under weaker conditions about the distribution of types. Intuitively, because the trust contract becomes relatively more attractive when the contingent contract is costly, the likelihood that it is actually chosen on the equilibrium path (which is a precondition for credible signaling) should increase.[12] But, on the other hand, the size of the effort effect as a result of norms signaling should be weaker as the choice of a fixed wage is a weaker signal about the average prosociality in the population.[13] There is some indication for these patterns in the data, as in the Costly contract choice treatment the effort effect of the Norms intervention is driven by an increase in the *proportion* of agents who exert strictly more effort under the fixed wage contract, whereas in the Costless contract choice treatment it is rather driven by an increase in the *level* of efforts chosen by the agents who exert strictly more effort under the fixed wage contract.[14] However, we also acknowledge that the inference process about what exactly the contract choice reveals about the social norm when the principal may act strategically requires a higher level of common knowledge of rationality. In other words, the idea underlying the main result—that "if the principal offers a fixed wage, he cannot have seen many shirking agents"—does not require a very sophisticated reasoning. Preferences for conformity then directly imply a positive effort reaction. But the question "what exactly does the contract choice tell me about the extent of shirking given that the principal chose the contract strategically?" seems to be more difficult.

Figure 2. Effort Distributions

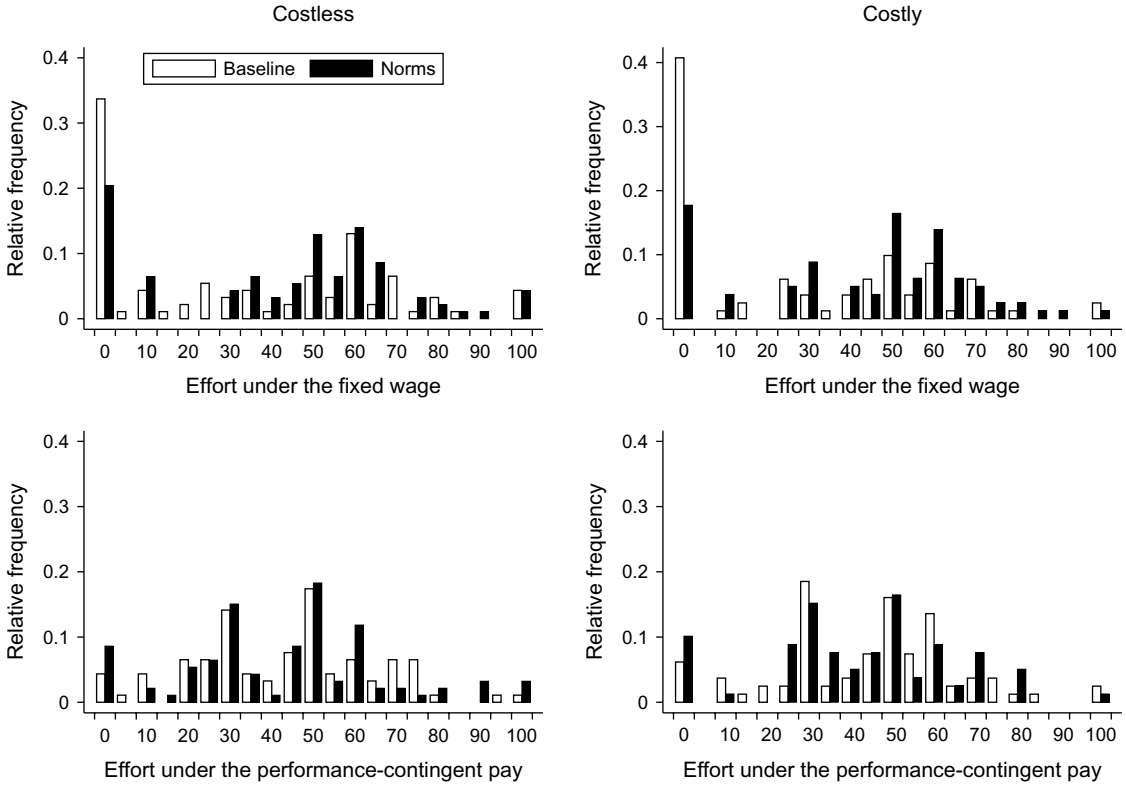


Figure 2 shows histograms of the effort choices for the fixed wage (upper panels) and the contingent wage (lower panels) in bins of five effort units. The most striking difference is the substantial reduction in the very low efforts (0–4) under the fixed wage in the Norms treatments (black bars) compared with the baseline setting (white bars). For example, in the Costly contract choice setting, the fraction of agents choosing these very low efforts drops from about 40.7% to 17.5% in the Norms treatment ($p = 0.0012$, two-sided proportion test). In the Costless treatment, the proportion of efforts in the lowest category also drops significantly from 33.7% in the Baseline to 20.4% in the Norms treatment ($p = 0.0422$). Hence, many of the otherwise very selfish agents apparently choose to exert substantially higher efforts when they know that an informed principal had chosen the fixed wage. In turn, the effort variance under the fixed wage is lower in the Norms treatments compared with the Baseline, both in the Costless setting and in the Costly setting (the respective p-values are 0.0583 and 0.0413, Levene's robust test of variances; see Table A.2 in Appendix A for more details). One interpretation of this result is that norms signaling leads to more consistent behavior of the agents.[15]

Finally, we turn our attention to the profits of principals and overall welfare. In line with the effort effect, the Norms intervention leads to significantly higher profits under the fixed wage, with an increase from €5.15 to €6.20 in the Costless and from €4.69 to €6.22 in the Costly contract choice treatment ($p = 0.0575$ and $p = 0.0034$, respectively, two-sided MWU test; see Table A.3 in Appendix A for more details). However, this is still less than the profits resulting from the contingent contract (€9.11 and €7.18, $p = 0.0000$ and $p = 0.0724$, respectively, two-sided Wilcoxon signed-rank (WSR) test). The generated overall welfare is significantly smaller under a fixed wage than under a contingent contract in the Costless Baseline treatment (€14.35 versus €15.29, $p = 0.0003$), but the fixed wage achieves a nearly identical total welfare level to the contingent compensation in the Norms treatment (€14.97 versus €15.21, $p = 0.7672$, two-sided WSR test). In the Costly contract choice treatments, however, the fixed wage leads to a significantly higher welfare than the contingent contract (€14.22 and €13.33 with $p = 0.0001$ in Baseline, and €15.09 and €13.25 with $p = 0.0000$ in Norms).

The principal's contract choice should depend on the information about the behavior of agents. To be able to analyze the choices of principals, it is important to have variation in the information principals receive about norms of behavior. We address this issue in Section 5.2 in detail, introducing a new design element in a subset of the experimental sessions.

5. Disentangling the Signaling Mechanism

The main results from the experiment seem to be well in line with the hypothesis that contracts can signal

social norms and, in turn, affect behavior beyond the direct incentive effects. In particular, the mere fact that agents know that the contract is chosen by a principal informed about the behavior of others substantially increases efforts under a fixed wage. But, of course, there may be additional mechanisms that drive the observed behavioral patterns. One potential alternative mechanism could be guilt aversion (see, for instance, Battigalli and Dufwenberg 2007 or Ellingsen et al. 2010). Suppose that, in line with the idea of guilt aversion, agents choose higher efforts when they believe the principal expects them to do so. If a principal now proposes a fixed wage only when she has high expectations about the efforts exerted under that form of compensation, the choice of the fixed wage can reveal these expectations and indeed trigger higher efforts. However, this reasoning alone cannot explain why efforts under the fixed wage are higher in the Norms treatment, as the principal's expectations are similarly revealed in the Baseline treatment. Hence, it is apparently important that the principal has information about the behavior of other agents, as this triggers the additional behavioral response. Further, even though principals do not observe agents' efforts directly, agents may perceive that principals compare them to other agents in the Norms treatments and therefore exert higher efforts. A mere comparison effect should lead to higher efforts also under the contingent contract. However, we do not observe that efforts increase under the contingent compensation when an informed principal proposes it.

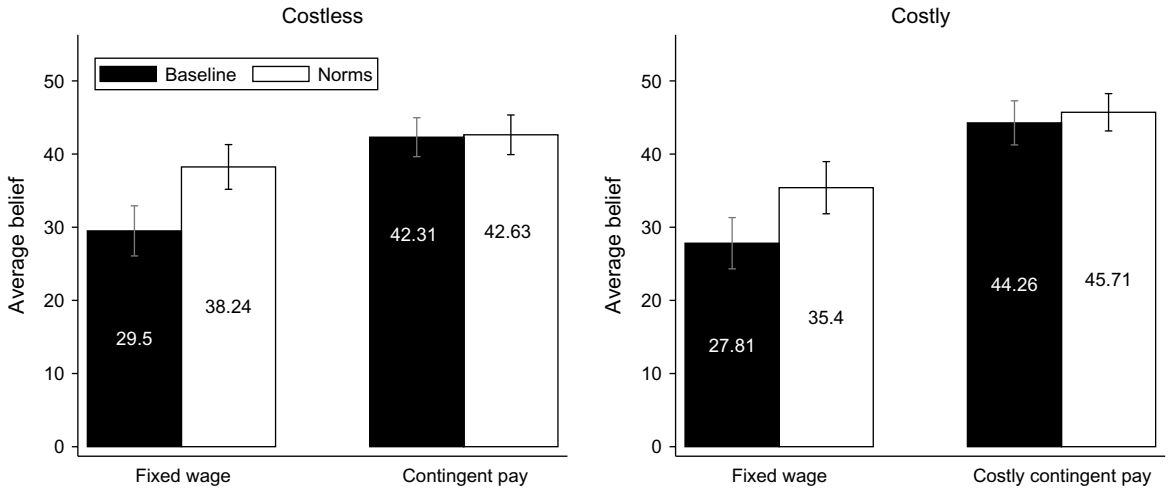
As laid out before, the mechanism suggested by the underlying theory rests on three premises: First, agents must infer information about social norms from principals' choices. Second, a principal's contract choice must be affected by the information observed about past behavior of other agents. And third, information about the behavior of others must affect the agents' own behavior. To see whether these are indeed core driving forces behind our results, we now investigate these three hypotheses individually.

5.1. Agents' Beliefs

An important element in the hypothesized mechanism is that agents adapt their beliefs about the norm of behavior based on the principal's contract choice. We investigate this question in two ways. In the last 16 of the 25 sessions, we elicited agents' beliefs about the behavior of other agents under both contract types after the decision-making stage. We used an incentivized procedure, such that agents received an additional payoff equal to €2 minus €0.01 per unit of the squared deviation between their estimate and the true session mean of efforts.[16] Although we caution that these beliefs are endogenous and potentially driven by false consensus effects or used self-servingly as a justification for prior actions,[17] it is still interesting to study how the contract choice affects beliefs differently in the Norms as compared to the Baseline treatments. As Figure 3 shows, the Norms intervention increases beliefs about efforts under the fixed wage in the Costless and the Costly contract choice settings ($p = 0.0424$ in Costless and $p = 0.0870$ in Costly, two-sided MWU test), but it does not significantly affect beliefs about efforts under the contingent wage ($p > 0.711$).[18] However, the difference-in-difference is not significant in either setting ($p = 0.156$ in Costless and $p = 0.3523$ in Costly).

To avoid false consensus effects and self-serving beliefs, we also conducted a separate (online) experiment that was designed purely to elicit incentivized beliefs about inferences from the principal's choices in our previous experiment. In the *Contingent Contract Beliefs* treatment ($N = 57$) and *Trust Contract Beliefs* treatment ($N = 61$), we investigate to what extent the

Figure 3. Agents' Beliefs and Standard Errors of Means



principal's contract choice affects observers' beliefs about this principal's information on the social norm. In these treatments, subjects were outside observers who received instructions from the Costless Norms treatment and had to estimate the information a principal had observed when selecting one of the two contracts. Subjects were paid for the correctness of their beliefs according to the quadratic scoring rule: they earned €3 minus €0.01 for each unit of quadratic deviation from the respective true value.[19] The experiment was conducted online, and participants were paid via bank transfer or with Amazon.de vouchers. Each subject participated only once and had not taken part in any of our prior treatments.

In the Contingent Contract Beliefs treatment, subjects estimated the average efforts actually observed by a principal who was randomly chosen from those who had selected the contingent contract. In the Trust Contract Beliefs treatment, they estimated the average efforts observed by a principal who had chosen the trust contract. In each of these two treatments, subjects stated two numbers, one for their estimate of the average efforts *observed by this principal* under the fixed wage and one for average efforts observed under the performance-contingent compensation.[20] A comparison between these two treatments allows another test of the idea that the actual contract choice of a principal affects beliefs about this principal's knowledge about the behavior of others, this time by "impartial" outside observers.[21]

The results show that third-party beliefs are affected by contract choices. In the treatment in which participants estimate what the principal had seen before proposing a contingent compensation contract, observers indeed believe that this principal observed significantly higher efforts under the contingent contract (59.28 instead of 49.12, $p = 0.0178$, two-sided WSR test). This picture is reversed when participants estimate what the principal had seen before proposing the fixed wage. Here, subjects expect higher effort under the fixed wage than under the performance-contingent compensation (62.36 instead of 49.13, $p = 0.0018$, two-sided WSR test).

We caution that these third-party beliefs may not be very accurate, given that it is likely difficult for outside observers in an online experiment to not only put themselves in the situation of a subject in the experiment but also make indirect inferences about behavior observed by principals who made a certain contract choice.[22] But the qualitative picture is of interest: contract choices affect inferences made about what principals have seen irrespective of whether they are elicited from directly affected agents or outside observers.

5.2. Principals' Contract Choice Behavior

So far, we have been mainly interested in differences in agents' reaction to the principal's contract choice. But it is also important to see whether principals choose different contracts when the social information varies. An affirmative answer to this question is an important precondition for the suggested mechanism: if principals do not react to information about the norm of behavior, their contract choice cannot reveal such information. To answer this question, we elicited principals' contract choices using the strategy method in six sessions of the Costly contract choice experiment.[23] In total, 92 principals (46 in the Baseline setting and 46 in the Norms setting) went through a different decision procedure than in our initial design, leaving the agents' decision completely unaffected. In both the Baseline and the Norms treatments, principals saw five different tables, each showing a *possible* behavior of 10 agents from a previous session (again containing each agent's efforts under fixed wage and contingent contracts). Each of the tables had exactly the same format as the table used in our initial design. The principals were informed that one of these tables corresponded to actual effort choices from a previous experimental session, but they did not know which one was the "true table" (see Table D7 in Online Appendix D for details).

The tables were designed such that effort levels and thus the relative profitability of the contracts varied between the different scenarios. In two of the five tables, the principals' profits were higher under the contingent contract, and in the other three tables, the fixed wage contract led to higher profits.[24] Principals only saw the vectors of chosen efforts and not the implied profits. To study whether principals used their contract choice strategically, principals were explicitly informed in the Baseline treatment that the agents were unaware that principals had learned the past behavior of other agents. In the Norms treatment, on the other hand, principals knew that the agents' were notified about their superior information—just as in the initial experiment. Hence, in the Norms treatment there was common knowledge that principals had information about social norms of behavior.

The principals were asked to make a contract choice for each of the five tables, which were presented in random order on the screen.[25] For the analysis we ranked tables according to the difference in average profits between the trust and contingent contracts, with the table where the trust contract "outperforms" the contingent contract by the highest amount occupying rank 1. The table with the true efforts from the previous experiment had rank 4.

We ran simple linear probability and probit regressions to study whether principals indeed adapt their contract choice according to information on agents' behavior in the previous experiment. To do that, we pooled the data from both treatments and include a dummy variable indicating whether an observation comes from the Norms treatment, i.e., the treatment

Table 2. Principals' Contract Choice

	Model (1)—OLS Dummy fixed wage	Model (2)—Probit (marginal effects) Pr(*Fixed wage* = 1)
Norms treatment	−0.07	−0.09
	(0.06)	(0.07)
Table rank 1	0.13**	0.13***
	(0.05)	(0.05)
Table rank 2	0.12**	0.12**
	(0.05)	(0.05)
Table rank 4	−0.43***	−0.41***
	(0.06)	(0.05)
Table rank 5	−0.42***	−0.40***
	(0.06)	(0.05)
Female	0.05	0.07
	(0.06)	(0.07)
Age	0.00	0.00
	(0.01)	(0.01)
Constant	0.52**	
	(0.21)	
Observations	460	460
R-squared	0.27	
Pseudo *R*		0.22

Note. Robust standard errors are in parentheses.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$ (two-sided).

where there is common knowledge that the principal has superior information.

As the regressions reported in Table 2 show, principals indeed react strongly to the information about behavior in the previous experiment. Compared with the reference table with rank 3, principals choose the fixed wage significantly more often when observing more prosocial behavior (i.e., for tables with ranks 1 and 2). They chose it significantly less often when observing more selfish behavior, such as in the (actually truthful) table 4 and the even more selfish table 5. Note that the dummy for the Norms treatment is not significantly different from zero. A potential interpretation is that principals do not anticipate that agents choose higher efforts under the fixed wage in the norms treatment. However, we caution that principals might have perceived that the efforts shown in the table (which came from a session from the Baseline treatment) already reflected the effect of the Norms intervention.

5.3. Effects of Explicit Norms Disclosure

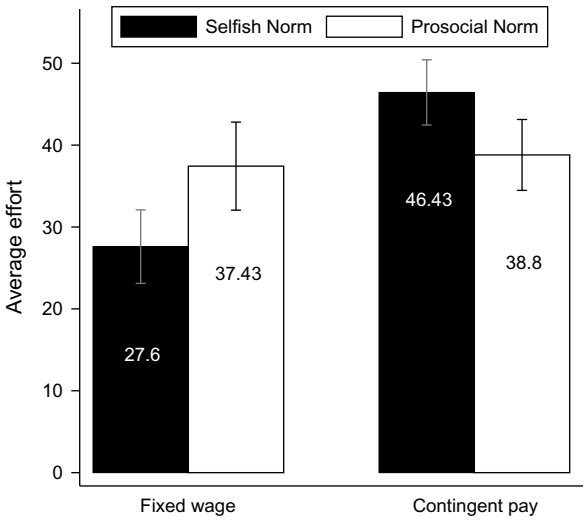
To investigate the last element of the proposed mechanism, we provide a direct test of whether information about others' behavior affects agents' effort choices in our framework. We explore the extent to which explicit information about actions of others affects agents' decisions. Note that our key conjecture is well in line with the existence of "social history" or "conformity effects" in experiments. A large number of experiments have already established that knowledge about others' past actions in the same situation affects behavior (examples include Berg et al. 1995, Bardsley and Sausgruber 2005, Bicchieri and Xiao 2009, Servátka 2009, Gächter et al. 2012, and Gürerk 2013). The purpose of this section is thus not so much to present novel evidence but rather to study whether conformity effects also occur in the setting of our main experiment. To do so, we designed an additional simple experiment in which we induce different norms of behavior in a rather straightforward way.

The design of this experiment is similar to our Costless Baseline treatment. The only difference is that now the principals and not the agents see a table similar to the one in the Norms treatment, with efforts of 10 selected employees from a previous experiment.[26] We ran two different treatments, *Selfish Norm* and *Prosocial Norm*. For each of the two treatments, we selected a different sample of actual effort contributions from the Baseline treatment in our first experiment. In the Prosocial Norm treatment, we displayed a sample of 10 selected agents with very high contributions under the fixed wage. The average effort of the selected sample was 60.1 under the fixed wage and 47.9 under the contingent compensation. To obtain a ceteris paribus comparison, the displayed sample of subjects in the Selfish Norm treatment (also taken from real observations in the Baseline) was very similar with respect to the efforts under the contingent compensation but with substantially lower efforts under the fixed wages. To be specific, the average effort was 47.1 under the contingent contract and 19 under the fixed wage.[27] Hence, we expect subjects to exert higher efforts after the fixed wage in the Prosocial Norm treatment than in the Selfish Norm treatment, whereas the efforts under the contingent compensation should not differ between the treatments.

The principals knew that the agents had seen the decisions of the 10 selected agents from the previous sessions but did not know the efforts themselves, and the agents were aware of this. A total of 120 subjects took part in this experiment (30 in the role of principal and 30 in the role of agent in each treatment). The experiment lasted about one hour. All payments were made individually and anonymously. The average earnings were €10.57 per subject, including a show-up fee of €2.50.

As can be seen from Figure 4, we observe a 36% higher average effort under the fixed wage in the Prosocial Norm treatment (37.43) than in the Selfish Norm treatment (27.6), but this difference is only significant when we use a one-sided test (the two-sided MWU p-value is 0.1092). However, the between-treatment difference in the span of efforts between the contingent compensation and fixed wage is highly significant ($p = 0.0129$, two-sided MWU test). This highly significant difference-in-difference is apparently also par-

Figure 4. Average Efforts and Standard Errors of Means in the Explicit Norms Disclosure Experiment



tially driven by a difference in effort under the contingent compensation (which is 46.43 in the Selfish Norm treatment and 38.8 Prosocial Norm treatment, but this difference is not significant; $p = 0.3396$, two-sided MWU test). Finally, the within-subjects difference in the effort choices for the respective compensation form is highly significant in the Selfish Norm treatment ($p = 0.0044$) but not in the Prosocial Norm treatment ($p = 0.8933$, two-sided WSR test).[28]

To sum up, we indeed find evidence for the three elements of the conjectured mechanism: (i) beliefs about the behavior of others are affected by knowledge about the contract choice of an informed principal, (ii) principals' contract choices are affected by information about behavior in the population, and (iii) information about other agents' choices affects agents' behavior.

5.4. Costs of Control and the Restriction Game

We also apply our experimental manipulation in the context of a different experiment, the "Cost of Control" experiment by Falk and Kosfeld (2006). The results are reported in detail in Online Appendix C. Here, the principal decides whether to impose a minimum on the agent's contribution (i.e., to control the agent) or not (i.e., to trust him). As Falk and Kosfeld have shown, the choice of control can have "hidden costs" and leads to lower efforts. Hence, in contrast to our initial setting, "control" should be detrimental already in the Baseline setting. But the mechanism proposed in this paper suggests that we may observe stronger costs of control when the principal is informed of the behavior of others, as imposing a restriction can reveal that most others behaved very selfishly.

We find a statistically significant effort reduction when control is imposed in the Norms setting where principals are informed about behavior of other agents but not in the Baseline setting were the principal is uninformed. We also find that it again led to a reduction in the variance of behavior. However, there are no significant differences between Baseline and Norms with respect to average efforts. Hence, the Norms intervention here did not lead to economically stronger average costs of control but made the effect statistically more stable because of the increased consistency of agents' behavior.

6. Conclusion

We have shown in a series of lab experiments that contract choices can convey information about the behavior of other agents previously observed by the contract designer, and this information can have an impact on agents' behavior. Individuals may react very differently to an identical contract when they know that its selection is based on richer information about the prior reactions of others. Contract choices thus can reveal information about prevalent social norms and also indirectly shape behavior beyond direct material incentives. We find substantial effort effects in our main contract choice experiments where the use of a fixed wage led to significantly higher efforts when it was chosen by an informed principal.

It is important to stress that in our experiments these signaling effects occur even though agents' behavior is not observed by peers and that ex post they do not even receive information on the distribution of choices. Hence, the mechanism relies on an apparent intrinsic tendency for conformity and not on technological complementarities or image concerns. It is thus applicable to, and should be relevant for, a broader number of contexts—namely, all situations in which a first mover's choice can reveal information about behavior in a broader population, which, in turn, can affect the behavior of second movers beyond their direct economic motives.

Our results also have implications for the design of incentive schemes in practice. A direct implication is that when employees (or citizens) are not well informed about norms of behavior but the designer of an incentive scheme (or a law) is, the choice of the scheme can have signaling effects as it reveals information about prevalent norms.[29] This seems particularly important when uncertainty about the norm is large, such as in newly founded companies or those formed through a merger. A particular view about the set of norms that form a firm's culture may become "self-fulfilling" when this view shapes the design of the incentive structure and thereby reveals itself to the employees. Granting a high degree of autonomy, for instance, can reveal that the employer is convinced that employees will not exploit this trust—and this signals to employees that the exploitation of trust is apparently

not the "usual" behavior in the firm. It may thus reinforce a culture of trust.

Of course, many important questions still need to be addressed. A key challenge is to study the consequences of changes in incentive structures on social norms in field settings, for instance, by exploiting information from employee surveys or using lab experiments in firms to elicit social norms before and after a change (see, e.g., Burks and Krupka 2012 for an approach to elicit norms in firms). Moreover, in smaller firms or communities, people may have rather precise information about norms of behavior in their direct environment of colleagues or neighbors but not on broader groups of all employees in a large firm or most members of a society. It seems important to study the extent to which contract choices can affect norms of behavior in subgroups that can mutually observe each other.

The interplay between contracts and social norms in organizations is an important field for further research. Whereas it is often easy to change formal rules in organizations, changing the complex system of informal rules is typically a much more demanding endeavor. But as we have pointed out in this study, changes in formal rules affect perceptions about informal rules of behavior and thus shape these social norms. If our aim is to give better advice to practitioners on how to optimally design incentives, these indirect effects should be taken into account because they have significant potential to alter the way in which changes in the formal rules affect behavior and, in turn, the overall performance of organizations.

Appendix A. Additional Tables and Figures

Table A.1. Treatment Overview and Sample Size

	Baseline (Agents *do not know* that the principals have information concerning other agents)	Norms (Agents *know* that the principals have information concerning other agents)	Total
Costless			
Principals bear no costs for the contingent pay	Agents: 92 (62) Principals: 92	Agents: 93 (63) Principals: 93	370
Costly			
Principals pay €2 for the contingent pay	Agents: 81 (53) Principals: 81 [46]	Agents: 80 (52) Principals: 79[a] [46]	321
			691

Notes. The numbers in parentheses indicate the size of the subsample where we collected agents' beliefs. The numbers in brackets indicate the size of the subsample where principals' contract choice via the strategy method (see Section 5.2 for more details). The respective agent was matched with one randomly chosen principal at the end of the experiment. As the experiment was a one-short decision, we kept the data of the agent.

[a]In one session, one principal was missing because one computer terminal was left empty by mistake.

Table A.2. Mean Efforts in Costless and Costly Treatments

	Treatments	No. of independent observations	Fixed wage	Contingent pay	Δ Effort = Effort fixed wage − Effort contingent pay	*p*-values Fixed wage vs. contingent pay WSR
Costless	Baseline	92	34.66 (31.32)	44.49 (21.87)	−9.83 (28.46)	0.0048
	Norms	93	43.31 (28.45)	43.91 (23.54)	−0.60 (29.78)	0.9586
	p-values Baseline vs. Norms	MWU Test of variances[a]	0.0575 0.0583	0.8087 0.6210	0.0552	
Costly	Baseline	81	30.73 (28.85)	45.43 (21.60)	−14.70 (28.06)	0.0000
	Norms	80	43.53 (25.95)	43.23 (21.60)	0.30 (25.12)	0.6045
	p-values Baseline vs. Norms	MWU Test of variances[a]	0.0034 0.0413	0.5031 0.9964	0.0004	

Notes. Standard deviations are reported in parentheses. All reported *p*-values are two-sided.

[a]Levene's robust tests for equality of variances. There is no significant difference in efforts between the Costless and Costly treatments (all *p*-values are above 0.50, as tested with the two-sided MWU test). There is no significant difference in the variance of efforts between Costless and Costly treatments ($p > 0.35$).

Table A.3. Descriptive Statistics of Principals' Payoffs and Total Welfare

	Treatments	Fixed wage	Contingent pay	Δ = Fixed wage – Contingent pay	*p*-values Fixed wage vs. contingent pay WSR
		A: Principals' payoffs			
Costless	Baseline	5.16	9.11	−3.95	0.0000
		(3.76)	(1.53)	(3.32)	
	Norms	6.20	9.07	−2.88	0.0000
		(3.41)	(1.65)	(3.24)	
p-values	MWU	0.0575	0.8087	0.0292	
Baseline vs. Norms	Test of variances[a]	0.0583	0.6210		
Costly	Baseline	4.69	7.18	−2.49	0.0000
		(3.46)	(1.53)	(3.15)	
	Norms	6.22	7.03	−0.80	0.0724
		(3.11)	(1.51)	(2.78)	
p-values	MWU	0.0034	0.5031	0.0008	
Baseline vs. Norms	Test of variances[a]	0.0413	0.9964		
		B: Total welfare			
Costless	Baseline	14.35	15.29	−0.95	0.0003
		(1.81)	(1.11)	(1.80)	
	Norms	14.97	15.21	−0.24	0.7672
		(1.62)	(1.20)	(1.93)	
p-values	MWU	0.0504	0.5178	0.0217	
Baseline vs. Norms	Test of variances[a]	0.0078	0.6313		
Costly	Baseline	14.22	13.33	0.88	0.0001
		(1.83)	(1.14)	(1.96)	
	Norms	15.09	13.25	1.844	0.0000
		(1.53)	(1.22)	(1.66)	
p-values	MWU	0.0028	0.6374	0.0014	
Baseline vs. Norms	Test of variances[a]	0.0000	0.8093		

Notes. Standard deviations are reported in parentheses. All reported *p*-values are two-sided.
[a]Levene's robust tests for equality of variances. There is no significant difference in fixed wage payoffs between the Costless and Costly treatments (all *p*-values are above 0.50, as tested with the two-sided MWU test). There is no significant difference in the variance of payoffs and welfare between Costless and Costly treatments ($p > 0.30$).

Table A.4. Mean Beliefs in Costless and Costly Treatments

	Treatments	No. of independent observations	Fixed wage	Contingent pay	Δ Belief = Belief fixed wage – Belief contingent pay	*p*-values Fixed wage vs. contingent pay WSR
Costless	Baseline	62	29.5	42.31	−12.81	0.0093
			(26.97)	(20.95)	(31.90)	
	Norms	63	38.24	42.63	−4.40	0.2961
			(24.25)	(21.47)	(31.50)	
	p-values	MWU	0.0424	0.8723	0.1562	
	Baseline vs. Norms	Test of variances[a]	0.2719	0.8132		
Costly	Baseline	53	27.81	44.26	−16.45	0.0005
			(25.50)	(21.94)	(30.39)	
	Norms	52	32.40	45.71	−10.31	0.0394
			(25.70)	(18.39)	(30.68)	
	p-values	MWU	0.0870	0.7115	0.3523	
	Baseline vs. Norms	Test of variances[a]	0.6733	0.0973		

Notes. Standard deviations are reported in parentheses. All reported *p*-values are two-sided.
[a]Levene's robust tests for equality of variances. There is no significant difference in beliefs between the Costless and Costly treatments (all *p*-values are above 0.36, as tested with the two-sided MWU test). There is no significant difference in the variance of beliefs between Costless and Costly treatments ($p > 0.125$).

Table A.5. Online Belief Elicitation Experiment

Treatments	No. of independent observations	Fixed wage	Contingent pay	Δ Belief = Belief fixed wage – Belief contingent pay	*p*-values Fixed wage vs. contingent pay WSR
Contingent Contract Beliefs	57	49.12 (28.24)	59.28 (18.04)	−10.16 (30.05)	0.0178
Trust Contract Beliefs	61	62.36 (20.29)	49.13 (21.30)	13.23 (30.92)	0.0018
p-values	MWU	0.0006	0.0088	0.0002	
Contingent Contract Beliefs vs. Trust Contract Beliefs	Test of variances[a]	0.2534	0.4161		

Notes. Standard deviations are reported in parentheses. All reported *p*-values are two-sided.
[a]Levene's robust tests for equality of variances.

Table A.6. Explicit Norms Experiment

Treatments	No. of independent observations	Fixed wage	Contingent pay	Δ Effort = Effort fixed wage – Effort contingent pay	*p*-values Fixed wage vs. contingent pay WSR
Selfish Norm	30	27.6 (24.59)	46.43 (21.83)	−18.83 (32.18)	0.0044
Prosocial Norm	30	37.43 (28.45)	38.8 (23.71)	−1.367 (24.32)	0.8933
p-values	MWU	0.1092	0.3396	0.0129	
Contingent wage vs. fixed wage	Test of variances[a]	0.1089	0.1836		

Notes. Standard deviations are reported in parentheses. All reported *p*-values are two-sided.
[a]Levene's robust tests for equality of variances. As for the comparison between the Explicit Norms treatment and the Costless treatment in the baseline experiment, there is no significant difference in efforts except that fixed wage effort under the Selfish Norm is significantly lower than the fixed wage effort in the Costless Norms treatment ($p = 0.0073$, two-sided MWU test).

Figure A.1. Information for Principals in the Norms Treatments (Example from Costless Treatment, Translated from German)

Here are decisions about effort from 10 participants of the last session of this experiment who were in the role of employees:

	Employee 1	Employee 2	Employee 3	Employee 4	Employee 5	Employee 6	Employee 7	Employee 8	Employee 9	Employee 10
Effort under trust compensation	60	34	0	20	69	60	0	18	25	0
Effort under contingent compensation	25	60	31	25	60	70	45	38	10	65

Please note that your assigned employee has never participated in this experiment before. Additionally, he is not informed about the levels of effort of the employees in the previous experiment. He knows, however, that you are informed about these.

Notes. In the Norms treatment principals observed decisions of 10 agents from the preceding session of the same experiment. This table is a screenshot with such information used in one Norms session of our main experiment. The displayed values refer to the actual decisions of (all) 10 agents from an earlier Baseline session. In the second session of the Norms treatment we used data from the second session of the Baseline treatment ("Effort under trust compensation: 50, 51, 25, 35, 0, 80, 24, 35, 70, 0"; "Effort under contingent compensation: 25, 48, 38, 70, 32, 50, 20, 58, 60, 28").

Figure A.2. Information for Agents in the Norms Treatments (Translated from German)

	Employee 1	Employee 2	Employee 3	Employee 4	Employee 5	Employee 6	Employee 7	Employee 8	Employee 9	Employee 10
Effort under trust compensation	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx
Effort under contingent compensation	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx

The employer has information about the efforts of 10 other employees from the previous experiment. He is participating for the first time.
The employer sees the following table where instead of xx, the decisions of the employees from the past experiment are displayed.

Notes. This table was presented to all agents in the Norms treatment who did not have any information on the actual effort choices by other agents. However, they knew that their principal was informed about decisions of 10 participants from the previous experimental session and would observe an identical table but with numbers instead of "xx."

Appendix B. A Formal Model

We consider a generalization of Sliwka (2007) to the case of continuous types, which we apply to the framework studied in the experiment. An agent A works for a principal P who only cares about her profits. Agents have a utility function

$$u_i(\pi_A, \pi_P) = \pi_A + \theta_i \cdot \pi_P$$

such that θ_i measures the degree of prosociality of an agent. Agents are either "steadfast" or "conformists." For steadfast agents, the θ_i is exogenously given and follows a density function $f_N(\theta)$ on $[0,1]$, which has the property that $N = E[\theta \mid N]$ such that N is the *average prosociality in the population*. Both principal and agents have a common prior belief on the distribution of N with mean N_0 and support $[0,1]$ with a continuous density $g(N)$. Principals observe the population mean N prior to their contract choice. A conformist's degree of prosociality θ_C is equal to his conditional expectation on θ given the prior expectation and the principal's action. Hence, conformists try to be as prosocial as the population average N. Let η be the fraction of steadfast agents in the population, which is assumed to be common knowledge.[30] After having observed s_N, the principal chooses between two exogenously given contracts $C \in \{t, c\}$, a "trust contract" or a "contingent contract." After the contract choice, the agents exert effort e_i, which affects the principal's earnings. With probability e, the principal receives a payoff of B (and 0 otherwise). When contract $C = t$ is chosen, the agent receives a fixed wage w irrespective of the performance outcome; when choosing $C = c$, the agent works under a performance-contingent contract, which pays out a bonus b only if the principal earns B. We assume that the use of the contingent contract may impose an additional cost $k < w$, which is borne by the principal.

Equilibrium Analysis

As the principal has private information on the population norm N, which is relevant for the decision of a conformist agent, the game is a signaling game. We now characterize perfect Bayesian equilibria of this game. First we determine the optimal effort choice of an agent under each contract type for given beliefs about the norm. Under the trust contract, an agent of type θ (where $\theta = E_A[N \mid C = t]$ if the agent is a conformist) maximizes

$$\max_{\theta} \{w + \theta(Be - w) - (c/2)e^2\}$$

and thus chooses an effort level of

$$e_t(\theta) = \frac{\theta B}{c},$$

which is strictly increasing in the "degree of prosociality." Under the contingent contract, the agent maximizes

$$\max_{e} \{eb + \theta((B-b)e) - (c/2)e^2\},$$

choosing an effort level

$$e_c(\theta) = \frac{b + \theta(B-b)}{c}.$$

Note that efforts are increasing in θ under both contracts, but the effect of θ on effort is larger under the trust contracts as

$$\frac{\partial e_t}{\partial \theta} = \frac{B}{c} > \frac{\partial e_c}{\partial \theta} = \frac{B-b}{c}.$$

Hence, the conformists' behavior is more sensitive to their beliefs about the norm under a trust than a contingent contract. Furthermore, $(b + \theta(B-b))/c > \theta B/c$ such that effort is always larger under the contingent contract when there is full information.

Now we can consider the principal's optimal choice given the agent's reaction. For ease of notation, let $N_C = E[N \mid C]$ be the agent's rational expectation about the norm after a contract choice $C \in \{t, c\}$ on the equilibrium path—which, in turn, determines the conformists efforts. When choosing a trust contract, the principal's expected profits are

$$\begin{aligned} E[eB - w \mid C = t] &= E[e \mid C = t]B - w \\ &= (\eta N + (1-\eta)N_t)\frac{B^2}{c} - w, \end{aligned} \tag{B1}$$

and under the contingent contract

$$\begin{aligned} &E[e(B-b) - k \mid C = c] \\ &\quad = E[e \mid C = c](B-b) - k \\ &\quad = \frac{b + (\eta N + (1-\eta)N_c)(B-b)}{c}(B-b) - k. \end{aligned} \tag{B2}$$

Note that the principal's profits are always increasing in N_t and N_c; i.e., the principal benefits when the agent believes that the norm N is high. Given the agent's beliefs and optimal response, the principal prefers to choose a trust contract iff (B1) is larger than (B2), which after some rearrangement is equivalent to

$$N > \frac{b(B-b) + (1-\eta)(N_c(B-b)^2 - N_t B^2) + (w-k)c}{\eta b(2B-b)}. \tag{B3}$$

Using this, we can show the following result.

Proposition 1. (i) *If $B(b/c) > w - k$ and there are sufficiently many steadfast agents (i.e., η is sufficiently large), there is a unique separating equilibrium characterized by a cutoff value $\bar{N}$ such that the principal proposes the trust contract if and only if $N > \bar{N}$. The agents' expectation about the norm and thus the conformists' degree of prosociality is larger after the choice of a trust contract*:

$$N_t = E[N \mid N \geq \bar{N}] > N_0 > N_c = E[N \mid N < \bar{N}].$$

(ii) *The cutoff value $\bar{N}$ is strictly decreasing in k.*

Proof. To establish claim (i), we have to show that the fixed point equation

$$\begin{aligned} F(\bar{N}, k) = \bar{N} - \big(&b(B-b) + (1-\eta)\big(E[N \mid N \geq \bar{N}](\bar{N})(B-b)^2 \\ &- E[N \mid N \geq \bar{N}](\bar{N})B^2\big) + (w-k)c\big) \\ &\cdot (\eta b(2B-b))^{-1} = 0 \end{aligned}$$

has a unique solution. We will show that under the stated conditions, $F(1,k) > 0$, $F(0,k) < 0$, and $\partial F(\bar{N}, k)/\partial \bar{N} > 0$. By continuity of $F(\bar{N}, k)$, the result then follows.

First note that $F(1,k) > 0$ iff

$$1 - \frac{b(B-b) + (1-\eta)(N_0(B-b)^2 - B^2) + (w-k)c}{\eta b(2B-b)} > 0,$$

which is equivalent to

$$\frac{1}{1-N_0}\left(1 + \frac{Bb - (w-k)c}{(B-b)^2} - N_0\right) > \eta. \tag{B4}$$

As $(Bb-(w-k)c)/(B-b)^2>0$ iff $B(b/c)>w-k$, condition (B4) always holds (even for $\eta\to 1$) if $B(b/c)>w-k$.

Note that $F(0,k)<0$ is equivalent to

$$b(B-b)+(w-k)c>(1-\eta)N_0B^2,$$

which holds if η is sufficiently large. To show that $\partial F(\bar N,k)/\partial\bar N>0$ if η is large enough, we use the fact that

$$\frac{\partial E[N\mid N\ge\bar N]}{\partial\bar N}=\frac{f(\bar N)}{1-F(\bar N)}(E[N|N\ge\bar N]-\bar N)\quad\text{and}$$
$$\frac{\partial E[N\mid N<\bar N]}{\partial\bar N}=\frac{f[\bar N]}{F[\bar N]}(\bar N-E[N\mid N<\bar N])$$

to obtain

$$\frac{\partial F(\bar N,k)}{\partial\bar N}$$
$$=1-\left((1-\eta)\left(\frac{f[\bar N]}{F[\bar N]}(\bar N-E[N\mid N<\bar N])(B-b)^2\right.\right.$$
$$\left.\left.-\frac{f(\bar N)}{1-F(\bar N)}(E[N\mid N\ge\bar N]-\bar N)B^2\right)\right)\cdot(\eta b(2B-b))^{-1},$$

which will be strictly positive whenever

$$\frac{\eta b(2B-b)}{(1-\eta)B^2}>\frac{f[\bar N]}{F[\bar N]}(\bar N-E[N\mid N<\bar N])\frac{(B-b)^2}{B^2}$$
$$-\frac{f(\bar N)}{1-F(\bar N)}(E[N\mid N\ge\bar N]-\bar N),$$

which will always hold when

$$\frac{\eta b(2B-b)}{(1-\eta)B^2}$$
$$>-\left[\frac{f(\bar N)}{1-F(\bar N)}(E[N|N\ge\bar N]-\bar N)-\frac{f[\bar N]}{F[\bar N]}(\bar N-E[N\mid N<\bar N])\right].$$

The right-hand side is bounded on $[0,1]$ (see the "Properties of the Δ function" section in the appendix of Bénabou and Tirole 2012). As $\lim_{\eta\to1}(\eta b(2B-b)/((1-\eta)B^2))=\infty$, there is a unique separating equilibrium when η is sufficiently large.

We establish claim (ii) by implicit differentiation:

$$\frac{\partial\bar N}{\partial k}=-\frac{\partial F(\bar N,k)/\partial k}{\partial F(\bar N,k)/\partial\bar N}.$$

We know already that $\partial F(\bar N,k)/\partial\bar N>0$ (if η is large enough) and

$$\frac{\partial F(\bar N,k)}{\partial k}=\frac{c}{\eta b(2B-b)}>0,$$

which completes the proof. □

We can also compare this to a situation without norms signaling (i.e., when the principal is uninformed), where the conformists stick to their prior beliefs about the norm $N_0=E[N]$, Proposition 1 then directly implies the following.

Corollary 1. *In any separating equilibrium in a game where norms signaling is feasible, efforts are higher under a trust contract and lower under a contingent contract compared with a setting where the contract choice reveals no information about the norm.*

Recall that the principal is always better off with a contingent contract when all agents are selfish. In a next step, we explore under which conditions pooling equilibria exist, in which the principal will always choose the contingent contract even when conformity matters and agents are uncertain about the norm. Of course, the existence of pooling equilibria hinges on assumptions about out-of-equilibrium beliefs. We take a conservative approach and impose no restriction on these beliefs and characterize the largest set of potential pooling equilibria.

Proposition 2. *A pooling equilibrium can exist in which the principal always chooses the contingent contract if*

$$\eta\le\frac{b(B-b)+c(w-k)+N_0(B-b)^2}{b(2B-b)+N_0(B-b)^2}.\tag{B5}$$

If $w-k<B(b/c)$, no pooling equilibrium exists if η is sufficiently large.

Proof. When the principal chooses the trust contract, her profits are

$$\eta N\frac{B^2}{c}-w.$$

A necessary condition for the existence of this equilibrium is that it must be beneficial to choose the contingent contract, even if agents believe that $N=0$ after a deviation to the trust contract. In that case, profits under the contingent contract are

$$\frac{b+(\eta N+(1-\eta)N_0)(B-b)}{c}(B-b)-k.$$

The principal thus prefers the contingent contract iff

$$\frac{b+(\eta N+(1-\eta)N_0)(B-b)}{c}(B-b)-k\ge\eta N\frac{B^2}{c}-w\iff$$
$$b(B-b)+c(w-k)+(1-\eta)N_0(B-b)^2\ge\eta N(B^2-(B-b)^2).$$

A pooling equilibrium exists if this is the case even for $N=1$ or

$$b(B-b)+c(w-k)+(1-\eta)N_0(B-b)^2\ge\eta(B^2-(B-b)^2)$$
$$\iff\frac{b(B-b)+c(w-k)+N_0(B-b)^2}{b(2B-b)+N_0(B-b)^2}>\eta.$$

Note that the cutoff is strictly positive. It is straightforward to check that it is strictly smaller than 1 iff

$$w-k<B\frac{b}{c}.\quad\square$$

Endnotes

[1] See, for instance, Akerlof (1980), Elster (1989), Bernheim (1994), Lindbeck et al. (1999), Kübler (2001), Fehr et al. (2002), Fehr and Fischbacher (2004a, b), Fischer and Huddart (2008), Bicchieri (2006), Krupka and Weber (2009), Huck et al. (2012), and Krupka et al. (2016). See also Young (2008) for an overview.

[2] Examples include Ichino and Maggi (2000), Clark (2003), Stutzer and Lalive (2004), and Bradler et al. (2016).

[3] For further experimental evidence on related crowding-out effects, see Gneezy and Rustichini (2000b), Fehr and Falk (2002), Fehr and Rockenbach (2003), Fehr and List (2004), Falk and Kosfeld (2006), and Mellström and Johannesson (2008). For a broader overview on the issue, see Bowles (2008).

[4] Several other theoretical models explore the detrimental effects of sanctions or performance-contingent pay schemes and give potential (behavioral) economic explanations. See, for instance, Bénabou and Tirole (2003, 2006) and Ellingsen and Johannesson (2008).

[5] See Selten (1967).

[6] The latter is well in line with previous studies on "social history effects." For instance, Berg et al. (1995) show that information about the behavior of others has a positive effect on reciprocity in an investment game. Bardsley and Sausgruber (2005), Bicchieri and Xiao (2009), Servátka (2009), Gächter et al. (2012), and Gürerk (2013) find a positive correlation between contributions in public-good, dictator, and gift-exchange games and information on decisions of unrelated individuals in the same situation.

[7] See also Hart and Moore (2008, Section V), Kessler and Leider (2012), and Bartling and Schmidt (2015) for more recent studies on the interaction of social norms and contracts that do not focus on the signaling effect of contracts.

[8] The second-best effort (maximizing the individual payoff) is 0 for the trust and 30 for the performance-contingent pay, and the first-best effort (maximizing the joint payoff) is 72 in both cases. Under the trust compensation, principals and agents earn equal (expected) payoffs at an effort of 59, and the principals (on average) do not make any loss starting from effort level 42. Agents had access to an on-screen computation tool, where they could insert effort values for a particular incentive scheme and learn the costs of effort and (expected) payoffs for both parties. They could use this tool for as many trials as they wanted before determining their final decision.

[9] In particular, principals see a table with 10 columns and 2 rows similar to Figure A.1 in Appendix A. To ascertain that agents understand the principal's information structure, agents see the same table but with "xx" instead of the actual efforts (see Figure A.2 in Appendix A). See Online Appendix D for the instructions.

[10] Our initial setup encompassed nine sessions. At the request of the referees, we collected more observations and ran another 16 sessions. In each session two treatments were conducted at the same time with subjects being randomly assigned to one of two treatments. In these sessions we also elicited agents' beliefs about the behavior of others and (in the costly contract choice treatments) principals' choices by the strategy method.

[11] It is important to note that an increase in average efforts is consistent with the theory. The model predicts that efforts increase under the fixed wage and decrease under the contingent wage. However, efforts under the contingent wage are less sensitive to information about the norm as, here, agents have a selfish incentive to work in both settings (in the model, the first derivative of conformists' effort choices with respect to their beliefs about the norm is always smaller under the contingent contact than under the fixed wage). Hence, overall efforts may well increase in equilibrium with norms signaling.

[12] To be precise, in the model, the condition $B(b/c) > w - k$, which is satisfied in the Costly contract choice setting but not in the Costless setting, guarantees the existence of a separating equilibrium if only the fraction of conformists is not too large; the same condition then also rules out the existence of a pooling equilibrium where only the contingent wage is offered. If this condition is not met, a separating equilibrium exists only under stronger restrictions on prior beliefs. See Appendix B for details.

[13] To see that, note the following: if there is a separating equilibrium in the model, informed principals choose the trust contract if and only if there is sufficiently strong average prosociality in the population. The higher the costs of the performance-contingent contract, the lower is the required level of prosociality that makes offering a trust contract more attractive. In the model, a costly contingent compensation option therefore leads to a weaker increase in efforts compared with the setting where the contingent contract can be chosen at no costs (provided that a separating equilibrium is played in both of them).

[14] In the Costly contract choice treatment, the effort effect of the Norms intervention the *proportion* of agents who exert strictly more effort under the trust than the contingent contract from 27% to 46% ($p = 0.0120$, two-sided proportion test). But the Norms intervention leaves the *spread in efforts* for these agents virtually unchanged (those who exert more effort under the trust contract exert 20.2 more effort units in the Baseline and 20.6 in Norms). In the Costless contract choice, the opposite pattern occurs, as the proportion of agents exerting more effort under the trust than the contingent contract increases to a weaker extent from 38% to 44% ($p = 0.4036$, two-sided proportion test), but the spread in efforts increases in these cases from 18.8 to 25.9 effort units (i.e., by about 38%, $p = 0.0726$, two-sided MWU test). Hence, more agents positively react to the choice of a trust contract in the Costly setting (in line with the idea that the likelihood that the signal is seen as informative is larger), but the size of the positive reaction is stronger in the Costless setting.

[15] We thank an anonymous referee for this suggestion.

[16] When the deviation between the guess and the actual mean was above 14, agents received no payment for these questions (i.e., no fines were imposed for wrong guesses).

[17] See Zizzo (2010) for discussion about demand effects in belief elicitation and Costa-Gomes et al. (2014) for a discussion on the endogeneity of elicited beliefs.

[18] See Table A.4 in Appendix A for more details.

[19] The subjects received nothing for answers with a deviation of more than 17 points from the true value.

[20] After showing the instructions of the Norms treatment and asking test questions to ensure that agents understood the setting, the subjects received the following statement: "We have randomly drawn one of the prior participants in the role of an employer. This employer has observed contributions of 10 employees from the prior experiment for both the trust and the contingent contract. This employer has chosen the [Treatment Contingent Pay Beliefs] contingent contract/[Treatment Fixed Wage Beliefs] trust wage after having seen the table of the form shown in the above. What is your best estimate about the average effort under the trust contract in the table above? What is your best estimate about the average effort under the contingent contract in the table above?"

[21] See also Table A.5 in Appendix A for descriptive statistics and p-values.

[22] We also conducted a "Baseline Beliefs experiment" in which instructions from the Baseline treatment of the Costless Contract Choice experiment were presented to a new sample of 60 subjects in the role of outside observers. Subjects had to estimate the average efforts under the fixed wage and contingent compensation. We can use this experiment to compare first- and third-party beliefs. As Figure C1 in Online Appendix C shows, the observers' average estimate of efforts under the fixed wage is closer to the true effort than the first-party estimate, which was substantially smaller than true efforts (mean effort: 34.66, third-party belief: 36.40, $p = 0.4933$, first-party belief: 29.50, $p = 0.0864$, two-sided WSR test). This supports the conjecture that agents adapt first-party beliefs in a self-serving manner. But, by contrast, outside observers substantially overestimate effort under the contingent wage (mean effort: 44.49, third-party belief: 53.22, $p = 0.0001$, first-party belief: 42.31, $p = 0.8252$, two-sided WSR test). This may be because self-serving beliefs play a weaker role here, but on the other hand, it may be easier to predict effort choices for agents who had actually made that same effort choice decision before.

[23] In the first sessions, we did not use the strategy method for the principal's choices, and hence, there was hardly any variation in observed behavior, which makes it impossible to evaluate causal

effects of information about the norm on contract choices. In these sessions 15.22% of principals choose the fixed wage contract in the Costless Baseline and 46.43% in the Costly Baseline treatments. Because of the low prosociality shown in the Costless Norms treatment, principals offered the fixed wage only in 4.3% of the cases in this treatment and in 25% in the Costly Norms treatment.

[24] We implemented this design element in the Costly contract choice setting since this setting allowed us to construct realistic tables close to real decision behavior, in which either the contingent contract or the trust contract is more profitable.

[25] To check the credibility of the different tables, principals were asked after their choices to state for each of the tables on a five-point Likert scale whether they thought it was likely that the table "originated from a previous experiment" (from 1 = "very unlikely to originate from a previous experiment" to 5 = "very likely to originate from a previous experiment"). The ratings varied only between 2.7 and 3.3, such that none of the tables was considered to be unlikely.

[26] We intentionally spoke of "selected agents" to avoid deception but made no information available on the specific selection procedure.

[27] The exact individual values are reported in Tables D8 and D9 in Online Appendix D.

[28] See Table A.6 in Appendix A for details.

[29] See Bénabou and Tirole (2012, Section 4) for a related discussion on "expressive law," i.e., the role of law in conveying a society's norms of behavior, which may lead to the choice of "softer" laws in order to signal that, for instance, only very disreputable people do not follow the norm, and hence, the need to induce tough sanctions is low. See also the discussion in Bowles (2008).

[30] It is, of course, also conceivable that knowledge about the proportion of steadfast types is not common knowledge. In particular, when the principal does not observe the prior norm N perfectly but observes behavior in a sample of agents, this leads to an additional inference problem, as behavior in the sample reflects not only the behavior of steadfast agents but also the beliefs of conformist types.

References

Akerlof GA (1980) A theory of social custom, of which unemployment may be one consequence. *Quart. J. Econom.* 94:749–775.

Bardsley N, Sausgruber R (2005) Conformity and reciprocity in public good provision. *J. Econom. Psych.* 26:664–681.

Bartling B, Schmidt KM (2015) Reference points, social norms, and fairness in contract renegotiations. *J. Eur. Econom. Assoc.* 13: 98–129.

Battigalli P, Dufwenberg M (2007) Guilt in games. *Amer. Econom. Rev.* 97:170–176.

Bénabou R, Tirole J (2003) Intrinsic and extrinsic motivation. *Rev. Econom. Stud.* 70:489–520.

Bénabou R, Tirole J (2006) Incentives and prosocial behavior. *Amer. Econom. Rev.* 96:1652–1678.

Bénabou R, Tirole J (2012) Laws and norms. IZA Discussion Paper 6290, IZA, Bonn, Germany.

Berg J, Dickhaut J, McCabe K (1995) Trust, reciprocity, and social history. *Games Econom. Behav.* 10:122–142.

Bernheim BD (1994) A theory of conformity. *J. Political Econom.* 102:841–877.

Bicchieri C (2006) *The Grammar of Society: The Nature and Dynamics of Social Norms* (Cambridge University Press, New York).

Bicchieri C, Xiao E (2009) Do the right thing: But only if others do so. *J. Behavioral Decision Making* 22:191–208.

Bowles S (2008) Policies designed for self-interested citizens may undermine "the moral sentiments": Evidence from economic experiments. *Science* 320:1605–1609.

Bradler C, Dur R, Neckermann S, Non A (2016) Employee recognition and performance: A field experiment. *Management Sci.* 62: 3085–3099.

Burks SV, Krupka EL (2012) A multimethod approach to identifying norms and normative expectations within a corporate hierarchy: Evidence from the financial services industry. *Management Sci.* 58:203–217.

Clark AE (2003) Unemployment as a social norm: Psychological evidence from panel data. *J. Labor Econom.* 21:323–351.

Costa-Gomes MA, Huck S, Weizsäcker G (2014) Beliefs and actions in the trust game: Creating instrumental variables to estimate the causal effect. *Games Econom. Behav.* 88:298–309.

Ellingsen T, Johannesson M (2008) Pride and prejudice: The human side of incentive theory. *Amer. Econom. Rev.* 98:990–1008.

Ellingsen T, Johannesson M, Tjøtta S, Torsvik G (2010) Testing guilt aversion. *Games Econom. Behav.* 68:95–107.

Elster J (1989) Social norms and economic theory. *J. Econom. Perspect.* 3:99–117.

Falk A, Kosfeld M (2006) The hidden costs of control. *Amer. Econom. Rev.* 96:1611–1630.

Fehr E, Falk A (2002) Psychological foundations of incentives. *Eur. Econom. Rev.* 46:687–724.

Fehr E, Fischbacher U (2004a) Social norms and human cooperation. *Trends Cognitive Sci.* 8:185–190.

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

Fehr E, List JA (2004) The hidden costs and returns of incentives—Trust and trustworthiness among CEOs. *J. Eur. Econom. Assoc.* 2:743–771.

Fehr E, Rockenbach B (2003) Detrimental effects of sanctions on human altruism. *Nature* 422:137–140.

Fehr E, Fischbacher U, Gächter S (2002) Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature* 13:1–25.

Fischbacher U (2007) z-Tree: Zurich toolbox for ready-made economic experiments. *Experiment. Econom.* 10:171–178.

Fischer P, Huddart S (2008) Optimal contracting with endogenous social norms. *Amer. Econom. Rev.* 98:1459–1475.

Friebel G, Schnedler W (2011) Team governance: Empowerment or hierarchical control. *J. Econom. Behav. Organ.* 78:1–13.

Gächter S, Nosenzo D, Sefton M (2012) The impact of social comparisons on reciprocity. *Scand. J. Econom.* 114:1346–1367.

Galbiati R, Schlag KH, van der Weele JJ (2013) Sanctions that signal: An experiment. *J. Econom. Behav. Organ.* 94:34–51.

Gneezy U, Rustichini A (2000a) A fine is a price. *J. Legal Stud.* 29:1–18.

Gneezy U, Rustichini A (2000b) Pay enough or don't pay at all. *Quart. J. Econom.* 115:791–810.

Greiner B (2004) An online recruitment system for economic experiments. Kremer K, Macho V, eds. *Forschung und wissenschaftliches Rechnen* (Gesellschaft für wissenschaftliche Datenverarbeitung, Göttingen, Germany), 79–93.

Gürerk Ö (2013) Social learning increases the acceptance and the efficiency of punishment institutions in social dilemmas. *J. Econom. Psych.* 34:229–239.

Hart O, Moore J (2008) Contracts as reference points. *Quart. J. Econom.* 123:1–48.

Huck S, Kübler D, Weibull J (2012) Social norms and economic incentives in firms. *J. Econom. Behav. Organ.* 83:173–185.

Ichino A, Maggi G (2000) Work environment and individual background: Explaining regional shirking differentials in a large Italian firm. *Quart. J. Econom.* 115:1057–1090.

Kessler JB, Leider S (2012) Norms and contracting. *Management Sci.* 58:62–77.

Krupka E, Weber RA (2009) The focusing and informational effects of norms on pro-social behavior. *J. Econom. Psych.* 30:307–320.

Krupka EL, Leider S, Jiang M (2016) A meeting of the minds: Informal agreements and social norms. *Management Sci.*, ePub ahead of print May 31, http://dx.doi.org/10.1287/mnsc.2016.2429.

Kübler D (2001) On the regulation of social norms. *J. Law, Econom., Organ.* 17:449–476.

Lindbeck A, Nyberg S, Weibull JW (1999) Social norms and economic incentives in the welfare state. *Quart. J. Econom.* 114:1–35.

Mellström C, Johannesson M (2008) Crowding out in blood donation: Was Titmuss right? *J. Eur. Econom. Assoc.* 6:845–863.

Selten R (1967) Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes. Sauermann H, ed. *Beiträge zur experimentellen Wirtschaftsforschung* (J.C.B. Mohr (Paul Siebeck), Tübingen, Germany), 136–168.

Servátka M (2009) Separating reputation, social influence, and identification effects in a dictator game. *Eur. Econom. Rev.* 53:197–209.

Sliwka D (2007) Trust as a signal of a social norm and the hidden costs of incentive schemes. *Amer. Econom. Rev.* 97:999–1012.

Stutzer A, Lalive R (2004) The role of social work norms in job searching and subjective well-being. *J. Eur. Econom. Assoc.* 2:696–719.

van der Weele J (2012) The signaling power of sanctions in social dilemmas. *J. Law, Econom., Organ.* 28:103–126.

Young HP (2008) Social norms. Durlauf SN, Blume LE, eds. *The New Palgrave Dictionary of Economics* (Nature Publishing Group, Basingstoke, UK), 647–651.

Zizzo DJ (2010) Experimenter demand effects in economic experiments. *Experiment. Econom.* 13:75–98.