Organizational Culture as Equilibrium?

Rules versus Principles in Building Relational Contracts

Robert S. Gibbons Manuel Grieder Holger Herz Christian Zehnder

January 15, 2020

Abstract

Effective organizations are able not only to coordinate their members on efficient strategies but also to adapt members' strategies to unforeseen change in an efficient manner. We explore whether part of organizational culture—namely, relational contracts that facilitate both coordination and adaptation—enable organizations to achieve these ends. In a novel experiment, we explore how parties establish such relational contracts, whether they achieve efficient cooperation, and how they adapt to exogenous shocks. Specifically, we test the hypothesis that basing a relational contract on general principles rather than specific rules is more successful in achieving efficient adaptation. In our Baseline condition, we observe that pairs who articulate general principles achieve significantly higher performance than those who rely on specific rules. The mechanism underlying this correlation is that pairs with principle-based agreements are more likely to expect their pair to take actions that are consistent with what their relational contract prescribes. To investigate whether there is a causal link between principle-based agreements and performance, we implement a "Nudge" intervention to foster principle-based relational contracts. The Nudge succeeds in motivating more pairs to formulate principles and in making pairs significantly more likely to select efficient initial choices. However, the intervention fails to increase performance in the long run. Our results suggest that principle-based relational contracts may improve organizational performance, but our results also illustrate the difficulty of building such an organizational culture, which is consistent with the idea that high-performing relational contracts constitute a competitive advantage only if they are difficult to imitate.

1 Introduction

In an early and influential statement, the social psychologist Edgar Schein defined organizational culture as "a pattern of basic assumptions—invented, discovered or developed by a given group as it learns to cope with its problems of external adaptation and internal integration ..." (1985: 9). At about the same time, but from a very different discipline, the economist David Kreps suggested that an aspect of organizational culture—the shared understanding of "how we do things around here"—might be modeled as an equilibrium of a repeated game (1990; conference presentation in 1984).

Two terms used by Kreps and Schein are "equilibrium" and "adaptation," respectively. In game theory, equilibrium means agreement on self-enforcing strategies: in this sense an organizational culture exists only if people agree on how things should be done and are willing to do things that way. Adaptation, on the other hand, suggests that organizational culture may help people adjust their strategies when their environment changes in unexpected ways.

Inspired by Kreps and Schein, in this paper we study "relational contracts"—shared understandings of the parties' roles in and rewards from collaborating together—as *part* of organizational culture. Although our approach is rooted in economics, we see ourselves as drawing importantly from Schein, and we hope that our results will prompt new theorizing about issues that were elided in Kreps' seminal statement.

We use experimental methods to explore whether parties can build relational contracts that not only provide guidance on how to behave in a given environment (equilibrium) but also facilitate coordinated changes in behavior after an unforeseen shock (adaptation). We thus hope to contribute to understandings of how the relational-contract part of organizational culture may be built initially and also updated when conditions change. Specifically, we explore the hypothesis that basing a relational contract on general principles rather than specific rules is more successful in ensuring efficient adaptation.

Although there is a large theoretical literature on relational contracting (see, e.g. Malcomson, 2013, for a survey), all the models we know assume that the parties are in equilibrium from the beginning of their relationship. Put differently, these models ignore the potential challenges that real parties face when creating a shared understanding of desired behaviors before their relationship

has begun. Furthermore, although some of the most interesting and important recent papers on relational contracts—such as Chassang (2010) or Andrews & Barron (2016) or Li et al. (2017)—analyze settings with learning or dynamics or adaptation, these models abstract from the fact that it is exactly in such non-stationary settings where real parties may have most difficulty creating shared understandings in the first place.

To see the potential challenge that building this kind of organizational culture may entail, it is important to note that there are very substantial information requirements underlying relational contracts. Gibbons & Henderson (2012) illustrate this "relational knowledge" using Lincoln Electric, Toyota, and Merck as examples. For instance, the parties need a shared understanding about which actions (under which circumstances) constitute "cooperation," which "defection," and which are available as "punishment" should the other party fail to cooperate. That is, in successful relational contracts the parties have solved not only the credibility problem of "Can I trust that you will do what we agreed to do?" but also the clarity problem of "Do we have a shared understanding of what we agreed to do?" The former is the central focus of the large theoretical literature mentioned above; the latter is missing from that literature (because the models assume that the parties start in equilibrium) and an important motivation for this paper.

These issues of building a shared understanding arise not only at the beginning of a relationship but also during successful and long-standing relational contracts when learning, dynamics, or adaptation create related challenges. As just one example, at Lincoln Electric the historical practice was that the total bonus pool was about half the firm's pre-tax, pre-bonus earnings, but then the firm invested heavily overseas, leading to a year when its domestic operation had record earnings but losses overseas were so large that the firm as a whole lost money, raising the question of whether the bonus pool for the domestic operation should be of record size (consistent with domestic earnings) or instead should be zero or even negative (consistent with the firm as a whole losing money). This example illustrates how a change in the environment can create uncertainty about what it means to comply with an existing relational contract.

Motivated by many cases like Lincoln Electric, we study the emergence and change of relational contracts in settings where there are shocks to the environment and the parties need to adapt (as opposed to stationary settings that may include routine fluctuations but where nothing fundamentally novel ever occurs). We conjecture that parties who develop agreements based on general

principles rather than on specific rules of behavior may perform better when their world changes. For example, the principle and the rule might give identical prescriptions for what should be done in a stationary setting familiar from shared experience, but the potential advantage of a principle is that it might give useful guidance even after the environment changes, whereas a rule might then prescribe completely inappropriate behaviors.

We hope to identify useful steps towards building a relational contract that facilitates both cooperation and adaptation. We are particularly interested in the impact of different forms of non-binding agreements (rules vs. principles) on building and updating an equilibrium. While there are fascinating ethnographies that can be interpreted as studying how new equilibria may emerge (or not) in organizations (see, e.g., Kellogg, 2009, 2011), it is difficult to use field data to isolate the determinants of successfully established relational contracts. Doing so requires that different approaches to building an equilibrium be observed and compared in detail within a given environment. We therefore see laboratory experiments as particularly well-suited to exploring organizational culture as equilibrium. First, a laboratory setting allows us to control previous history and the structure and dynamics of the environment. Second, because we not only observe participants' actions but also elicit their beliefs about others' likely actions, we can measure whether they have built an equilibrium. And finally, we can record their communications and relate what has been said and agreed upon to what kinds of equilibria result.

1.1 Overview of Experiment

In our experiment we implement a repeated buyer-seller game with a random stopping rule. Participants know the initial structure and parameters of the game, but they are also aware that the game will change in an unknown way after a few periods. Before interactions start, we give participants the possibility to communicate in order to find an agreement on how to play the entire game and to formulate a written statement capturing this agreement.

In addition to observed choices, we also measure the extent to which pairs reached a shared understanding about appropriate behavior in new environments. To collect these data, we interrupt interactions before the game changes. We confront participants with different scenarios of how the game might change for the coming periods and ask them to indicate for each scenario what choices their agreement prescribes and what choices they expect the pair to actually implement.

Subsequently, we implement one of the scenarios as the actual game that participants play for the remaining periods.

A first set of results explores the nature of endogenously emerging agreements in this Baseline setting and investigates how different types of agreements correlate with performance. We find that pairs whose non-binding ex-ante agreement articulates a broad principle tend to reach higher levels of performance when their world changes than do pairs whose ex-ante agreement articulates a narrow rule. A deeper analysis reveals that the mechanism underlying this correlation is that pairs with principle-based agreements are more likely to expect their pair to take the actions that are normatively prescribed by their agreement in different scenarios. In short, pairs with principle-based agreements are more likely to be in equilibrium.

As a next step we then investigate whether there is a causal link between principle-based agreements and performance. To this end, we implement a "Nudge Treatment" designed to exogenously foster the emergence of principle-based agreements. Our Nudge succeeds in motivating more pairs to formulate principles instead of rules in their agreements. In addition, the Nudge also makes pairs significantly more likely to select efficient initial choices in the new scenarios with which we confront them. However, the treatment fails to increase long-run performance in the actual game that participants play for the remaining periods. An important reason for the lack of an effect on overall performance is that, although the Nudge Treatment is effective in coordinating pairs on efficient initial quality, it does not lead to more consensus about the price choice. The fact that our treatment does not achieve the latter is in line with the view that building a relational contract is difficult in practice—both in the world and, apparently, in the lab.

1.2 Related Literature

Beyond organizational culture, our paper relates to several other prominent strands of organization theory. First, relational contracts are an important feature of transaction-cost economics (such as Williamson, 1979, 1991). Second, we see parallels with Cyert & March's (1963) quasi-resolution of conflict in the behavioral theory of the firm: relational contracts do not socialize organization members into having the same goals; instead, a relational contract may be one way to "deal successfully with the obvious potential for internal goal conflict inherent in a coalition of diverse individuals and groups" (1963: 31).

Turning to sociology, we see relational contracts as an aspect of informal organization. As Granovetter says "The distinction between the 'formal' and the 'informal' organization of the firm is one of the oldest in the literature" (1985: 502) and Blau & Scott say "It is impossible to understand the nature of a formal organization without investigating the [...] informal relations and unofficial norms" (1962: 6).

Finally, there are many new organizational forms that emphasize informal or relational aspects, including Foss' (2006) analysis of Oticon's "spaghetti organization" and Turco's (2016) "conversational firm." Both Foss and Turco explicitly reference the relational-contract model of empowerment by Baker et al. (1999), and Turco uses Gibbons & Henderson (2012) to discuss the difficulties in building such relational contracts.

While we see connections to these strands of organization theory, it may also be helpful to distinguish our view of organizational culture from the definitions of "culture" that appear in anthropology and elsewhere. As an example of the latter, Guiso et al. (2006: 23) define culture as "customary beliefs and values that ethnic, religious, and social groups transmit fairly unchanged from generation to generation." In the spirit of Kreps and Schein, we focus on an organization that has whatever members it has (who have whatever heritages they have), and we then ask what the organization might do to help its members to coordinate in the near term and to adapt in the long.

Methodologically, our experimental approach follows the traditions of experimental economics. More specifically, our study is related to laboratory experiments on repeated social dilemma games that study how cooperation evolves and what strategies emerge in such environments (see, e.g., Dal Bó, 2005; Dal Bó & Fréchette, 2011; Fudenberg et al., 2012 for infinitely repeated games, and, e.g., Andreoni & Miller, 1993; Brown et al., 2004; Fehr et al., 2009; Herz et al., 2019 for games with finite repetition). Our experiment advances this literature by studying how parties can prepare for unforeseen contingencies in an uncertain environment, a difficult and highly relevant challenge in many organizational settings.[1]

Our experiment also builds on previous experimental studies of how communication between partners facilitates cooperation (for example, see Sally, 1995; Ellingsen & Johannesson, 2004; Charness & Dufwenberg, 2006; Balliet, 2010). Our paper goes beyond the fact that communicating helps

[1]Fudenberg et al. (2012) explored some effects of uncertainty on cooperation by studying the case of a stable environment in which players' choices are implemented with noise.

(relative to not communicating) by studying what kind of non-binding agreements are useful to achieve robust cooperation in an unstable environment. Moreover, in our setting, different groups reach different agreements in advance and then have different interpretations of adaptation events in the future. This finding echoes experiments in common-interest settings (rather than divergent-interest games) where different groups develop different languages, with implications for how they adapt in the future (see, e.g., Weber & Camerer, 2003; Selten & Warglien, 2007).

In the next section we describe the experimental design, section 3 summarizes our hypotheses, and section 4 presents the results. In section 5 we discuss implications of our results and conclude.

2 Study Design

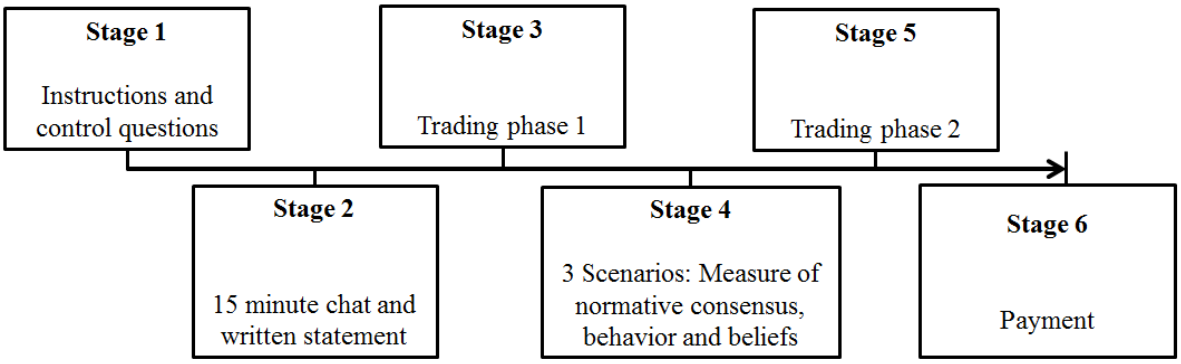
To study whether relational contracts can achieve both coordination in the short run and adaptation after shocks, we created a novel laboratory experiment in which a buyer (she) and a seller (he), engaged in a repeated trading game with uncertainty about the future. Participants knew that, after a fixed period of time, the parameters of the game would change in an unforeseeable way. To assess the role of communication for equilibrium building, before interactions started, we gave participants the possibility to communicate in order to find an agreement on how to play the game and to formulate a written statement capturing this agreement. After the first phase of interactions and before the game changed, we measured each group's shared understanding of how to behave in new environments. For this purpose we confronted participants with three possible scenarios of how the game might change. We asked them to indicate what they should do according to their initial agreement, and what they would actually do. Subsequently, one of these scenarios was implemented for the remaining interaction periods. This setup allowed us to observe what types of agreements endogenously emerged and how these agreements were associated with groups' shared understanding and performance.

To assess whether the nature of a group's agreement has a causal impact on their performance, we implemented an experimental treatment that fostered the emergence of principles in the agreement finding phase. In subsection 2.1, we describe the experimental set-up in our Baseline condition in more detail. In subsection 2.2, we explain the experimental treatment.

2.1 Baseline Experimental Set-up

Participants were randomly paired in groups of two that remained fixed for the entire experiment. In each group, participants were randomly assigned to the roles of either seller or buyer. Figure 1 represents the timeline of the experiment. Below, we discuss each stage in detail.

Figure 1: **Timeline and stages of the experiment**



Stage 1: Instructions and Control Questions. The instructions informed all participants of their role and of the game parameters in the first (five-period) trading phase (the full instructions are reproduced in Appendix A). Participants were made aware of the fact that the game parameters and structure would change after 5 periods, but no information was provided about the nature of the change. Participants also knew that they would continue playing with the same partner as in the first phase and that the second phase would be an infinitely repeated game with a stopping probability of 10% after each round. At the end of the instructions, participants had to correctly answer a set of control questions to enter the next stage.

Stage 2: Communication and Statement. Participants could exchange free-form text messages with their interaction partner via computer for 15 minutes. Chats were anonymous and participants were instructed not to reveal their identities. We told participants to use the chat function to find an agreement on how they intended to play the game and to write this down in a short joint statement. This statement was drafted by the seller in an entry window that was separate from the chat function. The seller could transmit the draft of the statement to the buyer, who could demand changes in the chat window. If the buyer agreed with the statement, she could confirm the statement with a click on a button. Once a statement was confirmed, the communication phase

ended. If no statement was confirmed within 15 minutes the groups continued without a statement.

Stage 3: First Trading Phase. All groups played five rounds of a simultaneous buyer-seller game in which the seller decided about the quality of a good to be delivered and the buyer determined the price.[2] The seller chose a quality $q \in [0, 10]$ at cost $c(q)$, where higher quality was associated with higher cost ($c'(q) > 0, c''(q) > 0$). Table 1 displays the cost function $c(q)$.[3] The value of the product to the buyer in the first trading phase was $v = 10q$ and the buyer picked a price $p \in [0, 100]$. In any period, the buyer's payoff was thus $\pi_B = 10q - p$ and the seller's payoff was $\pi_S = p - c(q)$. All parameters of the game were common knowledge.[4]

Table 1: **Cost function**

q	0	1	2	3	4	5	6	7	8	9	10
$c(q)$	0	1	3	6	9	13	18	23	28	33	40

Stage 4: Measuring Shared Understanding. To measure the extent to which the group's agreement had led to a shared understanding about appropriate behavior in new environments, we confronted participants with three possible scenarios for how the game might change in the second phase. For each scenario, subjects received detailed instructions explaining the new trading environment. To avoid order effects from systematically affecting the results, we randomized the order of presentation of the three scenarios. One of these scenarios, explained in detail below (see stage 5), was implemented in trading phase 2. Detailed descriptions of the other two scenarios can be found in the instructions reproduced in Appendix B.

For each scenario, we asked participants a set of questions that aimed at measuring the *normative* and the *behavioral* implications of the agreement. Specifically, to elicit the normative implications of the agreement, we asked participants for each scenario 1.a) "what should you do according to the agreement?" and 1.b) "what should your partner do according to the agreement?". We incentivized these two questions (1.a and 1.b) by paying participants a bonus of 20 experimen-

[2] In terms of real-life counterparts, the simultaneous nature of the game reflects a situation in which q is not known to the buyer when setting price p, and could, e.g., mirror a transaction of experience goods.

[3] Note that the cost function and all other parameters mentioned in the text are expressed in experimental points that were also used to explain the game to participants. The points had real value and were exchanged into the local currency at the end of the experiment. The exchange rate was: 25 Points = US$ 1.

[4] Note that given these parameters, a choice of $q = 10$ is efficient and maximizes group performance, as the marginal cost of providing the highest quality ($c'(10) = 7$) is smaller than the marginal benefit ($v' = 10$).

tal points for consensus with their partner (for each question in each scenario). For the questions targeting behavioral implications, i.e., actual choices, we asked 2.a) "what will you do in the first period?" and, to elicit the belief about the expected choice of the partner, 2.b) "what do you think that your partner will do in the first period?". Question 2.a) was incentivized by making the answer to that question the participant's binding first period choice in case that scenario was implemented. We decided not to incentivize question 2.b).

Stage 5: Second Trading Phase. Two elements in the game changed in the scenario that was actually implemented in the second trading phase compared to the first phase. First, the value of the product to the buyer doubled for every positive quality level. The new value was thus $v = 20q$. Second, we introduced an outside option for the buyer that was attractive in some periods. With probability 2/3 the outside option was worthless to the buyer ($v_o = 0$), but with probability 1/3 the outside option was attractive. Specifically, in this case the value of the outside option to the buyer was $v_o = 160$ (delivering a quality of $q^o_{hi} = 10$ at a price of $p^o_{hi} = 40$). If the buyer picked the outside option, the seller received a price of $p = 0$ for his product, but still incurred the production cost, in case he decided to provide a quality $q > 0$. In each period, the realized price and value of the outside option (p^o and q^o) were known to both the buyer and the seller before they made their decisions about what quality to provide and whether or not to take the outside option or to buy from the seller.

Increasing the value of the product to the buyer means that the payoff equalizing price was not the same anymore as in the first phase.[5] Adding the outside option means that the situation became riskier for the seller if he was unsure about whether the buyer would choose the outside option or whether she would choose to buy from him. For the buyer, the outside option was attractive (compared to buying from the seller) because once realized there was no uncertainty about q^o and p^o, whereas when buying from the seller, uncertainty remained about the quality q the seller decided to produce. These changes represent a tough test for the stability of relational contracts, because the sellers had to be quite certain about the buyers' actions in order to be willing to incur the cost of production in case an attractive outside option was present for the buyers.

[5] For instance, if the seller provided the highest quality ($q = 10$) and the buyer stuck to the price $p = 70$, which was payoff-equalizing in the first phase, the buyer earns $\pi_B = 140 - 70 = 70$ and appropriates the entire additional surplus from the increase in value, whereas the seller would not benefit at all from the additional surplus ($\pi_S = 70 - 40 = 30$).

Stage 6: Payments

Once the second trading phase was finished, participants were provided with a summary of their earnings in the experiment and received their payments in cash and in private.

2.2 The Nudge Treatment

Assessing whether principles causally affect parties' shared understanding and performance requires an exogenous intervention that fosters the emergence of principles. To this end, we added a "Nudge Treatment" with the following subtle experimental manipulation—merely a simple nudge—in the instructions and in the communication stage:

> "When finding an agreement, you should bear in mind that you do not yet know the exact situation you will encounter in the second part of the study. It may therefore be helpful to consider not only the first part of the study that you already know about, but also the second part that you do not yet have information about. For example, you could think about the principles on which you and the buyer would generally like to act during the study."

Importantly, no new information was added; the nudge only increased the salience of the fact that the game would change and suggested that participants think about the principles on which they would like to act.

2.3 Participants and Procedure

242 students (52% women, $M_{age} = 21.8$ years, $SD_{age} = 2.6$ years) from a university subject pool participated in our study and received the equivalent of US$ 10 as a show-up fee. Further payment depended on the performance in the study as outlined in the description of the experimental game above. Specifically, participants were paid for payoffs realized in the first and the second trading phase, as well as for aligned answers to the questions about the three scenarios. Average earnings amounted to US$ 49 per participant (including the show-up fee).

We conducted seven experimental sessions with a minimum of 32 and a maximum of 36 participants per session. Participants were randomly assigned to a computer cubicle upon arrival. Before each trading phase, participants received detailed written instructions and experimenters read aloud

a summary of the instructions before the start of each phase. Interactions were anonymous, i.e., buyers and sellers in a pair could not identify each other. The experiment was programmed in z-Tree (Fischbacher, 2007). The experimental manipulation was randomly assigned within each experimental session; there were 60 groups in the Baseline and 61 in the Nudge Treatment.

To implement the infinitely repeated game in the second trading phase, we followed Fudenberg et al. (2012) by randomly determining the number of periods to be played ex-ante and keeping it constant in all sessions. This procedure reduces variance and ensures a better comparability of results between sessions, without affecting the repeated-game incentives for participants. The random device set the number of periods to be played in the second phase to 12. In total, the experiment thus lasted for 17 periods, five periods in the first trading phase and 12 in the second phase.

2.4 Coding of Joint Statements and Chats.

To obtain numerical data on agreement content, three research assistants independently coded the joint statements as well as the chat protocols of each group. The coders coded for the presence of principles and rules. For the coding, we defined principles in the following way: "The participants formulate a principle that defines how to act in general. It is not based solely on a numerical definition of the quality to be delivered or the price to be paid ..., but provides overarching, general guidelines for action ... ." Rules were defined as clearly stating "numerically a quality and the price to be paid for it." The two coding categories were not mutually exclusive and groups could be coded as having formulated both a rule an a principle. Importantly, both rules and principles were only to be coded as present if there was a clear agreement between the parties on the respective rule or principle. In addition, the coders also coded for some sub-categories of rules and principles that we do not use in our main analyses in this paper. See the coder instructions reproduced in Appendix C for details.

The principles or rules could manifest themselves either explicitly in the joint statement, or be more implicitly formulated in the chat messages. For instance, to illustrate the latter case, a group could implicitly agree on a principle or a rule in the chat, but for some reason fail to write the principle or rule down in the statement. Coders coded separately for the appearance of principles and rules in the statement and in the chat protocol.

Table 2: **Percent agreement and inter-rater reliability of codings**

	Statements	*Chats*
Principle	.91	.80
	(.82)	(.59)
Rule	.98	.91
	(.96)	(.89)
N	121	

Notes: The table shows the percent agreement between raters. Gwet's AC (Gwet, 2008) is reported in parentheses.

Table 2 provides an overview of the percentage agreement and the reliability of the codings. We see that agreement among coders about the presence of principles and rules in statements and chats is generally very high (80-98%). Because of such high agreement, we use Gwet's AC score (Gwet, 2008) to assess inter-rater reliability. We can see that inter-rater reliability is very good for all categories of codings except the coding of principles in chats, where it is slightly lower.

3 Predictions

3.1 Framework and Terminology

Our study investigates whether relational contracts can enable organizations both to coordinate on efficient strategies and to adjust to unforeseen change in an efficient manner. To this end, we study a minimal organization—with just two members—in a repeated interaction that includes an exogenous change in their environment. The parties are aware that there will be a need for adaptation, but they cannot foresee the nature of the upcoming change.

After the shock, pairs face a game in which efficient cooperation can be sustained as an equilibrium of the infinitely repeated game.[6] A standard game-theoretic view would therefore be that the parties immediately adapt their strategies after the shock and jump to a new cooperative equilibrium.[7] In reality, however, there are good reasons to believe that the move to such a new equilibrium may not be trivial. Gibbons & Henderson (2012), for example, describe real-life adaptations com-

[6] If we assume that the players believe that the gains from trade will not diminish too much after the change, cooperation can also be sustained in equilibrium before the change.

[7] The folk theorem implies that there are many equilibria (including inefficient equilibria), but existing experimental work suggests that efficient equilibria that split the surplus equally are plausible focal points (see, e.g., Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000).

plicated by misunderstandings and coordination failure. They argue that a successful collaboration requires not only that there are equilibrium cooperation strategies available (the "credibility problem"), but also that the partners succeed in building a shared understanding of such strategies (the "clarity problem").

In the context of our experiment, a successful adaptation requires that the partners align their beliefs and actions about both the quality to be provided and the price to be paid. If this requirement is satisfied, we define a pair as *equilibrated*. If the agreed upon actions maximize the total surplus, we use the term *efficiently equilibrated*.

We argue that two elements are crucial for a pair to reach equilibriation once the environment has changed. First, partners must have a shared understanding of *what should be done* in the new situation. We call this requirement *normative consensus*. Second, a normative consensus leads to equilibration only if both parties actually *do what they think should be done*. In other words, normative consensus leads to equilibration only if it is self-enforcing. We use the term *normative-behavioral consistency* to describe a party's behavior that is in accordance with what that party thinks is the normatively correct thing to do.

All the game-theoretic models of cooperation that we know are silent about the process that may lead to equilibration.[8] In these models equilibration is taken to be so simple that it does not appear in the model. In contrast, we argue that real-life equilibration is a difficult process related to communication and the establishment of agreements among partners. Especially in changing environments, it is important to understand how an agreement can be formulated such that the parties know what to do even after the world around them has changed. We therefore conjecture that agreements that establish general principles rather than specific rules are desirable in unstable environments. Rather than prescribing detailed actions that may be inappropriate under new conditions, principles define guidelines that can be applied to govern interactions even when circumstances change.

[8]To be clear, we mean applied models of cooperation, not abstract models of fictitious play, rationalizability, evolutionary games, or the like.

3.2 Testable Hypotheses

Our first hypothesis describes our predictions regarding the expected performance of different types of agreements that emerge endogenously in our Baseline setting. Our thinking is that pairs who succeed in agreeing on a principle rather than relying on a rule have a better chance to produce both normative consensus and normative-behavioral consistency. As a consequence, we expect that those pairs who establish principle-based agreements will outperform the pairs who govern their relationship with a rule. Hypothesis 1 summarizes this chain of arguments:

Hypothesis 1 (Effects of Endogenously Emerging Principle-based Agreements)

a) ***Performance:*** *Buyer-seller pairs with principle-based (rather than rule-based) agreements achieve higher levels of performance after an exogenous change in the environment.*

b) ***Mechanisms:*** *Buyer-seller pairs with principle-based (rather than rule-based) agreements are more likely to reach both normative consensus and normative-behavioral consistency and are therefore more likely to be efficiently equilibrated after an exogenous change in the environment.*

Our first hypothesis refers to correlations between endogenous variables within our Baseline setting. Accordingly, evidence supporting this hypothesis will not allow us to make any causal claims regarding the effects of principle-based agreements on outcomes. Our Nudge Treatment therefore aims at exogenously triggering the emergence of more principle-based agreements that allow the parties to reach higher performance levels. Hypothesis 2 summarizes our expectations.

Hypothesis 2 (Causal Effects of the Nudge Treatment)

a) ***Agreement Types:*** *Buyer-seller pairs in the Nudge Treatment are more likely to establish principle-based (rather than rule-based) agreements than pairs in the Baseline setting.*

b) ***Performance:*** *Buyer-seller pairs in the Nudge Treatment are more likely to achieve higher levels of performance after an exogenous change in the environment than pairs in the Baseline setting.*

c) ***Mechanisms:*** *Buyer-seller pairs in the Nudge Treatment are more likely to reach both normative consensus and normative-behavioral consistency and are therefore more likely to be*

efficiently equilibrated after an exogenous change in the environment than pairs in the Baseline setting.

4 Results

In the first part of this section, we present analyses testing Hypothesis 1 focusing on the effects of endogenously emerged principles in the Baseline condition. In the second part, we analyze the results of our Nudge treatment (Hypothesis 2).

4.1 Baseline Outcomes

In this subsection, we investigate (a) the frequencies of rule-based and principle-based agreements in our experimental environment and (b) how the different types of agreements correlate with performance. In addition, we shed light on the mechanisms underlying the observed correlation of agreement type and performance. As there is no exogenous variation and agreements are established endogenously, all results in this subsection are purely correlational.

4.1.1 Rules vs. Principles: Emergence and Performance

Our first result establishes the relative frequency with which different types of agreements emerge and how they correlate with the performance of the corresponding groups:

Result 1

a) A large majority of groups (70%) establishes rule-based rather than principle-based agreements.

b) Groups who formulate principle-based agreements tend to achieve significantly higher levels of performance than those who rely on rule-based agreements.

Table 3 provides support for the first part of Result 1. The table is based on our coding data and shows an overview of the relative frequency (in percent of all groups) with which principle-based and rule-based agreements endogenously emerged in the baseline condition, separately for the statement, the chat as well as when combining both of them.

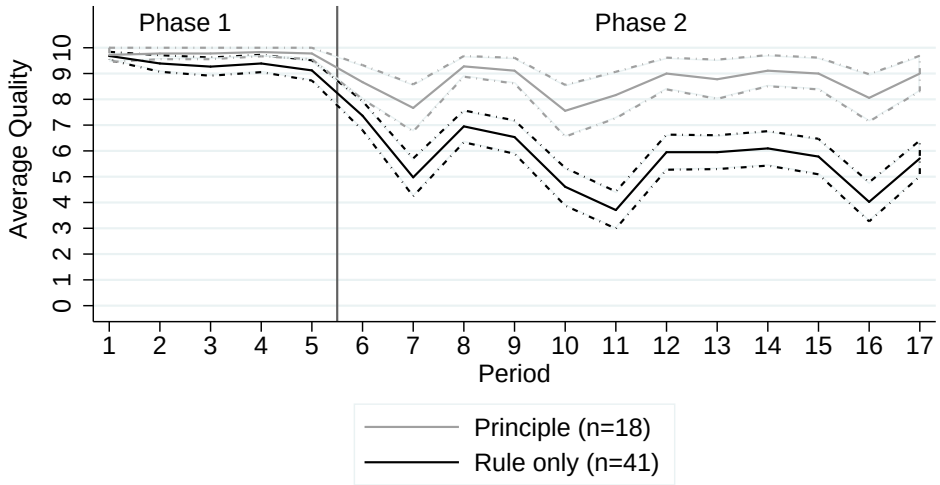
Table 3: **Relative Frequency of Rules and Principles in Baseline**

	Statements	*Chats*	*Combined*
Principle	.133	.250	.300
Rule	.883	.950	.983
N		60	

Notes: The table shows the coded relative frequency of groups (in percent of all groups) who included principles and rules in their statements in the baseline. The first column contains only codings for principles and rules in the final written statements. The second column contains only codings in chat messages. The third column combines all codings in statements and chat messages.

We consider it likely that principles and rules from both the statements and the chats could have an impact on the implicit agreement. Consequently, in our analyses we rely on the combined coding from both statements and chats. This means that, in what follows, we say that a group agreed on a principle whenever a principle was mentioned in the statement, the chat, or in both. When considering the combined data, we observe that (with the exception of one single pair) all groups wrote down a rule, whereas only 30% of all groups also articulated a principle. Interestingly, there are no groups who governed their interaction with a principle alone—i.e., all groups who formulated a principle also specified its implication for the current setting in the form of a rule.

Figure 2: **Quality in Groups with Principles vs. Rules (Baseline Only)**



Notes: Only data from the baseline included. One of the 60 groups in the baseline had neither a principle nor a rule and is thus not included in the graph. The dashed lines represent plus/minus one standard error of the mean.

Figure 2 illustrates the second part of Result 1. The Figure shows the development of average quality provided by sellers over time and distinguishes between groups with an agreement that only contained a rule and groups who also formulated a principle. For the interpretation of the figure it is important to note that in our setting the seller's quality choice fully determines overall performance of the relationship. The figure reveals that groups with principle-based agreements clearly outperformed groups that relied solely on rules. Indeed, and in line with Hypothesis 1a, when using total group payoffs in Phase 2 as an overall measure of performance, we find that groups with principles had average earnings of 1752 points, whereas groups with rules earned 1348 points on average. This difference is statistically significant ($t(57) = 3.43$, $p < .01$). In contrast, when comparing average earnings in Phase 1, we find no significant difference ($t(57) = 0.90$, $p = .37$).

4.1.2 Rules vs. Principles: Equilibration and Normative Consensus

As discussed in Section 3, a fruitful buyer-seller collaboration requires two elements: First, beliefs and actions about chosen quality and price need to be equilibrated. Second, the trading partners need to be equilibrated on an efficient outcome. We operationalize equilibration in the following way. In the quality dimension the seller chooses a quality that is at or above the buyer's quality expectation. In the price dimension, the requirement is that the buyer chooses a price that is at or above the seller's price expectation. For efficient equilibration, we additionally require that the seller chooses a quality of 10. Note that every group that is efficiently equilibrated is also equilibrated, but not vice versa.

To measure the degree of equilibration, we construct an *Equilibration* score and an *Efficient Equilibration* score, capturing in how many of the three scenarios presented to participants before Phase 2 a group satisfied the respective requirements. Recall that the behavioral choices in these scenarios were incentivized by rewarding agreement within a pair (for normative consensus) and by making the answer the first-period choice in the second trading phase if that scenario was implemented (for behavioral choices). We then explore whether groups with principle-based agreements were more likely to achieve (efficient) equilibration, and whether these (efficient) equilibration scores are correlated with high performance. Our findings are summarized in our second result:

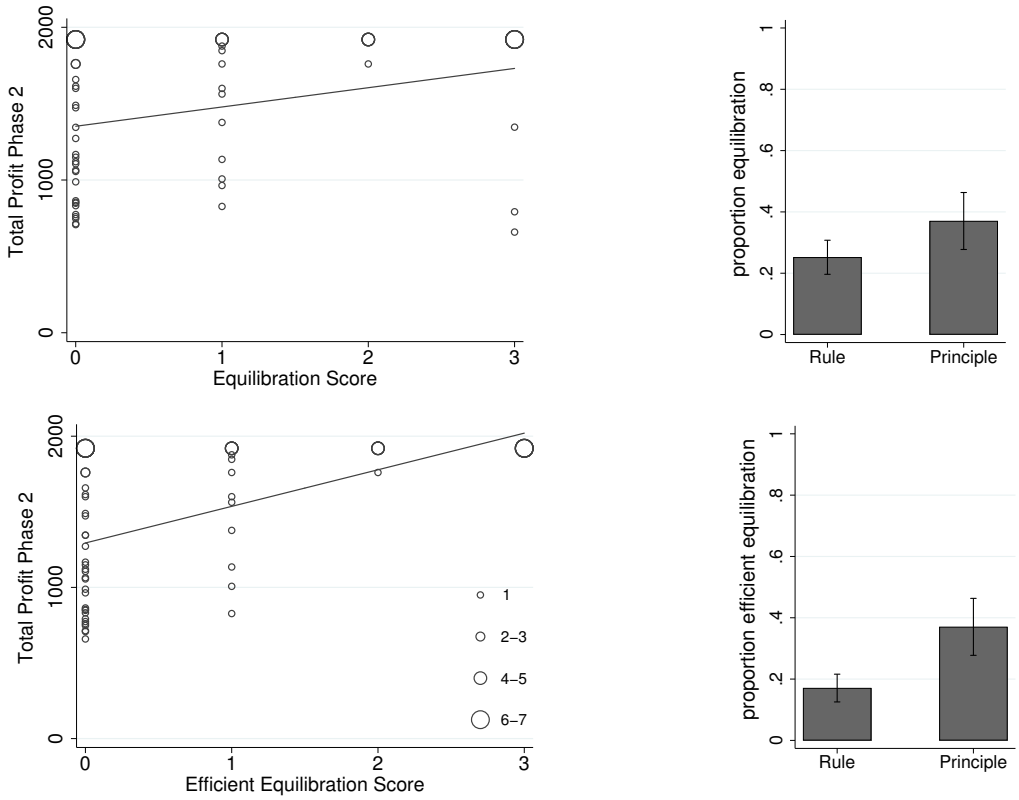
Result 2

a) Equilibrated groups reach significantly higher performance levels, particularly if they are efficiently equilibrated.

b) Groups with principle-based agreements are significantly more likely to be efficiently equilibrated.

The left side of Figure 3 shows evidence for the first part of Result 2. There is a clear correlation between performance, measured in terms of aggregate profits in Phase 2, and the degree of equilibration as measured by the equilibration scores ($r = 0.30$, $p = .02$ for equilibration, see top-left panel; and $r = 0.53$, $p < .01$ for efficient equilibration, see bottom-left panel).

Figure 3: **Equilibration, Performance and Agreement Type**



Notes: The top panels focus on equilibration. The top-left panel shows the correlation between equilibration and performance. The top-right panel shows the degree of equilibration for groups with rule-based vs. principle-based agreements. The bottom panels focus on efficient equilibration. The bottom-left panel shows the correlation between efficient equilibration and performance. The bottom-right panel shows the degree of efficient equilibration for groups with rule-based vs. principle-based agreements. Only data from the Baseline included. The left panels are based on all 60 groups. In the right panels, the one group without a rule- or principle-based agreement is excluded (n=59). The bubble size legend in the bottom-left panel also corresponds to the plot in the top-left panel.

The two graphs provide further interesting insights: First, the top-left panel illustrates that equilibration alone is only a partial predictor of performance: there are three groups that are fully equilibrated, but do not achieve high profits. A comparison with the bottom-left panel shows that these were inefficiently equilibrated groups—i.e., groups whose actions and expectations were aligned on an inefficient quality level. Second, the bottom-left panel reveals that efficient equilibration is a very strong predictor of performance. In fact, 10 out of the 11 groups who achieved an efficient equilibration score of at least 2 achieved the maximum profit of 1920 points in phase 2 of the experiment. Once the efficient equilibration score falls below 2, however, many groups do not achieve efficient performance.

It is important to note that the correlation between our measures of equilibration and performance is by no means tautological. The efficient equilibration measure is based on beliefs and actions that participants indicated when asked about behavior in the three different scenarios. Only one of these scenarios was eventually implemented and the collected responses only concern behavior in the first period of a given scenario. Performance, in contrast, is measured over all 12 periods of the implemented scenario in which buyers and sellers interacted in the second phase of the experiment. Still, a more conservative approach is to look at the correlation between performance in Phase 2 and the equilibration score based only on the two non-implemented scenarios. The results are basically unchanged. The correlations then are $r = 0.28$ ($p = .03$) between performance and equilibration, and $r = 0.46$ ($p < .01$) between performance and efficient equilibration.

The right-hand side of Figure 3 provides support for the second part of Result 2. The two panels show the proportion of groups that achieved equilibration and efficient equilibration, conditional on having a rule-based agreement versus a principle-based agreement. In line with Hypothesis 1b, the bottom-right panel shows that groups with principles were 20 percentage points more likely to achieve efficient equilibration ($t(57) = 2.18$, $p = .03$). The effect on equilibration only amounts to 12 percentage points and is not significant.

If principles are correlated with equilibration, which in turn is correlated with performance, a logical next step is then to ask what the necessary requirements are for a pair to be equilibrated. As discussed in Section 3, two elements are crucial in this context. First, trading partners need to have a shared understanding of *what should be done* in a particular situation (*normative consensus*). Second, normative consensus is useful only if both parties actually *do what should be done* according

to the agreement (*normative-behavioral consistency*).

The next result confirms that there are important correlations between normative consensus and normative-behavioral consistency, on the one hand, and equilibration on the other. It also shows why groups with principle-based agreements are more likely to reach equilibration.

Result 3

a) Equilibration is rarely reached without the joint presence of normative consensus and normative-behavioral consistency.

b) Groups with principles are not significantly more likely to develop an (efficient) normative consensus about the actions their agreement prescribes in a new situation, but they do display stronger normative-behavioral consistency.

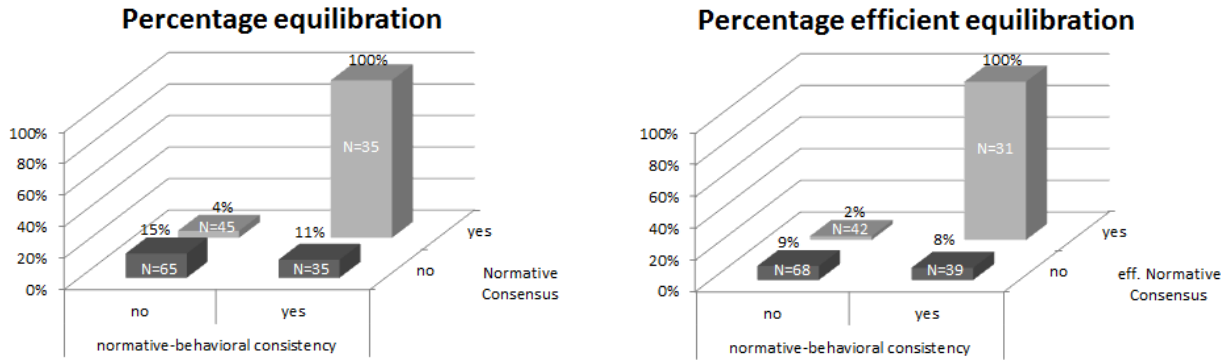
According to our definition of equilibration, normative consensus and normative-behavioral consistency are sufficient but not necessary conditions for groups to be equilibrated. Theoretically, it is possible to achieve equilibration without normative consensus and normative-behavioral consistency, but Figure 4 illustrates that such equilibrations are rarely observed in the data, and hence provides the evidence for Result 3a.[9] Thus, we find that both normative consensus and normative-behavioral consistency are (almost) required for equilibration.[10]

Evidence for Result 3b stems from Table 4. The table shows marginal effects from logit regressions of a principle dummy on dummy dependent variables indicating whether a group (1) achieved normative consensus, (2) achieved efficient normative consensus, and (3) showed normative-behavioral consistency at the group level. There are 59 groups with a rule or a principle, and each group replied to three scenarios, so there are 177 observations. The first two rows show

[9]Those groups who were equilibrated and had normative-behavioral consistency, but lacked normative consensus are usually cases in which, for example, the seller's belief about the buyer's price choice was correct, but the seller stated a higher normative price. Those groups that had neither normative consensus nor normative-behavioral consistency but are equilibrated are usually groups in which one party interpreted the agreement (usually rule) literally but acted according to the spirit of the agreement (the buyer stating a price of 70 but then paid 120). Finally, those groups that had equilibration with normative consensus but without normative-behavioral consistency are usually groups in which the buyer lowered the price relative to her normative price statement, but both the actual as well as the normative price were above the expectation of the seller.

[10]Normative-behavioral consistency for one party is strongly correlated with beliefs in the normative-behavioral consistency of the other party. Among the groups with normative consensus, parties who do not believe that the other party will act consistently with that consensus choose to act inconsistently themselves in 85% (39 of 46) of the cases. Parties who believe that the other party will act consistently, in contrast, act consistently themselves in 82% (94 of 114) of the cases.

Figure 4: **Normative Consensus, Normative-Behavioral Consistency, and Equilibration**



Notes: Only data from the Baseline included ($n = 60$). The horizontal axes categorize groups based on whether they reached normative-behavioral consistency and (efficient) normative consensus. The vertical axes plot the percentage of groups who reached (efficient) equilibration.

that although groups with principles were more likely to achieve normative consensus and efficient normative consensus (by 7 and 13 percentage points, respectively), these effects are not statistically significant. However, row (3) shows that groups with principles were 28 percentage points more likely to display normative-behavioral consistency, a highly significant effect.

Table 4: **Marginal Effects of Principles on Normative Consensus and Normative-Behavioral Consistency**

	Marginal Effect	Standard Error
(1) Normative Consensus	0.07	(0.10)
(2) Efficient Normative Consensus	0.13	(0.10)
(3) Normative-Behavioral Consistency	0.28***	(0.10)
Observations	177	

Notes: The table reports marginal effects from logit regressions of a principle dummy on the the dummy dependent variable indicated in the first column. Data stem from 59 groups that replied to 3 scenarios each. Standard errors clustered at the group level are in parentheses. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .10$

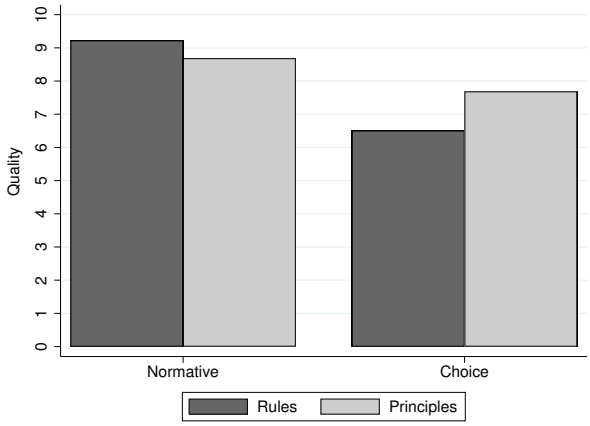
The higher degree of normative-behavioral consistency implies that the normative consensus in groups with principle-based agreements was more likely to be self-enforcing after the exogenous shock, whereas in groups with rules there seemed to exist substantial incentives to deviate. As a consequence, groups with principle-based agreements were more likely to be equilibrated.

To illustrate, consider the situation in the implemented scenario, in which the value of the prod-

uct doubles in comparison to the first phase (and the outside option is not attractive for the seller). If we compare the normative consensus in this situation between groups with agreements based on rules versus principles, we find that they differed mainly with regard to the price. Specifically, 92% of the groups with agreements based on principles and efficient normative consensus normatively agreed that the price should be raised to the level that guarantees an equal split of profits when the seller provides maximum quality. In contrast, groups with agreements based on rules and efficient normative consensus only did so in 64% of the cases (this difference is marginally significant: $z = 1.77$, $p = .08$).[11] Thus, groups with agreements based on principles were more likely to build an efficient normative consensus that ensured an equal split of profits.

Importantly, sellers also expected buyers to follow through on the normative consensus, meaning that they would actually receive an equal share of the increased profit, and vice versa buyers did expect the sellers to provide the highest quality. The effects of these expectations on performance can be seen in Figure 5, which displays the average quality levels (across the three scenarios) the sellers indicated they *normatively* should deliver and the quality they *actually* decided to deliver for groups with rules versus groups with principles.

Figure 5: **Rules vs. Principles: Normative and Actual Average First-Period Quality Choices**



Notes: Only data from the Baseline included. The chart shows average normative quality choices and actual quality choices for groups with rules and principles separately.

The left side of Figure 5 shows that average assessments of the normative quality to be provided

[11] Note that the price that induced an equal split of profits in response to maximum quality was 70 in Phase 1 and 120 in Phase 2. No group had normative consensus on a price higher than 120 in the described situation. Moreover, 20% of the groups with rules agreed on a normative price of 70, whereas none of the groups with principles did.

according to the initial agreement were relatively close to the efficient level of 10 and did not differ significantly between groups with agreements based on rules versus principles ($t(57) = 1.10$, $p = .28$). In contrast, the right side shows two things about the quality actually provided in the first period: (1) it was lower than stated in the normative consensus in both groups with agreements based on rules ($t(40) = 6.47$, $p < .01$) and groups with agreements based on principles ($t(17) = 2.80$, $p = .01$), but (2) the difference from the normative consensus is smaller for groups with agreements based on principles ($t(57) = 2.53$, $p = .01$). Thus, although the normative assessment of the quality to be provided in groups with agreements based on rules was not less efficient than in groups with agreements based on principles, sellers in groups with agreements based on rules were less likely to act upon that normative assessment and were more likely to deviate downwards in their actual quality choices. Again, this illustrates the deviation incentives in groups with rule-based agreements: their normative consensus was not self-enforcing. The normative consensus in groups with principle-based agreements was more likely to be self-enforcing, leading sellers to follow through with the efficient quality choice in the first period.

4.2 Experimental Manipulation: Nudging the Emergence of Principles

So far we have shown that groups who endogenously succeed in building principles are more likely to be equilibrated and more successful in terms of performance in phase 2 of the experiment. However, these results are purely correlational. In this subsection we attempt to establish a causal link between principle-based agreements and performance using an exogenous intervention. Toward this end, we compare the data of our "Nudge Treatment" with the data of our Baseline condition. Recall that the experimental manipulation in the Nudge Treatment consisted of only a short reminder in the experimental instructions, making the fact that the game was going to change in the second trading phase more salient, and telling participants it could be beneficial to take this into account when communicating, for instance by trying to include a general principle about how to act in their written statement. Our fourth result summarizes the observed impact of our treatment:

Result 4 *The Nudge Treatment successfully triggered the emergence of more principle-based agreements and had a significant impact on quality choices of sellers in the first period after the shock. However, in Phase 2, our intervention did not significantly increase overall performance.*

First, in line with Hypothesis 2a, the nudge treatment was successful in inducing subjects to develop principle-based agreements. Coders found far more principles in the Nudge Treatment (in 52 of 61 or 85.2% of groups) than in the Baseline (in 18 of 60 or 30.0% of groups). This difference is highly significant ($z = 6.15$, $p < .01$).

But how did the treatment affect the actions taken by sellers and buyers? We begin our analysis by focusing on sellers' quality choices in the three scenarios. Table 5 shows the results of a regression of average quality chosen in each of the 3 scenarios on a Nudge Treatment dummy. The dataset is comprised of 363 observations (121 sellers, 3 observations per seller). Standard errors are clustered at the group level. Column (1) shows a marginally significant treatment effect. In the Nudge treatment, sellers on average chose a quality that was 0.92 points, or 14 percent, higher than in the Baseline treatment ($p = .08$). Column (2) shows the same regressions, but additionally includes fixed effects for the different scenarios. The result is basically unchanged.

Table 5: **Treatment Effect on Sellers' Quality Choices in the Three Scenarios**

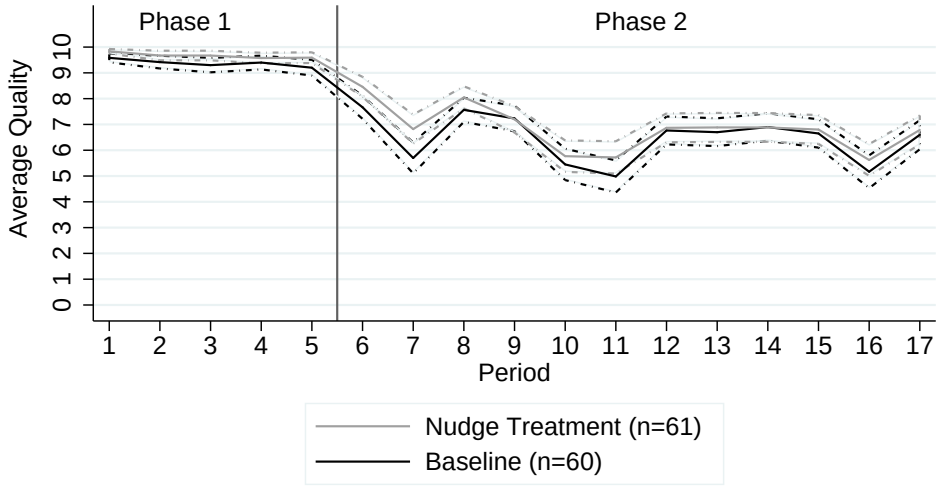
	(1)	(2)
Nudge Treatment	0.92*	0.92*
	(0.52)	(0.52)
Constant	6.78***	6.60***
	(0.36)	(0.40)
Scenario FE's	No	Yes
Adj. R^2	0.01	0.01
Observations	363	363

Notes: OLS regressions based on the quality decisions of 121 sellers in the 3 scenarios. Standard Errors clustered at the group level are in parentheses. ***$p < .01$, **$p < .05$, *$p < .10$

Having established that the treatment induced sellers to provide a marginally significant higher quality in the three scenarios, we can next test whether the treatment also had a lasting effect on the long-run performance of the groups in Phase 2. Figure 6 displays the performance in the Baseline and the Nudge Treatment over the entire experiment. First, groups in the Baseline and the Nudge Treatment did not show performance differences in Phase 1 ($t(119) = 1.04$, $p = .30$). Moreover, the Nudge Treatment also did not significantly affect performance in Phase 2, after the unforeseen contingency materialized. Average quality was slightly higher in the first two periods of Phase 2, but this difference is not significant. Thereafter, the difference in average performance becomes

very small. Overall, average quality levels were very similar in the Nudge Treatment (6.82) and the Baseline treatment (6.45; $t(119) = 0.58$, $p = .56$), and average earnings turn out to be 26 points higher in the Nudge Treatment ($t(119) = 0.29$, $p = .77$); both these differences go in the expected direction, but they are not statistically significant, thus providing no support for Hypothesis 2b.

Figure 6: **Quality in Baseline vs. Nudge Treatment**



Notes: The dashed lines represent plus/minus one standard error of the mean.

Why do we observe a marginally significant treatment effect on first-period quality choices in the three scenarios, but only a small and insignificant effect on overall performance? To answer this question, we next analyze whether the increase in principles affected the degree of equilibration in the treated groups. In the spirit of Table 4, Table 6 shows the marginal effects (obtained via logit regressions) of the Nudge Treatment on dummy dependent variables capturing whether a group (1) achieved normative consensus, (2) achieved efficient normative consensus, (3) was normative-behaviorally consistent at the group level, (4) was equilibrated, and (5) was efficiently equilibrated. The first two rows show that the Nudge Treatment had a small and insignificant effect on normative consensus. The third row shows a significant treatment effect on normative-behavioral consistency at the group level: groups in the Nudge Treatment were 16 percentage points more likely to act consistently. However, the fourth and fifth row show no significant effect on equilibration, thus providing no evidence in support of Hypothesis 2c.

How can these findings be reconciled? The fact that the treatment increased efficient equilibra-

Table 6: **Marginal Effects of the Nudge Treatment on Normative Consensus, Consistency, and Equilibration**

	Marginal Effect	Standard Error
(1) Normative Consensus	0.01	(0.06)
(2) Efficient Normative Consensus	0.03	(0.06)
(3) Normative-Behavioral Consistency	0.16**	(0.07)
(4) Equilibration	0.02	(0.06)
(5) Efficient Equilibration	0.04	(0.06)
Observations	363	

Notes: The table reports marginal effects from logit regressions of a Nudge Treatment dummy on the dummy dependent variable indicated in the first column. Standard errors clustered at the group level are in parentheses. ***$p < .01$, **$p < .05$, *$p < .10$

tion by only an insignificant 4 percentage points is consistent with the lack of a treatment effect on performance, because we hypothesized that an increase in performance would be driven by increased efficient equilibration. But why did the increase in normative-behavioral consistency not translate into increased equilibration? Table 7 provides further insights into this question. In this table, we decompose equilibration and normative consensus into their price and quality components. Normative consensus on quality means the parties normatively agree on a quality level. Equilibration in quality means the quality level actually chosen matches the buyer's belief. Normative consensus and equilibration on price are defined analogously. Based on logit regressions, the table shows the marginal effects of the Nudge Treatment on equilibration (1) and efficient equilibration (2) in the quality dimension, normative consensus (3) and efficient normative consensus (4) in the quality dimension, as well as equilibration (5) and normative consensus (6) in the price dimension.

Table 7 provides several interesting insights. First, it shows that the Nudge Treatment did cause groups to achieve significantly higher equilibration on quality (1 and 2). This finding reveals that the above reported increase in normative-behavioral consistency was associated with more efficient equilibration in the quality dimension. However, row (5) reveals that the Nudge Treatment failed to increase equilibration in the price dimension. Instead, we even see a (statistically insignificant) decrease. But since efficient equilibration requires equilibration on both quality and price, the failure of the treatment to increase equilibration in price begins to explain the lack of an overall

Table 7: **Effects of the Nudge Treatment on Normative Consensus and Equilibration on Quality and Price**

	Marginal Effect	Standard Error
(1) Equilibration on Quality	0.13**	(0.07)
(2) Efficient Equilibration on Quality	0.16**	(0.07)
(3) Normative Consensus on Quality	0.00	(0.04)
(4) Efficient Normative Consensus on Quality	0.00	(0.06)
(5) Equilibration on Price	-0.05	(0.04)
(6) Normative Consensus on Price	0.01	(0.06)
Observations	363	

Notes: The table reports marginal effects from logit regressions of a Nudge Treatment dummy on the dummy dependent variable indicated in the first column. Standard errors clustered at the group level are in parentheses. ***$p < .01$, **$p < .05$, *$p < .10$

treatment effect on efficient equilibration.[12]

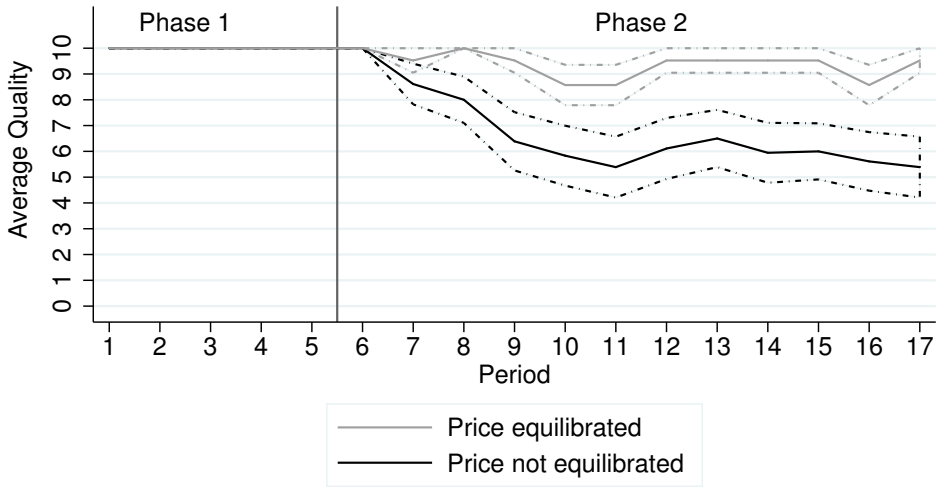
To illustrate the importance of achieving equilibration in *price*, Figure 7 displays average *quality* over the course of the experiment for all groups in the Nudge Treatment that achieved efficient equilibration on quality ($N = 39$) in the scenario that was eventually implemented in Phase 2 of the experiment. The figure then displays the development of quality separately for groups who achieved equilibration on price versus groups who did not. Groups that achieved equilibration on price stayed very close to the efficient quality level throughout (average quality was 9.4 over the 12 periods of Phase 2). Groups who were not equilibrated on price also started out at the efficient quality level of 10 in period 6 (by construction of the sub-sample), but then experienced a constant deterioration of average quality over time (average quality was 6.6 over the 12 periods of Phase 2). As a consequence, groups without equilibration in price achieved significantly lower profits in Phase 2 of our experiment ($t(37) = 2.73$, $p < .01$).[13]

In sum, our Nudge Treatment caused groups to reach more principle-based agreements and in causing sellers to provide significantly higher quality in the first period of the three scenarios. But the treatment was not effective in significantly increasing equilibration in the price dimension (and hence also failed to increase equilibration in general). In particular, groups that failed to achieve

[12]Empirically, it turns out that equilibration on price was harder to achieve. Overall, we observe equilibration on price in 53% of the cases, whereas equilibration in quality was achieved in 67% of the cases. Arguably, increasing equilibration in price would therefore have been the more important dimension to affect overall equilibration.

[13]This result also holds when including observations from the control treatment ($t(59) = 3.31$, $p < .01$) as well as when conditioning on equilibration instead of efficient equilibration ($t(42) = 2.00$, $p = .05$).

Figure 7: **Average quality conditional on efficient equilibration in quality, by equilibration in price**



Notes: All groups that achieved efficient equilibration on quality in the Nudge Treatment are included. The sample is then split into groups who also achieved equilibration in prices vs. groups who did not. 39 groups are included in this subsample, 21 of which achieved equilibration in price. The dashed lines represent plus/minus one standard error of the mean.

equilibration in the price dimension experienced significant decreases in quality over Phase 2, even though most of these groups did have principle-based agreements. As a consequence, we observe only a small and insignificant increase in average long-run performance in the treatment.

5 Discussion

5.1 Summary

In this paper—inspired by Kreps' and Schein's conceptions of organizational culture—we study relational contracts that not only specify how to behave in a given environment but also help the group to adapt their behaviors after an unforeseen shock. We distinguish between rule-based relational contracts that define specific actions versus principle-based relational contracts that provide general guidelines. In our Baseline data, agreements that include a principle rather than just a rule were more successful. Although groups with rule-based agreements achieved a similar level of shared understanding regarding what should be done in a new situation (normative consensus), they were more often unwilling to act accordingly (lack of normative-behavioral consistency). Groups that

relied on principles, in contrast, more often reached a normative consensus that was self-enforcing, even in unanticipated situations. As a result, the latter groups were more likely to be efficiently equilibrated after an exogenous shock and to transition their relational contract smoothly to a new environment.

To assess causality, we tried to stimulate the emergence of principle-based agreements using a subtle salience Nudge. Our treatment produced significantly more principle-based agreements and—when measured immediately after the change in environment—also led to higher levels of normative-behavioral consistency and an increase in equilibration in the quality dimension. However, the principles induced by our treatment failed to increase equilibration on price. In other words, whereas the Nudge increased coordination on efficient outcomes (the quality dimension), it failed to align groups on the distribution of payoffs (the price dimension). These conflicting views about distribution may have been the cause of the deterioration of cooperation over time that we observed in the post-shock game. As a result, the Nudge Treatment did not have a significant effect on performance in the long run and hence failed to provide evidence for causal links from principle-based agreements to equilibration or performance.

5.2 Causal Effects of the Nudge Treatment on Equilibration and Performance

To derive implications from our results it is important to understand why our treatment did not increase long-run performance but did induce more pairs to produce principle-based agreements.

A pessimistic interpretation of our findings is simple: there is no causal link between principles and performance. Instead, the correlation between principles and performance observed in the Baseline could reflect omitted-variables bias—i.e., groups that endogenously articulated principles may have performed better because other unobserved factors affected both principles and performance. Exogenously increasing the frequency of principles, without affecting these other factors, would then not produce an increase in performance.

This pessimistic interpretation is inconsistent with the fact that we do observe significant treatment effects on some factors that contribute to performance, such as equilibration in quality and normative-behavioral consistency. If the correlation between principles and performance in the Baseline were only a consequence of omitted-variable bias, the Nudge Treatment should not have had an impact on any performance-related variables. Our findings therefore indicate that principles

can play some role in the equilibration process in unstable environments.

A second interpretation involves the measurement of principles. Ideally, we would like to measure whether pairs agreed on an effective principle. Of course, we can measure only whether a principle—defined according to our codebook—appeared in the chat or in the written statement. As a consequence, there may be two types of errors in our data: (1) We may code pairs as not having a principle when in fact they have agreed to one implicitly, without writing it down. Or (2) we may code pairs as having a principle when in fact they have not really reached a shared understanding of a general principle but instead have only written language that meets the criteria in our codebook. We discuss these two errors in turn.

First, explicitly articulating an effective principle may be hard, so some pairs in the Baseline may have concentrated on discussing a rule, while implicitly understanding that their rule is a manifestation of a broader principle. Even the most careful codings of text can capture only what people actually write down, not what they may mean or think. Thus, in the Baseline, pairs that are coded as having only rules may in fact have also agreed implicitly on effective principles. Because the Nudge Treatment pushes participants to consider general principles, some pairs may have been more likely to make an implicit principle explicit and hence would be coded as having a principle. In short, our intervention could have caused only a relabeling of some pairs from having only a rule to having a principle, when in fact they had an effective if implicit principle regardless of labeling.

Second, again because formulating an effective principle may be hard and even the most careful codings of text cannot capture what may be in people's heads, the Nudge Treatment may have induced some groups to write an ineffective principle. If some pairs wrote down a principle simply because they felt that they had to, it is possible that those principles were not effective in equilibrating actions in unforeseen contingencies. This would lead to an increase in measured principles but those groups would be unlikely to change their behaviors in useful ways. In short, our intervention could have caused an increase in ineffective principles but no increase in effective principles.

Both of these errors imply that our Nudge Treatment may have been weaker than it seems from our coding. Ultimately, many fewer groups may have been treated, in the sense of the Nudge Treatment causing them to agree on an effective principle. Any combination of relabeling and/or ineffectiveness could explain the lack of a treatment effect on equilibration and long-run performance. At the same time, we can reject that the relabeling and ineffectiveness interpretations

explain the whole treatment effect on the emergence of effective principles because we did find treatment effects on equilibration in quality and on normative-behavioral consistency.

5.3 Implications for Practice and Research

While our Nudge Treatment did positively affect normative-behavioral consistency, one of the main motivations of our research program was not achieved in this initial paper: the treatment failed to help parties achieve better long-run performance. Perhaps one should not be surprised at this failure. For example, Barney (1986) argued long ago that if organizational culture is to create sustained competitive advantage then the culture must be hard to imitate. Of course the same is true for relational contracts: if the success stories described by Gibbons & Henderson (2012) were easy to imitate, presumably they would be run-of-the-mill stories rather than success stories. So it may be no surprise to practitioners that our (rather muted) intervention did not have a significant effect on long-run performance, which in turn can be seen as tough news for researchers hoping for causal inference: if researchers are interested in something that practitioners find hard to achieve in the world, it may be hard to achieve it even in the lab.

Accepting these concerns about the difficulty of our long-run research program, we nonetheless see this initial paper as having made a useful contribution. More specifically, while there are ethnographies and other studies that provide rich field data about people's efforts to build and change equilibria, our paper is the first laboratory study we know to provide data on what people are thinking as they try to build an equilibrium in a controlled setting. That is, although the use of principles in our Baseline data is endogenous, the facts that pairs using principles achieve not just higher performance but also higher normative-behavioral consistency seem to us to be correlations worthy of further attention from both academics and (if suitably translated) practitioners.

Regarding future research, especially in organization theory and organizational economics, we hope that those seeing themselves as contributing to the research streams described in Section 1—concerning not just organizational culture but also the relevant aspects of transaction-cost economics, the behavioral theory of the firm, informal organization, and new organizational forms—will consider not just the potential role of relational contracts once they are built but also the challenges in building and changing relational contracts over time. In addition, we hope for a close dialogue between ethnographic work on organizational change and the issues we have emphasized

here.

In addition to these literatures concerning organizations, we also hope to influence the three other (more abstract) literatures in economics mentioned in the Introduction: experiments on repeated games, experiments on communication, and game theory itself.

Our results illustrate that it is in no way guaranteed that parties in a repeated game can reach an equilibrium—let alone an efficient one—and that there exist systematic differences in this regard between groups (see also Proto et al., 2019). The experimental literature on equilibrium selection is informative on how differences across game environments drive equilibrium selection (see, for example, Van Huyck et al., 1990; Goeree & Holt, 2001), but it is largely silent about differences across groups within the same environment. Our paper makes a step towards a better understanding of what drives such heterogeneity, even within the same experimental conditions, by showing that different shared understandings with systematic performance implications can develop even under identical initial conditions.

Similarly, we advance the study of communication effects in social-dilemma settings by shedding light on what kind of non-binding agreements help sustain cooperation in the face of future uncertainty. To do so, we developed a novel experimental design that implements considerable uncertainty for participants and allows measuring how well they adapt to this uncertainty. By showing that principles are a useful tool to achieve equilibration in such uncertain settings, our results go beyond showing that communication helps: we provide an indication of how successful agreements should be shaped.

More generally, our results also point towards the need to understand behavioral heterogeneity in repeated-game settings. Much of the literature at least implicitly argues that changes in parameters should be evaluated against the comparative-statics predictions of how the efficient equilibrium changes when parameters change, which is akin to an interpretation that all groups in a given condition behave homogeneously. But given the heterogeneity across groups in a given setting we observed, a focus on heterogeneous treatment effects is warranted.

Finally, all of these observations about experimental economics have analogs in game theory. As noted in Section 1, essentially all repeated-game models assume the parties to be in equilibrium from the beginning of their relationship. Such models may capture shared understandings developed over time in relationships that began long ago, but that does not imply that real parties can

easily create such shared understandings before their relationship has begun. Moreover, repeated games have a huge number of equilibria, ranging (under suitable parameters) from efficient to quite the opposite, so if the parties do reach an equilibrium, there is then the question of which of these many equilibria they reach. Most models assume that parties not only reach (in fact, begin in) an equilibrium but also select an efficient one. We find that different shared understandings with systematic performance implications can develop even under identical initial conditions. A theory of repeated games that can incorporate, and eventually explain, such differences—in shared understandings as well as in performance—would be a major step forward.

5.4 Limitations and Future Directions

While we believe our experimental results provide interesting and novel insights about a research question that is important for organizational theory and practice, our experimental approach has several limitations that should be addressed by future research. First, our results stem exclusively from laboratory experiments conducted with student participants. The laboratory setting provided great advantages in terms of the measurement—of agreements, degree of equilibration, and performance. However, it would of course be desirable to conduct similar studies with different and more diverse subject pools, as well as in field settings. Especially in field settings, the methodological toolbox likely may need to be diverse to advance this agenda; for example, one might need to combine qualitative approaches that aim at capturing subjects' understandings (before and after shocks) with quantitative data on the subjects' actions and the performance of the relationship.

Moreover, as noted above, although it had significant effects in the short-run, our treatment did not lead to a significant increase in long-run performance. Our evidence suggests that this was due to a failure to increase equilibration in price. Arguably, agreement on price is harder to achieve than agreement on quality in our setting. Future research should examine whether other manipulations can be more effective at increasing all dimensions of efficient equilibration and, hence, performance.

5.5 Conclusion

In this paper we have started to explore whether organizational culture—part of which consists of relational contracts—can enable organizations to coordinate on efficient strategies in the near term and to adapt to unforeseen changes when they occur. Our results suggest that relational contracts

based on general principles can improve long-run performance. However, we also find that it is difficult to trigger the right kind of principles—those that ensure successful adaptation. While in retrospect the latter finding is consistent with the logic of competitive advantage—namely, if it were simple to build such relational contracts why isn't everyone doing it?—our results do leave important open questions. As such, we hope that our paper is the starting point of an empirical and theoretical literature that will explore how parties within and between organizations develop shared understanding to achieve coordination and adaptation in unstable environments.

References

Andreoni, J. & Miller, J. H. (1993). Rational cooperation in the finitely repeated prisoners dilemma: Experimental evidence. *Economic Journal*, *103*(418), 570–85.

Andrews, I. & Barron, D. (2016). The allocation of future business: Dynamic relational contracts with multiple agents. *American Economic Review*, *106*(9), 2742–59.

Baker, G., Gibbons, R., & Murphy, K. J. (1999). Informal authority in organizations. *Journal of Law, Economics, & Organization*, 56–73.

Balliet, D. (2010). Communication and cooperation in social dilemmas: A meta-analytic review. *Journal of Conflict Resolution*, *54*(1), 39–57.

Barney, J. B. (1986). Organizational culture: can it be a source of sustained competitive advantage? *Academy of Management Review*, *11*(3), 656–665.

Blau, P. M. & Scott, R. W. (1962). *Formal Organizations: A Comparative Approach.* Chandler Publishing Company.

Bolton, G. E. & Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American economic review*, *90*(1), 166–193.

Brown, M., Falk, A., & Fehr, E. (2004). Relational contracts and the nature of market interactions. *Econometrica*, *72*(3), 747–780.

Charness, G. & Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, *74*(6), 1579–1601.

Chassang, S. (2010). Building routines: Learning, cooperation, and the dynamics of incomplete relational contracts. *American Economic Review*, *100*(1), 448–65.

Cyert, R. M. & March, J. G. (1963). *A Behavioral Theory of the Firm*. Prentice-Hall.

Dal Bó, P. (2005). Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games. *American Economic Review*, 1591–1604.

Dal Bó, P. & Fréchette, G. R. (2011). The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review*, *101*(1), 411–429.

Ellingsen, T. & Johannesson, M. (2004). Promises, threats and fairness. *The Economic Journal*, *114*(495), 397–420.

Fehr, E., Brown, M., & Zehnder, C. (2009). On reputation: A microfoundation of contract enforcement and price rigidity. *The Economic Journal*, *119*(536), 333–353.

Fehr, E. & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*(3), 817–868.

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

Foss, N. J. (2003). Selective intervention and internal hybrids: Interpreting and learning from the rise and decline of the oticon spaghetti organization. *Organization Science*, *14*(3), 331–349.

Foster, L., Haltiwanger, J., & Krizan, C. J. (2006). Market selection, reallocation, and restructuring in the US retail trade sector in the 1990s. *Review of Economics and Statistics*, *88*(4), 748–758.

Fudenberg, D., Rand, D. G., & Dreber, A. (2012). Slow to anger and fast to forgive: cooperation in an uncertain world. *American Economic Review*, *102*(2), 720–749.

Gibbons, R. & Henderson, R. (2012). Relational contracts and organizational capabilities. *Organization Science*, *23*(5), 1350–1364.

Goeree, J. K. & Holt, C. A. (2001). Ten little treasures of game theory and ten intuitive contradictions. *American Economic Review*, *91*(5), 1402–1422.

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, *91*(3), 481–510.

Guiso, L., Sapienza, P., & Zingales, L. (2006). Does culture affect economic outcomes? *Journal of Economic perspectives*, *20*(2), 23–48.

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, *61*(1), 29–48.

Herz, H., Schmutzler, A., & Volk, A. (2019). Cooperation and mistrust in relational contracts. *Journal of Economic Behavior & Organization*.

Kellogg, K. C. (2009). Operating room: Relational spaces and microinstitutional change in surgery. *American Journal of Sociology*, *115*(3), 657–711.

Kellogg, K. C. (2011). Hot lights and cold steel: Cultural and political toolkits for practice change in surgery. *Organization Science*, *22*(2), 482–502.

Kreps, D. M. (1990). Corporate culture and economic theory. *Perspectives on Positive Political Economy*, *90*, 109–10.

Li, J., Matouschek, N., & Powell, M. (2017). Power dynamics in organizations. *American Economic Journal: Microeconomics*, *9*(1), 217–41.

Malcomson, J. M. (2013). Relational incentive contracts. In R. Gibbons & J. Roberts (Eds.), *Handbook of Organizational Economics* (pp. 1014–1065). Princeton NJ: Princeton University Press.

Proto, E., Rustichini, A., & Sofianos, A. (2019). Intelligence, personality, and gains from cooperation in repeated interactions. *Journal of Political Economy*, *127*(3), 1351–1390.

Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and society*, *7*(1), 58–92.

Schein, E. H. (1985). *Organizational Culture and Leadership: A Dynamic View.* San Francisco: Jossey-Bass Inc., U.S.

Selten, R. & Warglien, M. (2007). The emergence of simple languages in an experimental coordination game. *Proceedings of the National Academy of Sciences*, *104*(18), 7361–7366.

Turco, C. J. (2016). *The conversational firm: Rethinking bureaucracy in the age of social media.* Columbia University Press.

Van Huyck, J. B., Battalio, R. C., & Beil, R. O. (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *The American Economic Review*, *80*(1), 234–248.

Weber, R. A. & Camerer, C. F. (2003). Cultural conflict and merger failure: An experimental approach. *Management science*, *49*(4), 400–415.

Williamson, O. E. (1979). Transaction-cost economics: the governance of contractual relations. *Journal of Law and Economics*, *22*(2), 233–261.

Williamson, O. E. (1991). Comparative economic organization: The analysis of discrete structural alternatives. *Administrative Science Quarterly*, 269–296.

Online Appendix

A Instructions First Trading Phase

Starting on the next page, we reproduce the original experimental instructions for the first trading phase handed out to participants. We provide the instructions given to sellers as an example. Instructions for buyers contained the exact same information and were presented in the same way. The experimental manipulation in the instructions in the Nudge treatment is contained on the final page of these instructions.

General Instructions for Sellers

You are now taking part in an economic study funded by various research funding agencies. Depending on your decisions, you can earn money. It is therefore very important that you read these explanations carefully.

These instructions are for your private information only. **There is an absolute ban on communication during the study**. If you have any questions, please contact us. Non-compliance with this rule leads to exclusion from the study and from all payments.

At the beginning of the study you will receive a show-up fee of 10 Dollars. During the study, we are not talking about Dollars, but about points. In the course of the study, you can earn a further amount of money by earning points. Your total income will therefore be calculated in points first. The total number of points you earn during the study is then converted into Dollars at the end, with

1 Point = 0.04 Dollars.

At the end, you get the points earned during the study plus the 10 Dollars show-up fee in cash.

At the beginning of the study, the participants were randomly divided into two groups: Buyer and seller. **You are a seller throughout the entire study**. Each seller was randomly assigned to a buyer. **The study consists of 2 parts**. The first part consists of 5 rounds. The second part consists of an unknown, randomly determined number of rounds. You will be assigned to the same buyer in all rounds, both in part 1 and in part 2. None of the other study participants will know your exact assignment or your decisions. Anonymity will therefore be preserved.

On the following pages you will first be informed about the exact procedure of the first part of the study. Information on the second part of the study will be provided towards the end of the instructions.

Course of the first part of the study

At the beginning of the study, each seller is randomly assigned to a buyer. Part 1 of the study comprises 5 rounds. You will be assigned to the same buyer in all 5 rounds of the first part, and also in the following second part of the study.

In each round, each seller carries out a transaction with his assigned buyer: The seller determines the quality of a product for which the buyer pays a price. The buyer makes a profit from the transaction if she pays less for the product than the product is worth to her. The value of the product to the buyer depends on the quality of the product. The seller makes a profit on the transaction if he receives a price that exceeds his production costs. Production costs depend on product quality. Higher quality is always associated with higher production costs.

Before the first part of the study starts, you also have the opportunity to **formulate an agreement in the form of a written statement together with the buyer**. How you can draft this statement will be explained later.

Course of a round

You and your assigned buyer complete one transaction in each of the 5 rounds. The process is organized as follows:

Stage 1: You choose the product quality that you want to deliver to your buyer. At the same time, your buyer chooses the price she wants to pay you, without knowing what product quality you deliver.

Stage 2: The buyer receives the product in the quality chosen by you and you receive the price determined by your buyer. Both will be informed about the income in the current round. Then the next round begins.

Stage 1: Determination of quality and price

At the beginning of each round you determine the product quality. The quality must be an integer value between 0 and 10:

$$\mathbf{0 \leq \ Quality \leq 10.}$$

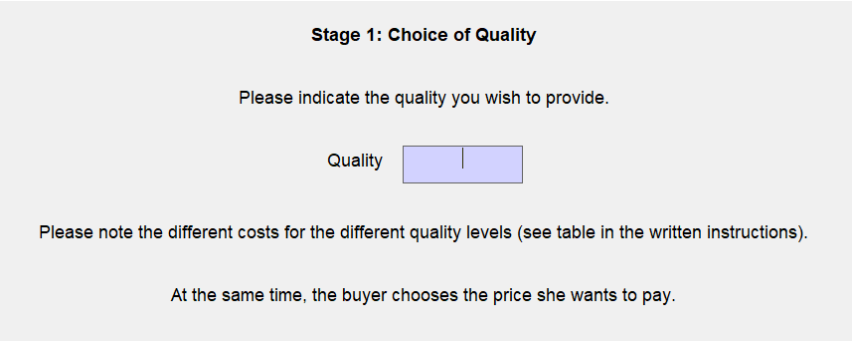
Quality and production cost

The choice of product quality is associated with costs. The higher the quality, the higher the production cost. The production cost for each product quality is given by the table below.

Product Quality	0	1	2	3	4	5	6	7	8	9	10
Cost of Production	0	1	3	6	9	13	18	23	28	33	40

Hence, a product quality of 0 costs nothing. A product quality of 1 costs one point. A product quality of 10 costs 40 points. The values for all other qualities are given in the table above. You choose the product quality and thus the production cost without knowing the price that the buyer will pay.

You can select the quality using the following screen display:

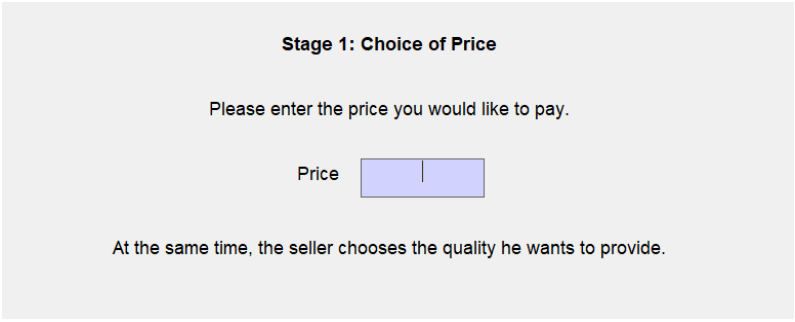


Price

While you choose the product quality, the buyer determines what price she wants to pay, without knowing the quality you will deliver. The price cannot be lower than 0 and not higher than 100:

$$\mathbf{0 \leq Price \leq 100.}$$

The buyer enters the price via the following screen display:



Stage 2: Calculation of incomes

Your income depends on the price the buyer pays and the quality of the product you choose to provide.

Your income in a round is calculated as follows:

Your Income = Price – Production Cost of the Provided Quality

The higher the price chosen by the buyer, the higher your income will be. In addition, the lower the production cost, the higher your income will be. As described above, more quality always costs more. The exact values can be found in the table above.

The buyer's income depends on the price she has paid and the value of the product for her. The value of the product is determined by the product quality you choose to provide as follows:

Value of the Product = 10 * Provided Quality.

With a quality of 0, the value of the product for the buyer is 0. Each additional quality step increases the value of the product by 10 points. For a quality of 1, the value is 10 points, for a quality of 2, the value is 20 points, and so on. For a quality of 10, the value is 100 points.

The income of the buyer is thus as follows:

Buyer's Income = Value of the product – Price
= 10 * Provided Quality – Price

The higher the quality of the product you deliver, the higher the buyer's income, as a higher quality results in a higher value of the product. At the same time, the lower the price the buyer pays for the product, the higher her income will be.

Note that buyers and sellers can also make losses in any period. These must be paid out of the respective show-up fee or income earned in other periods.

Examples

The following (arbitrarily chosen) examples illustrate how incomes are determined.

Example 1:
The seller chooses a quality of 8. The buyer pays a price of 50.
Buyer's income: 8*10-50=30.
Seller's income: 50-28=22.

Example 2:
The seller chooses a quality of 4. The buyer pays a price of 33.
Buyer's income: 4*10-33=7.
Seller's income: 33-9=24.

Example 3:
The seller chooses a quality of 7. The buyer pays a price of 30.
Buyer's income: 7*10-30=40.
Seller's income: 30-23=7.

At the end of each round, you and the buyer can see a summary of the current round's outcomes on the screen. This includes

- The quality you chose to provide.
- The price paid by the buyer.
- Your income.
- The buyer's income.

After that, the current round is completed and the next round begins. In total, the first part of the study consists of 5 rounds. After the fifth round, the second part of the study begins.

The Second Part of the Study

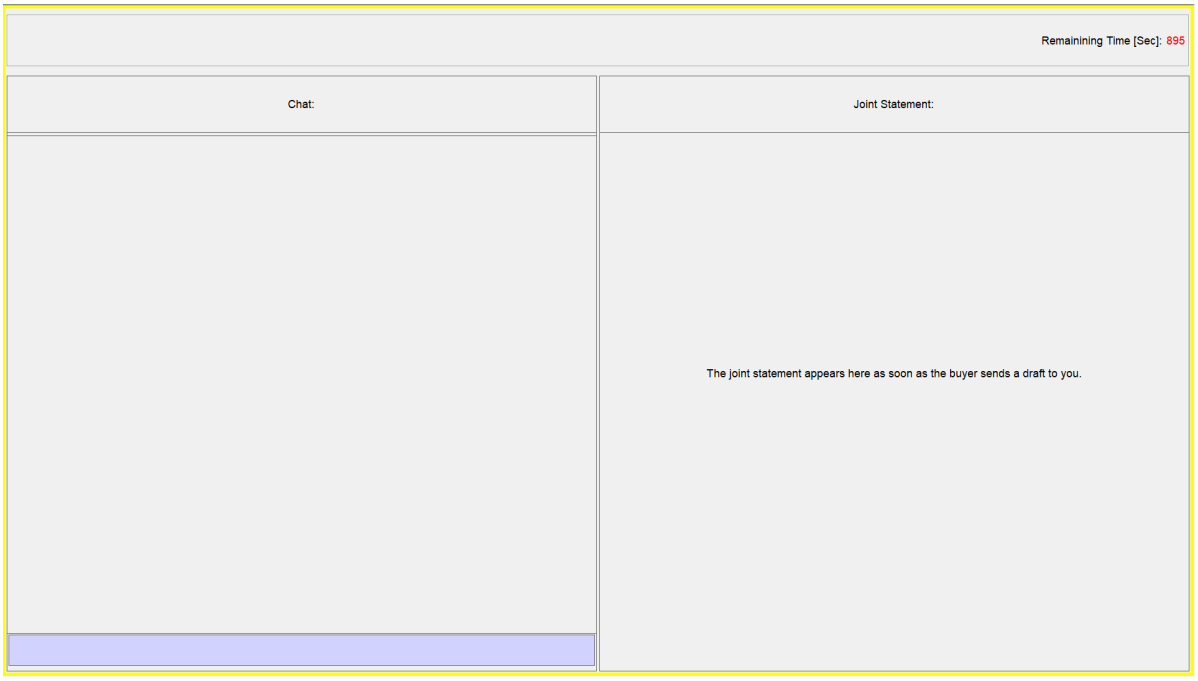
In the second part of the study you are still assigned to the same buyer. You therefore continue to interact with the same person as in the first part. The second part of the study consists of an unknown number of rounds. In the second part of the study, some elements concerning the course of a round and the calculation of income will change. You will only be informed about the concrete changes at the beginning of the second part of the study.

The second part of the study ends randomly. At the end of each round there is a 90% probability that another round will be carried out. With 10% probability, the second part of the study ends and so does the study.

The Agreement

Before the first round of the first part of the study, you have the opportunity to make a non-binding agreement with the buyer assigned to you. It is a written statement, in which you can describe how you and the buyer make their decisions in this study.

You can draft the written statement on the following screen:



The left window is a chat window. Here you can chat with the buyer and discuss your agreement. You have a total of 15 minutes to talk to the buyer and find an agreement. **If you do not come up with a written statement within these 15 minutes, the chat window will automatically close** and the study will continue without a statement.

Limit yourself to discussions concerning this study. Chat messages that have nothing to do with the study or identify a study participant may lead to exclusion from the study.

As soon as you have reached an agreement via chat, **the buyer writes a draft of the statement describing the agreement.** The buyer should write down your agreement as precisely as possible. Once the buyer has written the statement, it will be displayed in the right window. You can then accept or return it if you do not agree with the wording of the statement. In this case, you should again discuss adjustments to the statement via chat. The buyer can then adjust the statement accordingly. This process can be repeated within the 15 minute time limit until a statement of the agreement is drafted that is accepted by you and the buyer.

As soon as both you and the buyer have accepted the statement, or if the 15 minutes have expired, the first round of the first part of the study begins. You can display the statement on the decision screens as needed in each round of the first and second parts of the study.

[*In Nudge Treatment only:*
Please note:
When finding an agreement, you should bear in mind that you do not yet know the exact situation you will encounter in the second part of the study. It may therefore be helpful to consider not only the first part of the study that you already know about, but also the second part that you do not yet have information about. For example, you could think about the principles on which you and the buyer would generally like to act during the study.]

You have arrived at the end of these instructions. Thank you for your attention. **We now ask you to solve the exercises displayed on your screen.** The study can only begin when all participants have correctly completed the exercises. If you have any questions, please raise your hand.

B Instructions Measure of Shared Understanding and Second Trading Phase

Starting on the next page, we reproduce the original experimental instructions for the measure of shared understanding and the second trading phase handed out to participants. We provide the instructions given to sellers as an example. Instructions for buyers contained the exact same information and were presented in the same way. The instructions were identical for the baseline and the Nudge treatment.

Instructions for the Second Part of the Study for Sellers

In the second part of the study you are still assigned to the same buyer. In principle, the process of a round remains the same as in the first part of the study (the seller chooses the quality and the buyer decides on the price at the same time). However, some elements will change compared to the first part.

Before we begin with the second part of the study, we present three possible scenarios of how the situation might change compared to the first part. One of these scenarios is then actually implemented in the second part, i.e. in the second part you will interact with the buyer within one of the described scenarios.

Before the second part begins, and before you know which of the scenarios will actually be implemented, we ask you to answer a few questions for each of the scenarios. The first questions are about **what you and the buyer *should* do in a given scenario as agreed at the beginning of the study:**

1. a) **What quality should you deliver according to the agreement?**
   b) **What price should the buyer pay according to the agreement?**

In your answers, we ask you to **interpret your agreement in a broader sense**. It is possible that you and the buyer did not discuss a particular scenario or decision. In such a case, please consider what the overall meaning and content of the agreement was and base your answers on it.

If your answers to questions 1.a) and b) match those of the buyer, you will receive a bonus: For each question that you answer in the same way as the buyer, you will receive a bonus of 20 points. So think carefully about what decisions you and the buyer should make according to the agreement.

So while questions 1.a) and b) relate to what you and the buyer *should* do according to the agreement, the next questions, 2.a) and b) relate to **what *you will actually do* in the scenario and what *you believe the buyer will actually do.***

2. a) **What quality do you want to deliver in this scenario in the first round of the second part of the study?**
   b) **What price do you think the buyer in this scenario will pay in the first round of the second part of the study?**

Your decision regarding quality in the first round of the second part of the study is binding and will be implemented exactly as you indicated if the corresponding scenario occurs in the second part. Think carefully about your answers.

Your answers to questions 1) and 2) may be the same, or they may differ. For each of the two questions we will also ask you to indicate how confident you are about your answer. Your statement about the certainty of your answers does not affect the bonus or your income in the study in general.

On the following pages we present the three scenarios in detail.

After you have answered the above questions, the second part of the study begins, in which one of the three scenarios is actually implemented. **The second part of the study ends randomly. At the end of each round, there is a 10% probability that the second part of the study will end and the study will be over. With a probability of 90%, the study continues and another round follows.**

As soon as random chance determines that no more rounds will be carried out, the study is over. You will then receive your income from the five rounds of the first part of the study, your income from the bonus questions (1.a) and b)) via the three scenarios, your income from all rounds of the second part of the study, and your show-up fee of 10 Dollars in cash.

In order for the screens for entering your answers to the various scenarios to appear, please enter the following code on your computer: 31-72-93

Scenario 1: Higher Value of the Product and Alternative Source of Supply for the Buyer

Please read the scenario in its entirety (including the examples at the end) before entering your answers on the computer screen!

In this scenario, the value of the product for the buyer increases and doubles compared to the first part:

Value of the Product = 20 * Quality.

In this scenario, the price must not be less than 0 and not more than 200:
$0 \leq$ Price ≤ 200.

In addition, in this scenario the buyer **now has the option of purchasing the product from a different source than you**. The **income that the buyer earns when choosing the alternative source of supply fluctuates randomly**. Two situations can arise:

- Situation 1:
  The alternative source provides a **quality of 0 at a price of 0**, so the **buyer's income is 0 if he chooses the alternative source** (20*0 - 0 = 0). This situation occurs with a **probability of 2/3**.

- Situation 2:
  The alternative source provides a **quality of 10 at a price of 40**, so the **buyer's income is 160 if he chooses the alternative source** (20*10 - 40 = 160). This situation occurs with a **probability of 1/3**.

Important: Which of the two situations occurs is **decided randomly by the computer at the beginning of each period and you and the buyer are both informed before you make your decisions.**

So at the beginning of each period you and the buyer first learn how high the buyer's income is in that period if he chooses the alternative source of supply. Then, as in the first part, you determine the product quality without knowing whether the buyer will purchase the product from you or choose the alternative source. **The production costs for you are the same as in the first part.**

The buyer decides at the same time whether he wants to purchase the product from you or from the alternative source. If he purchases the product from you, he sets a price (which must be between 0 and 200) as in the first part. **If he obtains the product from the alternative source, you will receive a price of 0**, i.e. you will not receive any compensation in this case. **In this case, the buyer receives the income randomly determined by the computer**. All other parameters remain the same as in the first part of the study.

Your income will be calculated exactly the same as in part 1 of the study:

Your income = Price – Production cost of the provided quality

You always incur the cost of the chosen quality, regardless of whether the buyer chooses the alternative source of supply or purchases the product from you. If the buyer chooses the alternative source, the price you receive is 0.

The buyer's income depends on whether she purchases the product from you or chooses the alternative source:

Purchase of the product from you:

Buyer's Income = value of the product – price
= 10 * provided quality – price

Alternative source of supply:

Situation 1 (quality = 0; price = 0): **buyer's income = 20*0 - 0 = 0**
Situation 2 (quality = 10; price = 40): **buyer's income = 20*10 – 40 = 160**

Examples

The following (arbitrarily chosen) examples illustrate how incomes are determined in this scenario.

Example 1:
The computer randomly determines that the buyer's income is 0 (quality of 0, price of 0) if the buyer chooses the alternative source of supply. Buyer and seller are both informed of this. The seller chooses a quality of 5. Without knowing the quality provided by the seller, the buyer decides to purchase the product from the seller and pays a price of 70.
Income of the buyer: 5*20-70 = 30.
Seller's income: 70-13 = 57.

Example 2:
The computer randomly determines that the buyer's income is 160 (quality of 10, price of 40) if the buyer chooses the alternative source of supply. Buyer and seller are both informed of this. The seller chooses a quality of 3. Without knowing the quality provided by the seller, the buyer chooses the alternative source.
Buyer's income: 20*10 - 40 = 160.
Seller's income: 0-6 = -6.

PLEASE ENTER NOW YOUR ANSWERS TO THE QUESTIONS CONCERNING THE ABOVE SCENARIO ON THE COMPUTER SCREEN!

Scenario 2: Fluctuating value of the product

Please read the scenario in its entirety (including the examples at the end) before entering your answers on the computer screen!

In this scenario, the value of the product varies. Two different situations can occur:

- Situation 1:
  The value of the product to the buyer triples compared to the first part:
  Value of the product = 30 * Quality.
  This situation occurs with a **50% probability**.

- Situation 2:
  The product is worthless for the buyer for each quality provided:
  Value of the product = 0 * Quality.
  This situation occurs with a **50% probability**.

Both cases are equally likely in each round, and the computer randomly determines which situation occurs in each round.

As in the first part, you continue to choose the quality you want to provide.
Important: You make this choice without knowing the value of the product for the buyer in this round.

As in the first part, the buyer again determines the price without knowing what quality you are providing.
Important: In contrast to you, the buyer knows whether situation 1 or 2 has occurred in the current round and how high the value of the product is (depending on the quality).

In this scenario, the price must not be lower than 0 and not higher than 300:
$0 \leq \text{Price} \leq 300$.

All other parameters, especially the production costs, remain the same as in the first part of the study.

Your income is calculated exactly the same as in part 1 of the study:

Your income = Price - Production cost of the provided quality

The buyer's income now depends on the realized value situation:

Situation 1:	**Income of the buyer = value of the product - price** **= 30 * provided quality - price**
Situation 2:	**Income of the buyer = value of the product - price** **= 0 - price**

Examples

The following (arbitrarily chosen) examples illustrate the determination of income in this scenario.

Example 1:
The seller chooses a quality of 3. Situation 1 occurs and the value of the product is three times as high as in the first part. The buyer knows the value situation (but not the quality chosen by the seller) and pays a price of 35.
The buyer's income: 3*30-35 = 55.
Seller's income: 35-6 = 29.

Example 2:
The seller chooses a quality of 7. Situation 2 occurs and the value of the product is 0. The buyer knows the value situation that has occurred (but not the quality chosen by the seller) and pays a price of 40.
Income of the buyer: 0-40 = -40.
Seller's income: 40-23 =17.

PLEASE ENTER NOW YOUR ANSWERS TO THE QUESTIONS CONCERNING THE ABOVE SCENARIO ON THE COMPUTER SCREEN!

Scenario 3: Higher value of the product and possible loss of price

Please read the scenario in its entirety (including the examples at the end) before entering your answers on the computer screen!

In this scenario, the value of the product for the buyer increases and doubles compared to the first part:

Value of the product = 20 * Quality.

The price in this scenario must not be less than 0 and not higher than 200:
$0 \leq$ price ≤ 200.

In this scenario, it is also possible that the price paid by the buyer to the seller will not reach the seller.

All other parameters, especially production costs, remain the same as in the first part of the study.

So there are two possible situations in this scenario. One in which the price paid by the buyer does not arrive at the seller and another in which the price paid by the buyer flows into the seller's income (as in the first part).

- Situation 1:
  The price paid by the buyer **does not reach the buyer**. I.e. the buyer pays the price, but you do not receive it. This situation occurs with **50% probability**.

  In this situation, your income is calculated as follows:

  Your income = 0 - production cost of the provided quality

  The buyer's income is calculated similarly to the first part:

  Buyer's income = value of product - price
  = 20 * provided quality - price

- Situation 2:
  The price paid by the buyer **arrives as in the first part and increases your income**. This situation is **50% likely** to occur.

In this situation, your income is calculated as in the first part:

Your income = price - production cost of the provided quality

In this situation, the buyer's income is calculated exactly as before:

Buyer's income = value of the product - price
= 20 * provided quality - price

Both cases are equally likely in each round, and the computer randomly determines which situation occurs in each round.

As in the first part, you continue to choose the quality you want to provide and the buyer determines the price (between 0 and 200) he pays (without knowing the quality you provide).
Important: Before you make your decisions, neither you nor the buyer know whether situation 1 or 2 will occur and whether the price paid by the buyer will reach you.

Examples

The following (arbitrarily chosen) examples illustrate the determination of income in this scenario.

Example 1:
The seller chooses a quality of 5. The buyer pays a price of 60 (without knowing the quality provided by the seller). Situation 1 occurs and the price paid does not reach the seller.
Income of the buyer: 5*20-60 = 40.
Seller's income: 0-13 = -13.

Example 2:
The seller chooses a quality of 8. The buyer pays a price of 110 (without knowing the quality provided by the seller). Situation 2 occurs and the price paid arrives at the seller.
Income of the buyer: 8*20-110 = 50.
Seller's income: 110-28 = 82.

PLEASE ENTER NOW YOUR ANSWERS TO THE QUESTIONS REGARDING THE ABOVE SCENARIO ON THE COMPUTER SCREEN!

C Instructions to Coders

Starting on the next page, we reproduce the instructions provided to coders containing definitions of the different coding categories.

Coding of Statements and Chats

Background: Structure of the Experiment

In the experiment, each participant was randomly assigned a role, buyer or seller, and buyer-seller pairs were randomly formed. These pairs remained together throughout all periods of the experiment. In each period of the experiment, the buyer-seller pairs could complete a transaction, i.e. the buyer could purchase a product from the seller. The seller determined what quality he wanted to deliver (the higher the quality, the higher the production costs for the seller and the higher the value for the buyer) and the buyer determined what price she wanted to pay for the product.

The experiment was divided into two parts. For the first part, the participants received detailed instructions describing in detail all the rules and parameters of the buyer-seller interaction. Please read these instructions carefully, so that you know exactly how the experiment worked and what level of information the participants had. The participants did not receive any detailed information about the second part at the beginning of the experiment. They only knew that they would interact with the same buyer or seller in the second part.

Before the first part started, the participants could chat with each other for 15 minutes to find an agreement on how they would like to act during the experiment (see the relevant sections at the end of the instructions) and to write this down in a joint statement. Your task is to encode these chats and the statements.

Definitions of Coding Categories

Below are the definitions of the different categories according to which you should encode chats and joint statements. Before you start coding, the definition of each of these categories must be clear and familiar. The definitions refer to the possible content of the final agreements and previous chats. So you have to decide if there is any content in an agreement or chat that falls into one of these categories. The categories are not mutually exclusive, i.e. it is possible that an agreement or chat content contains multiple categories.

Fixed rule:
Unique rule that clearly defines numerically a quality and the price to be paid for it.
Examples:

- *The seller delivers a quality of 10 and the buyer pays a price of 70.*
- *Maximum quality and price of 80.*
- *Quality =7, Price = 44*

Flexible rule:

Rule that numerically defines quality and price, but leaves room for manoeuvre by leaving an acceptable range for quality and / or price.

Examples:

- Quality from 6-8, price = 44
- The quality must never fall below 2 and the price must always be higher than 50.

Suboptimal rule:

Agreed rule is not welfare maximizing, i.e. the rule provides for a quality < 10 (does not apply to flexible rules where 10 is a possible, but not the only possible, quality).

Example:

- All cases with quality < 10.

Unequal rule:

Agreed rule leads to unequal distribution of profit between buyer and seller. Attention: with certain qualities, due to the cost structure (and because the price could only be chosen as an integer), it is not possible to divide the profit exactly equally. In such cases, the code should only be used if the difference in profit is >1. In the case of flexible rules where the same distribution is a possible solution, please do not code here.

Examples:

- Price ≠ 70 with a quality of 10
- Price ≠ 61 or 62 with a quality of 9 (NB: this is a case in which an exactly equal distribution is not possible). At a price of 61 or 60, however, the difference in prize is only one point. 62 or 61, with a quality of 9, should therefore not be coded as unequal, but all other prices with a quality of 9 should be).
- Price ≠ 54 with a quality of 8

Principle:

The participants formulate a principle that defines how to act in general. It is not based solely on a numerical definition of the quality to be delivered or the price to be paid (but can be supplemented by such numerical definitions), but provides overarching, general guidelines for action that can be applied in various situations.

Examples:

- We always provide each other with the greatest possible benefit.
- We minimize the risk for the seller.
- We maximize the existing value and divide the profit equally.
- We look at the situation in Part 2 and then adjust quality and price so that it is acceptable to both.
- Mention of Fairness ("FairErw")
- Participants mention fairness considerations in the agreement / chat. This includes avoiding inequality or striving for equality or discussing what is unfair or fair behaviour.

Mention of Fairness:
Participants mention fairness considerations in the agreement / chat. This includes avoiding inequality or striving for equality or discussing what is unfair or fair behaviour.
Examples:

- We try to be fair to each other.
- We want to balance the profits.
- If someone acts unfairly, he will be punished.
- Not paying the agreed price would be unfair.

Mention of efficiency
The participants mention efficiency considerations in the agreement / chat. This is particularly about welfare maximization discussions, i.e. when participants discuss implicitly or explicitly that they want to maximize the sum of payoffs.
Examples:

- We optimize the sum of the profits.
- Optimal quality where marginal costs = marginal benefit

Mention of the second phase
The still unknown second phase of the experiment is mentioned or discussed.
Examples:

- In the second phase we will do xy.
- What happens later in the experiment?

No agreement
"No agreement concluded' (to be coded only for agreements).
Examples:

- You have not made an agreement
- Empty field

Coding Units and Data Structure

As mentioned before, you have to encode both statements and chats. These are located in different folders in the Excel sheet you will receive. Each statement and each chat is uniquely assigned to a group by a "Group_id" number. Please read and encode first the final agreement of a group and then the chat.

For the joint statements there is one line per group and the codes must be entered into the columns in the same line to the right of the statement. Please enter a "1" in the corresponding column if an agreement contains elements that fall into one of the given categories. You do not have to enter a "0" if the corresponding category does not appear in the agreement.

The chats have multiple lines per group. One line corresponds to one chat message. It is about encoding the chat of a group as a whole and not the individual chat messages. I.e. it is not absolutely necessary for us to know in which line (at which position in the chat) a certain coded element occurs, but only whether it occurs somewhere in the chat. Since we consider the chats as a whole, it is also

important that the encodings correspond to the "consensus" that has developed in the chat and that individual statements are not encoded in isolation. For example, if a participant suggests a suboptimal rule, but the two participants realize during the chat that there are better solutions, then "suboptimal rule" should not be coded for this chat. To ensure this, it is best if you read the chat completely before you add the codes and that you enter the codes in the last line of the chat. This of course means that you may have to read a chat (or certain parts of it) several times. Please do this if necessary. It is very important that the encoding is accurate and consistent.