

Building an Equilibrium:

Rules versus Principles in Relational Contracts

Robert S. Gibbons^{*†} Manuel Grieder^{‡§} Holger Herz[¶] Christian Zehnder^{||}

June 8, 2020

Abstract

Effective organizations are able to adapt members' strategies to unforeseen change in an efficient manner. We study when relational contracts enable organizations to achieve this. Specifically, in a novel experiment we explored the hypothesis that basing a relational contract on general principles rather than on specific rules is more successful in achieving efficient adaptation. In our Baseline condition, we indeed observe that, compared to pairs who relied on specific rules, those who articulated general principles achieved significantly higher performance after change occurred. Underlying this correlation, we also find that pairs with principle-based agreements were more likely both to expect and to take actions that were consistent with what their relational contract prescribed. To investigate whether there is a causal link between principle-based agreements and performance, we implemented a "Nudge" intervention intended to foster principle-based relational contracts. The Nudge succeeded in causing more pairs to articulate principles, but the intervention failed to increase performance after the shock because many of the pairs induced to articulate principles then did not take actions that were consistent with their relational contracts. In short, our results suggest that (1) principle-based relational contracts may improve organizational performance, but also that (2) high-performing relational contracts may be difficult to build.

^{*}Massachusetts Institute of Technology, Sloan School of Management and Department of Economics, rgibbons@mit.edu

[†]National Bureau of Economic Research (NBER)

[‡]Zurich University of Applied Sciences (ZHAW), School of Management and Law, manuel.grieder@zhaw.ch

[§]ETH Zurich, Chair of Economics

[¶]University of Fribourg, Department of Economics, holger.herz@unifr.ch

^{||}University of Lausanne, Faculty of Business and Economics (HEC), christian.zehnder@unil.ch

We thank seminar participants at the Stanford CASBS (2016) workshop on organizations and their effectiveness, LMU Munich, Innsbruck, Paris School of Economics, INSEAD, SIOE (2019) in Stockholm, and the Organization Science Special Issue Online Conference 2020. We also thank Lino Eden, Eleonora Egli, and Elise Moret, as well as Stefan Wehrli and the staff at the ETH Decision Science lab for excellent research assistance and administrative support, and we gratefully acknowledge financial support from MIT Sloan's Program on Innovation in Markets and Organizations.

1 Introduction

In this paper we use experimental methods to study parties building “relational contracts.” Following Kreps (1996), we define relational contracts informally as shared understandings of the parties’ roles in and rewards from collaborating together, and formally as equilibria of repeated games.

A large theoretical literature in economics ignores the challenge of creating such shared understandings, assuming instead that parties are in equilibrium from the beginning of their relationship. In contrast to this literature, we study how parties develop relational contracts when there will be shocks to the environment. In particular, we explore the hypothesis that basing a relational contract on general principles rather than on specific rules is more successful in achieving efficient adaptation after change occurs. Our results suggest that (1) compared to rule-based relational contracts, principle-based relational contracts facilitate adaptation and improve organizational performance, but also that (2) high-performing relational contracts are difficult to build.

Our work relies on literatures in multiple social sciences and fields of management concerning relational contracts both within and between organizations. The *within-organizations* view of relational contracts relates to several established strands of organization theory. For example, relational contracts are an important element of transaction-cost economics (such as Williamson, 1979). In addition, we see parallels with Cyert & March’s (1963) quasi-resolution of conflict in the behavioral theory of the firm, in that relational contracts may address “the obvious potential for internal goal conflict inherent in a coalition of diverse individuals and groups” (1963: 31). Finally, we see relational contracts as an aspect of informal organization (e.g., Granovetter, 1985, pp. 490 & 502). As Blau & Scott observed long ago, “It is impossible to understand the nature of a formal organization without investigating the [...] informal relations and unofficial norms” (1962: 6).

Beyond these established strands of organization theory, recent analyses continue to emphasize informal aspects, including Foss’ (2003) discussion of Oticon’s “spaghetti organization” and Turco’s (2016) “conversational firm.” Notably, both Foss and Turco reference the relational-contract model of empowerment by Baker et al. (1999), which envisions empowerment as informal: below the top of an organization, decision rights are “loaned, not owned” (1999: 56).

We see decades of such work conveying a consistent message: relational contracts are (a) common within organizations and (b) often an important determinant of organizational performance.

In addition, detailed analyses of specific organizations—by Foss, Turco, and others such as Kellogg (2009, 2011)—illustrate a further point: (c) relational contracts can be difficult to build.

In economics, some of the earliest discussions of relational contracts—such as Simon (1951) and Williamson (1975, Ch. 4)—considered interactions within organizations. But in sociology and law, some of the earliest discussions—such as Macaulay (1963) and Macneil (1978)—considered interactions between organizations. This *between-organizations* view of relational contracts again relates to established strands of organization theory (where we now construe “organization theory” to concern activities that are organized, even if they do not occur within a single organization). First, between just two entities, relational contracts are a central feature of the supply chains in Dore (1983), the “hybrids” in Williamson (1991), and the alliances in Gulati (1995). Second, among more than two entities, relational contracts are also important in the quasifirms in Eccles (1981), the joint ventures in Kogut (1989), and the networks in Powell (1990). In decades of such work we find the same consistent messages: relational contracts are (a) common between organizations and (b) an important determinant of the organizations’ performances, but (c) difficult to build.

Whether within or between organizations, our interest is in how parties *build* equilibria—not in how they select them from a pre-existing menu, nor in observing that different organizations (or alliances) seem stuck in different equilibria. To illustrate the challenges that building an equilibrium may entail, consider the information requirements underlying relational contracts. Gibbons & Henderson (2012) discuss this “relational knowledge” using Lincoln Electric, Toyota, and Merck as examples. The parties need a shared understanding about which actions (under which circumstances) constitute “cooperation,” which “defection,” and which are available as “punishment” if cooperation fails. That is, in successful relational contracts the parties solve 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 focus of a large theoretical literature in economics described below; the latter is missing from that literature.

Relational contracts require shared understanding not only at the beginning of a relationship but also when circumstances change. As an 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 change can create uncertainty about what it means to comply with an existing relational contract.

Motivated by many such examples, we study how parties build relational contracts when there will be shocks to the environment. We conjecture that parties who develop agreements based on general principles rather than specific rules may perform better after change occurs. For example, a principle and a rule might give identical prescriptions for what should be done in a setting familiar from shared experience, but after the environment changes the principle might still give useful guidance while the rule might prescribe unhelpful behaviors. We use a laboratory setting that allows us to control not only previous history but also the structure and dynamics of the environment. 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. In addition, we can record their communications and relate what has been said to what kinds of equilibria result.

To conclude this motivation for our work, we forge one last connection to organization theory: between relational contracts and organizational culture. More specifically, we connect our work not only to Kreps' (1990) interpretation of repeated-game equilibria as (part of) corporate culture ("how we do things around here"), but also to Schein's (1985) classic definition of organizational culture. In brief, we argue that Schein enriches Kreps, and vice versa.

Schein (1985: 9) 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" We take three inspirations from this definition, as follows.

Our first inspiration concerns "assumptions," which we relate to equilibria in repeated games. As Kreps (1990) emphasized, repeated games have many equilibria, corresponding to different assumptions about how the parties will react to each other's behaviors and to external events. Seeing equilibrium in terms of assumptions makes it natural to ask how such assumptions arise.

Our second inspiration is that assumptions—or, for relational contracts, shared understandings—are "invented, discovered or developed by a given group." Of course, many scholars in many disciplines have studied shared understandings.¹ Our interest is in *the joint development of the shared*

¹Perhaps the pithiest analysis of shared understandings is Geertz's (1973: 12) "Culture is public because meaning

understandings that underlie relational contracts. As we discuss in the literature review, this topic seems unexplored in organizational economics (OE) and almost so in organization theory (OT).

Finally, our third inspiration concerns the organizational issues of “integration” and “adaptation.” Like the general study of shared understandings that pre-dates Schein, the general study of such organizational issues also pre-dates Schein in both OT and OE—e.g., integration and differentiation in Lawrence & Lorsch (1967) and coordination and adaptation in Marschak & Radner (1972), respectively. The specific inspiration we take from Schein is that the details of and balance between such issues may result from path-dependent processes among organization members.

To summarize the connections we see between relational contracts and organizational culture, we follow Kreps in seeing (part of) organizational culture as an equilibrium among the parties involved, and we follow Schein in seeing such equilibria as developed over time “by a given group as it learns to cope.” Within organizations, we see such relational contracts addressing key issues such as differentiation, coordination, adaptation, and delegation; and between organizations we see cultures in supply chains, alliances, and the like, just as within organizations.

1.1 Overview of Experiment

We designed a laboratory experiment allowing us to study how parties build relational contracts and how they adapt when the environment changes. In particular, we implemented a repeated buyer-seller game with a random stopping rule, where participants knew the initial structure and parameters of the game, but were also aware that the game would change in an unknown way after a few periods. Before trading interactions began, we gave participants an opportunity to communicate about a non-binding agreement on how to play the entire game and to formulate a written statement of this initial agreement.

In addition to observed choices in games actually played, we also measured the extent to which pairs reached a shared understanding about appropriate behavior in possible post-shock environments. To collect these data, we interrupted interactions before the game changed. We confronted participants with different scenarios of how the game *might* change for the coming periods and asked them to indicate for each scenario (a) what choices their agreement prescribes and (b) what choices they expect the parties to actually implement. Subsequently, we implemented one of the

is.” See Gibbons & Prusak (2020) for an interpretation of Geertz in keeping with the present paper.

scenarios as the game that participants played for the remaining periods.

Our first set of results explores the endogenously emerging non-binding agreements in this Baseline setting. More specifically, we investigate how different types of initial agreements correlate with subsequent performance (defined as the sum of profits in a buyer-seller pair). We find that pairs whose initial agreement articulated a broad principle tended to achieve higher levels of performance when their world changed than did pairs whose agreement articulated only a narrow rule. Further analysis reveals that the mechanism underlying this correlation is that pairs with principle-based agreements were more likely to expect their pair to take the actions that were prescribed by their agreement; that is, *pairs with principle-based agreements were more likely than pairs with rule-based agreements to be in equilibrium after the shock.*

We then investigated whether there is a causal link between principle-based agreements and performance. To this end, we implemented a “Nudge Treatment” designed to exogenously foster the emergence of principle-based agreements. Our Nudge succeeded in inducing more pairs to articulate principles in their agreements. In addition, the Nudge also induced more pairs to select efficient choices in the scenarios with which we confronted them.

Importantly, however, the Nudge failed to increase long-run performance in the game participants played after the change. A reason for this result on long-run performance is that, although the Nudge effectively coordinated pairs on efficient initial *quality*, it did not lead to more coordination on the initial *price*. The Nudge was thus not successful enough in helping to avoid conflicts about the price that led to a deterioration of performance over time. As a result, the Nudge did yield more principle-based agreements, but it appears that those pairs that agreed on principles only because of the Nudge were no more likely to be equilibrated than if they had not been treated. Providing a potential explanation for this observation, we find that the agreements of these additional pairs with principles in the Nudge were less clear and less often had a shared interpretation. In short, we conjecture that many of the additional principles in the Nudge were of low caliber.

We see these findings from both the Baseline and the Nudge as consistent with the view that building relational contracts is difficult—certainly in the world and even in the lab. Put differently, in the Baseline some pairs were able to reach principle-based agreements that induce efficient adaptation when the world changes, but many pairs were not. The goal of the Nudge was to induce the latter kind of pairs to behave like the former. Apparently our simple Nudge was too weak

to achieve this result—even though our lab environment is enormously simpler than a real-world setting. More specifically, the Nudge induced more pairs to state that they had reached agreement on a principle, but in fact many of these agreements were of low caliber, and not equilibria: as the game was played, at least one party departed from what the other party thought had been agreed.

1.2 Related Literature

Having provided broad motivations from and connections to organization theory above, we now turn to specific papers related to ours.

In organizational economics, there is a large theoretical literature on relational contracting (see Malcomson (2013) for a survey), but such models assume that the parties start in equilibrium. That is, these models ignore the challenges that real parties face when creating a shared understanding 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 adaptation, it is exactly in such settings where real parties may have most difficulty creating shared understandings.

Organization theory has come closer than organizational economics has to providing evidence on learning and adaptation in relational contracts—especially relational contracts between organizations. For example, Gulati (1995) considers the role of prior alliances between the same partners in shaping the current alliance. Similarly, Poppo & Zenger (2002) explore the impact of whether “the buyer has worked with the vendor for years and years” (2002: 717). While suggestive, estimating the effect of the number or duration of prior interactions on the current interaction explores the *result* of the relationship’s early dynamics, not those dynamics themselves.

Studying parties *building* an equilibrium requires longitudinal data on fixed parties. In an innovative paper, Mayer & Argyres (2004) analyze a sequence of formal contracts between two firms. While their data focus on the dynamics of the formal contracts over time (which appear to result from the firms learning to work together), Mayer & Argyres also speculate about how contracts might affect the development of interorganizational trust—a topic close to relational contracting but again not providing direct evidence on the process of building relational contracts.

Finally, Lumineau & Malhotra (2011) and Malhotra & Lumineau (2011) use legal files to analyze fine-grained data on parties’ behaviors during and after contract disputes as a function of the formal

contract structure chosen for the relationship. Our work complements theirs by analyzing (a) the heterogeneous behaviors that may occur holding the formal structure constant and (b) how informal interventions besides the formal structure (such as our Nudge) may help parties avoid disputes in the first place—a possibility that cannot be analyzed in data concerning disputes that have already escalated to involving lawyers.

As for building relational contracts *within* organizations, the closest contributions we know are ethnographies like those by Kellogg and Turco cited above. For all their richness, ethnographies offer limited ability to measure parties' beliefs about appropriate behavior in hypothetical futures. Indeed, even in lab studies that provide longitudinal data on fixed parties, measuring outcomes over time again explores the result of the relationship's early dynamics, not those dynamics themselves.

Methodologically, our approach follows the traditions of experimental economics. More specifically, our study is related to laboratory experiments on repeated social dilemma games that study cooperation strategies (see, e.g., Dal Bó, 2005; Dal Bó & Fréchette, 2011; Fudenberg et al., 2012 for infinitely repeated games like in our experiment, and, e.g., Andreoni & Miller, 1993; Brown et al., 2004; Fehr et al., 2009; Camerer & Linardi, 2019; Herz et al., 2019 for games with finite repetition). Our experiment advances this literature by studying how parties prepare for unforeseen contingencies in an uncertain environment, a difficult and relevant challenge in many organizational settings.²

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; Skorbiansky, 2018). Our paper goes beyond the fact that communicating helps (relative to not communicating) by studying what kind of non-binding agreements are useful to achieve adaptation in an unstable environment. Moreover, in our setting, different groups reach different agreements in advance and then have different interpretations of appropriate adaptation 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).

²Fudenberg 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.

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 how parties build relational contracts and how they adapt after shocks, we created a 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 building equilibrium, before trading interactions started we gave participants the possibility to communicate via text chat in order to reach a non-binding agreement on how to play the game and to formulate a written statement capturing this agreement. After the first phase of trading 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 initial agreements endogenously emerged and how these agreements were associated with groups' shared understanding and performance after change occurred.

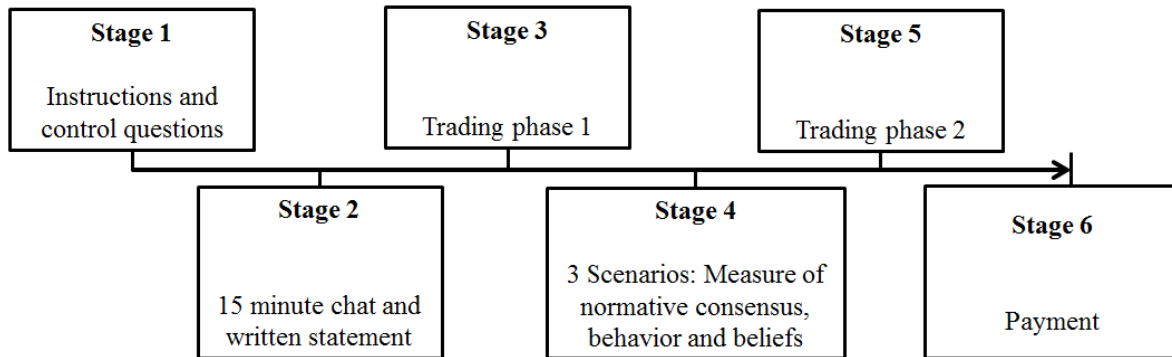
To assess whether the nature of a group's initial agreement has a causal impact on their subsequent performance, we implemented an experimental treatment that fostered the emergence of principles in the agreement-finding phase. In subsection 2.1, we describe the experimental set-up in our Baseline condition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

2.1 Baseline Experimental Set-up

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

Stage 1: Instructions and Control Questions. The instructions informed all participants of

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



their role and of the game parameters in the first (five-period) trading phase (the full instructions are reproduced in Appendix ??). 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. Thus, whereas participants were aware that something would change, they could not reasonably foresee what form that change would take exactly. 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 ensure understanding of the experimental set-up before they could enter the next stage.

Stage 2: Chat 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 a non-binding 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.³

³Note that an agreement might exist without the pair having been able to write down a statement (for instance, because they ran out of time in the communication phase). For our analyses, we therefore coded agreements not only in the statements but also in the chats (see Section 2.4).

Stage 3: First Trading Phase. All groups played five rounds of a simultaneous buyer-seller game in which the seller decided about the quality q of a good to be delivered and the buyer determined the price p .⁴ 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 $c(q)$.⁵ 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.⁶

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 a pair's initial agreement 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 ??.

For each scenario, we asked participants a set of questions aimed at measuring the *normative* and the *behavioral* implications of the pair's initial 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 experimental points for consensus with their partner (for each question in each scenario). For the questions

⁴In 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.

⁵Note 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.

⁶Note 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 ($v' = 10$).

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 $q_{hi}^o = 10$ at a price of $p_{hi}^o = 40$). 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.⁷ 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 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.

⁷For 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, corresponding to the profits made in the experiment, 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 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 and were randomly assigned the role of buyer or seller. 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, but they knew that their interaction partner was another participant present in the room during the same experimental session. 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 and 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 ?? 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**

	<i>Statements</i>	<i>Chats</i>
Principle	.91 (.82)	.80 (.59)
Rule	.98 (.96)	.91 (.89)
<i>N</i>	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 how parties build relational contracts that allow them to adapt to unforeseen change in an efficient manner. To this end, we study a repeatedly interacting buyer-seller pair that experiences an exogenous change in their trading environment. The pair starts their relationship in a setting with known parameters and they are aware that there will be a need for adaptation, but the nature of the upcoming change is unknown to them.

Our buyer-seller pairs faced a game in which efficient cooperation could be sustained as an equilibrium of the infinitely repeated game.⁸ From a standard game-theoretic perspective one would therefore expect that the parties not only cooperate from the outset but also immediately adapt their strategies to keep cooperating after the shock.⁹ In contrast, we believe that such

⁸We presume that the players (correctly) believe that the continuation payoffs will not diminish too much after the change, so that cooperation can also be sustained in equilibrium before the change.

⁹Of course there are also inefficient equilibria, but efficient equilibria have received almost all the attention in the literature (see, e.g., Malcomson, 2013).

a move to a new form of cooperation may not be trivial. For example, Gibbons & Henderson (2012) describe real-life adaptations complicated by misunderstandings and coordination failure. They argue that successful adaptation requires not only that collaborative equilibria exist (the “credibility problem”), but also that the partners succeed in building a shared understanding of the equilibrium strategies (the “clarity problem”).

Because the post-shock phase of our experiment is a repeated game (not a one-shot interaction), we believe that successful adaptation requires that the parties align their beliefs and actions about both the quality to be provided and the price to be paid after the shock.¹⁰ We use the term *equilibrated* to describe a buyer-seller pair that coordinates on a new price-quality combination that simultaneously meets the expectations of both parties. Specifically, a pair is equilibrated if the buyer’s chosen price (p) is equal or larger than the seller’s price expectation (\hat{p}_S) and the seller’s chosen quality (q) is equal or larger than the buyer’s quality expectation (\hat{q}_B). But it is possible to equilibrate on an inefficient outcome. If a pair is not only equilibrated but also coordinates on actions that maximize the total surplus (the sum of profits in the pair), we use the term *efficiently equilibrated*. In our experiment efficiency is entirely determined by the seller’s quality choice. Thus, an equilibrated pair is efficiently equilibrated if the seller chooses a quality of 10.

Our first goal therefore is to understand the process that produces equilibration after change occurs. Put differently, we are interested in how relational contracts that facilitate adaptation are built. All the game-theoretic models of cooperation that we know are silent about this issue.¹¹ In these models equilibration is taken to be so simple that it does not appear in the model.

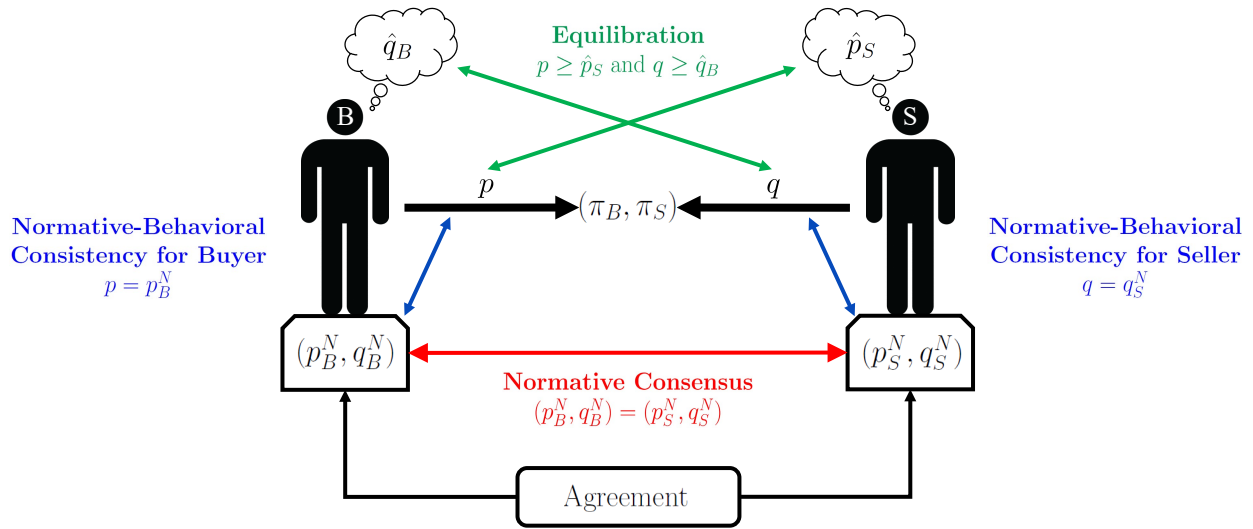
We hypothesize that two elements are important for a pair to reach equilibration after the environment has changed. First, it seems valuable that partners have a shared understanding of *what should be done* in the new situation. We say that a pair exhibits *normative consensus* if the two parties agree on the price-quality combination $((p_B^N, q_B^N) = (p_S^N, q_S^N))$ that their initial agreement implies in the new situation. If, in addition, they also both agree that the seller should choose the efficient quality of 10, we say that the pair reached *efficient normative consensus*.

¹⁰That is, in a one-shot interaction, it *might* be possible for parties to stumble into high performance without having their beliefs and actions aligned, but abundant theoretical, experimental, and real-life evidence about repeated games strongly suggests that collaboration withers once parties suffer what they deem to be reneging. In other words, failure to reach what we call “equilibration” causes conflicts between trading parties that harm efficiency.

¹¹To be clear, we mean applied models of collaboration, not abstract models of fictitious play, rationalizability, evolutionary games, or the like.

Second, an (efficient) normative consensus helps to produce (efficient) equilibration only if both parties actually *do what they think should be done*. We say that a given party exhibits *normative-behavioral consistency* if that party acts in accordance with what that party thinks is implied by the initial agreement in the new situation. More specifically, the buyer exhibits normative-behavioral consistency if she chooses the price that she thinks the agreement indicates ($p = p_B^N$), and the seller exhibits normative-behavioral consistency if he chooses the quality that he thinks the agreement indicates ($q = q_S^N$). See Figure 2 for a summary of this terminology.

Figure 2: **Framework and Terminology**



Notes: The figure illustrates our framework and terminology. The payoffs to the buyer (π_b) and the seller (π_s) are determined by the simultaneous choices of price (p) by the buyer and quality (q) by the seller. If the buyer and seller agree on *what should be done* in a given state, the pair has reached *normative consensus*: the price-quality combination that the buyer perceives as normatively appropriate (p_B^N, q_B^N) corresponds with what the seller perceives as normatively appropriate (p_S^N, q_S^N). For a given party, *normative-behavioral consistency* requires that the party act in accordance with what that party thinks is implied by the pair's informal agreement. For the buyer, normative-behavioral consistency means choosing the price ($p = p_B^N$) that she thinks should be chosen according to the agreement, and for the seller, normative-behavioral consistency means choosing the quality ($q = q_S^N$) that he thinks should be chosen. Finally, a buyer-seller pair is *equilibrated* if the buyer's price choice (p) and the seller's expectation of that price (\hat{p}_S) satisfy $p \geq \hat{p}_S$ and the seller's quality choice (q) and the buyer's expectation of that quality (\hat{q}_B) satisfy $q \geq \hat{q}_B$. An equilibrated pair is *efficiently equilibrated* if they coordinate on the efficient quality of $q = 10$.

We note that, strictly speaking, efficient normative consensus and normative-behavioral consistency are neither sufficient nor necessary conditions for pairs to be equilibrated.¹² However, we

¹²That is, pairs *could* be equilibrated (i.e., the actual price paid by the buyer meets or exceeds the seller's expectation and the quality delivered by the seller meets or exceed the buyer's expectation) without agreeing on the normative implications of their initial agreement and/or without acting in accordance with these normative implica-

hypothesize that in practice most pairs that fail to achieve normative consensus or that fail to have both parties display normative-behavioral consistency will not reach equilibration. Our second goal therefore is to explore empirically whether normative consensus and normative-behavioral consistency are important building blocks for equilibration.

Finally, given our conjecture that agreements involving general principles rather than specific rules may be useful even after circumstances change, our third goal is to understand how principle-based agreements help pairs (a) be efficiently equilibrated after a change and (b) reach efficient normative consensus and exhibit normative-behavioral consistency.

3.2 Testable Hypotheses

Our first hypothesis describes our predictions regarding the expected performance of different types of initial 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 for the pair (i.e., agreement on what should be done) and normative-behavioral consistency for each individual party (i.e., individual willingness to do it). As a consequence, we expect that those pairs who establish principle-based agreements are more likely to be efficiently equilibrated and will therefore 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 exhibit both normative consensus as a pair and normative-behavioral consistency as individuals, and such pairs are more likely to be efficiently equilibrated after an exogenous change in the environment.*

This 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 tions. Likewise, pairs can have reached normative consensus and act consistently with it, but one party *could*—for some reason—still expect the other party to take an action that exceeds the normatively appropriate choice (i.e., to choose a quality or a price that is higher than what the normative consensus implies).

claims regarding the effects of principle-based agreements on outcomes. Our Nudge Treatment therefore aims at exogenously triggering the emergence of 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) **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.*
- b) **Mechanisms:** *Buyer-seller pairs in the Nudge Treatment are more likely to exhibit both normative consensus as a pair and normative-behavioral consistency for both individuals, and they are 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 of the effects of endogenously emerged principles in the Baseline condition (Hypothesis 1). In the second part we analyze the results of the Nudge Treatment (Hypothesis 2).

4.1 Baseline Outcomes

In this subsection, we first investigate the frequencies of rule-based and principle-based initial agreements and analyze how the different types of initial agreements correlate with performance in Phase 2 (i.e., after the trading environment has changed).

We then explore the mechanisms underlying the observed correlation between initial agreement and performance after the shock, in two steps. First, we consider the association between equilibration and performance (and the association between principle-based agreements and equilibration). Second, we consider the association between efficient normative consensus and normative-behavioral consistency, on the one hand, and equilibration, on the other (as well as the association between principle-based agreements and the consensus and consistency concepts).

As there was no exogenous variation in the Baseline, agreements were established endogenously and 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 initial agreements emerged and how agreement type correlates with the post-shock performance:

Result 1

- a) *A large majority of pairs (70%) established rule-based rather than principle-based initial agreements.*
- b) *Pairs who formulated principle-based agreements tended to achieve significantly higher post-shock performance than those who relied 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 pairs) with which principle-based and rule-based agreements endogenously emerged in the Baseline condition—separately for the statement, the chat, and combining both of them.

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

	<i>Statements</i>	<i>Chats</i>	<i>Combined</i>
Principle	.133	.250	.300
Rule	.883	.950	.983
<i>N</i>		60	

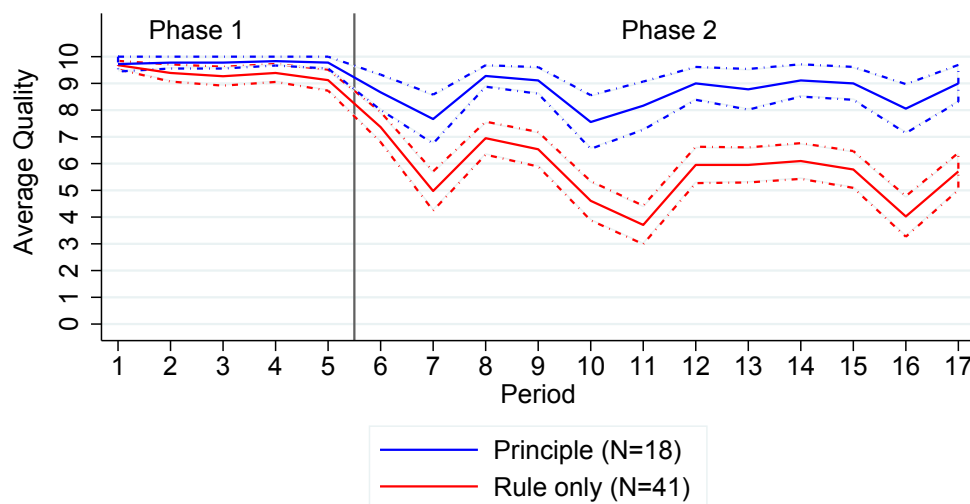
Notes: The table shows the coded frequency of pairs (in fraction of all pairs) who included principles or rules in their statements in the baseline. The first column contains only codings for the final written statements. The second column contains only codings for chat messages. The third column combines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

We consider it likely that both the statements and the chats could have an impact on the initial 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 pair relied on a principle whenever a principle was agreed upon in the statement, the chat, or in both. (See the Online Appendix¹³ for analyses of principles that were coded as present in the chats only versus in the statements.) When considering the combined data, we observe that (with the exception of one pair) all pairs wrote down a rule, whereas only 30% of all pairs also articulated a principle. Interestingly, there

¹³https://osf.io/ne7rd/?view_only=ee19b21676cc4366aed8c0729dae141b

are no pairs who governed their interaction with a principle alone—i.e., all pairs who formulated a principle also specified its implication for the current setting in Phase 1 (which was known at the time that the parties could communicate), in the form of a rule.

Figure 3: **Quality in Pairs with Rules vs. Principles (Baseline Only)**



Notes: Only data from the Baseline included. One of the 60 pairs in the Baseline had neither a principle nor a rule and is thus not included in the graph ($n = 59$). The solid lines display average quality in a given period. The dashed lines represent plus/minus one standard error of the mean.

Figure 3 illustrates the second part of Result 1. The Figure shows the average quality provided by sellers over time and distinguishes between pairs with an initial agreement that contained only a rule versus pairs who also formulated a principle. Note that in our setting the seller's quality choice fully determines overall performance of the relationship (i.e., joint profits). The figure reveals that pairs with principle-based agreements clearly outperformed pairs that relied solely on rules. Indeed, and in line with Hypothesis 1a, when using joint payoffs in Phase 2 as an overall measure of performance, we find that pairs with principles had average earnings of 1,752 points, whereas pairs with rules earned 1,348 points on average. This difference is statistically significant ($t = 3.95$, $p < .01$).¹⁴ In contrast, when comparing average earnings in Phase 1, we find no significant difference ($t = 1.21$, $p = .23$).

¹⁴All p -values reported in this paper are for two-tailed significance tests and all t -tests were conducted allowing for unequal variances between compared groups.

4.1.2 The Role of Efficient Equilibration for Long-Run Performance

We next explore the mechanisms underlying the observed correlation between initial agreement and performance after the shock. We begin by considering the association between efficient equilibration and performance (and the association between principle-based agreements and equilibration).

In Section 3, we hypothesized that a fruitful buyer-seller collaboration after the shock is facilitated if beliefs and actions about chosen quality and price are equilibrated on an efficient outcome.¹⁵ To measure the degree of efficient equilibration *before the shock*, we construct an *Efficient Equilibration Score* ranging from 0 to 3, capturing in how many of the three scenarios presented to participants before Phase 2 a pair was efficiently equilibrated. Recall that decisions in these scenarios were incentivized by making the answer the first-period choice in Phase 2 if that scenario was implemented. We then explore whether efficient equilibration in the scenarios is correlated with high performance throughout Phase 2 (i.e., during the full repeated game after the shock) and whether pairs with principle-based initial agreements were more likely to achieve efficient equilibration in the scenarios. The findings are summarized in our second result:

Result 2

- a) *Pairs that were efficiently equilibrated in the scenarios reached significantly higher performance levels in the subsequent trading phase 2.*
- b) *Pairs with principle-based initial agreements were significantly more likely to be efficiently equilibrated in the scenarios.*

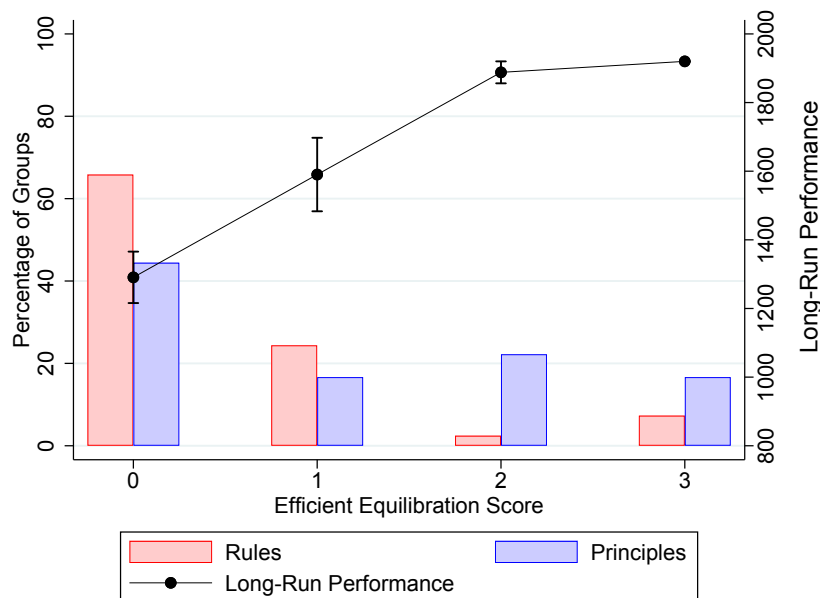
Figure 4 shows how the data support Result 2a. As indicated by the black line, there is a clear correlation between post-shock performance, measured in terms of aggregate profits throughout Phase 2 (right-hand axis), and the Efficient Equilibration Score in the scenarios ($r = 0.53$, $p < .01$).

Of the three scenarios discussed before Phase 2 began, one was implemented as the trading game in Phase 2, so a more conservative approach is to study the correlation between long-run performance and the Efficient Equilibration Score based on only the two non-implemented scenarios.

¹⁵Section 3 notes that it is possible for pairs to be equilibrated on inefficient outcomes. However, the data reveal that in our setting this phenomenon plays no important role: in 93.3% of the cases in which a pair reached equilibration, the pair is efficiently equilibrated. We therefore focus on efficient equilibration. Qualitatively, all our results hold when replicating the analysis for the less restrictive equilibration concept.

The result is essentially unchanged. The correlation then is $r = 0.46$ ($p < .01$) between long-run performance and efficient equilibration.

Figure 4: **Efficient Equilibration, Long-Run Performance and Agreement Type**



Notes: Only data from the Baseline included. The line shows average profits in Phase 2 (long-run performance) for a given Efficient Equilibration Score of a pair ($n = 60$). The error bars represent plus/minus one standard error of the mean. The bars represent the share of pairs who reached a particular Efficient Equilibration Score, separately for pairs with principle-based agreements and pairs with rule-based agreements ($n = 59$, because one pair agreed neither on a rule nor on a principle).

Figure 4 also provides support for Result 2b. The bars show the proportion of pairs that achieved the different possible levels of efficient equilibration, conditional on having a rule-based agreement versus a principle-based agreement. In line with Hypothesis 1b, we observe that pairs with principle-based agreements were substantially more likely to achieve high levels of efficient equilibration than pairs who relied on rules only. The average Efficient Equilibration Score is more than twice as large for pairs with principle-based agreements (1.11 vs. 0.51, $t = 1.93$, $p = 0.06$).

4.1.3 Building Equilibration: Efficient Normative Consensus and Normative-Behavioral Consistency

Result 1 established that pairs with principle-based initial agreements had better post-shock performance than did pairs relying on rules alone, and Result 2 established that post-shock performance is correlated with efficient equilibration in the scenarios. As discussed in Section 3, two concepts

seem important in achieving such equilibration. First, trading partners need to have a shared understanding of *what should be done* to reach an efficient outcome in a particular situation (*efficient normative consensus*). Second, normative consensus is useful only if both parties actually *do what should be done* according to their agreement (*normative-behavioral consistency*).

We now analyze the association between equilibration, on the one hand, and efficient normative consensus and normative-behavioral consistency, on the other (and the association between principle-based agreements and these consensus and consistency concepts). Note that all the measures in this analysis are drawn from the parties' responses to the scenarios, before the shock occurs.

The next result confirms that there are important correlations between efficient normative consensus within the pair and normative-behavioral consistency by both parties, on the one hand, and efficient equilibration, on the other. The result also shows that the effect of principle-based agreements works more through normative-behavioral consistency (doing what should be done) than through efficient normative consensus (agreeing on what should be done).

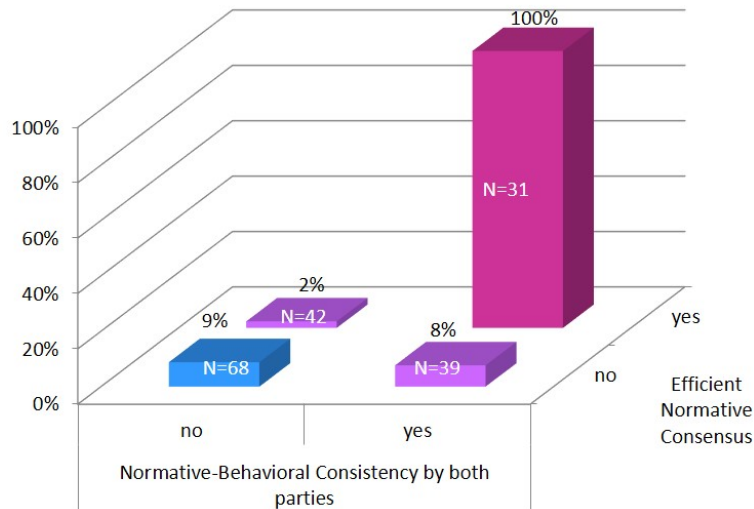
Result 3

- a) *Efficient equilibration is unlikely without the joint presence of efficient normative consensus and normative-behavioral consistency by both parties.*
- b) *Pairs with principle-based agreements are not significantly more likely to develop efficient normative consensus, but they do display significantly stronger normative-behavioral consistency (for both parties).*

Recall from Section 3 that, strictly speaking, efficient normative consensus and normative-behavioral consistency are neither sufficient nor necessary conditions for pairs to be efficiently equilibrated. Figure 5 illustrates that, in our data, the likelihood of efficient equilibration is extremely high (actually, in these data, certain) if efficient normative consensus and normative-behavioral consistency are both present and extremely low otherwise. This observation is the foundation for Result 3a.

Evidence for Result 3b stems from Figure 6. For an average pair, the figure shows in how many of the three scenarios presented to participants before the shock: (a) the pair was efficiently equilibrated; (b) the pair reached efficient normative consensus; and (c) both parties in the pair displayed

Figure 5: **Efficient Normative Consensus, Normative-Behavioral Consistency, and Efficient Equilibration**

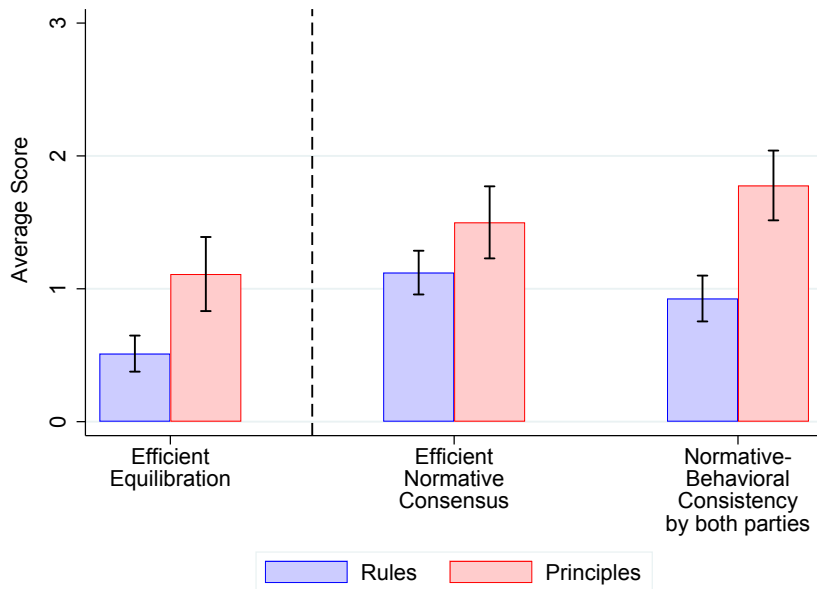


Notes: Only data from the Baseline included ($n = 180$ from 60 pairs in 3 scenarios). The horizontal axes categorize pairs based on whether both parties behaved normative-behaviorally consistent and whether the pair had efficient normative consensus at the scenario level. The vertical axis plots the percentage of pairs who reached efficient equilibration in the corresponding scenario.

normative-behavioral consistency. The figure shows these averages for pairs with principle- versus rule-based initial agreements.

As a benchmark, the data to the left of the dashed line repeats the finding discussed in subsection 4.1 that pairs with principle-based agreements are more likely to be efficiently equilibrated (see also Result 2b and Figure 4). New data is presented to the right of the dashed line. In the middle of the figure, pairs with principle-based agreements reached, on average, a 34% higher score for efficient normative consensus, but this effect is not statistically significant ($t = 1.19$, $p = 0.24$). On the right side of the figure, however, such pairs were 92% more likely to have both parties display normative-behavioral consistency, a highly significant effect ($t = 2.71$, $p = 0.01$). This higher degree of normative-behavioral consistency (i.e., actually doing what a party thinks should be done) by both parties in pairs with principle-based agreements is a key correlate of high performance in our Baseline data.

Figure 6: **Effects of Principles on Efficient Equilibration, Efficient Normative Consensus and Normative-Behavioral Consistency**



Notes: Only data from the Baseline included. The bars show the average Efficient Equilibration Score, Efficient Normative Consensus Score, and Score for Normative-Behavioral Consistency by both parties separately for pairs with principle-based agreements and pairs with rule-based agreements ($n = 59$, because one pair neither agreed on a rule nor on a principle). The error bars represent plus/minus one standard error of the mean.

4.2 Experimental Manipulation: Nudging the Emergence of Principles

So far we have shown that pairs who endogenously reached principle-based initial agreements had higher long-run performance after the shock, were more likely to be efficiently equilibrated in the scenarios considered before the shock, and were more likely to have both parties display normative-behavioral consistency in these scenarios. However, these results are purely correlational.

We now analyze whether our Nudge Treatment can provide causal evidence about any of these effects. Recall that the Nudge manipulation consisted of only a short reminder in the experimental instructions, making more salient the fact that the game was going to change in the second trading phase 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.

Subsection 4.2.1 parallels the analyses from subsection 4.1. While we find that the Nudge significantly increased the emergence of principle-based agreements, we then find a treatment effect

only on normative-behavioral consistency; we find no statistically significant increase in long-run performance, efficient equilibration, or efficient normative consensus.

To understand these results, in subsection 4.2.2 we show that the Nudge did create a significant increase in efficient equilibration on quality, which translated into a significant increase in short-run performance. But we also find that the Nudge failed to create a significant increase in equilibration on price, and lack of equilibration on price was in turn associated with lower long-run performance.

Finally, in subsection 4.2.3, we explore whether the absence of treatment effects on price equilibration and on long-run performance can potentially be explained by differences in the caliber of the principles that were exogenously induced by the Nudge relative to those that endogenously emerged in the Baseline.

4.2.1 Treatment Effects on Long-Run Performance, Efficient Equilibration, Efficient Normative Consensus, and Normative-Behavioral Consistency

Before discussing the impact of the Nudge on the outcomes of interest, we first report a manipulation check to establish that the treatment had the intended effect on observed agreements: inducing more pairs to establish principle-based initial agreements. We find that this is the case. Coders, who were blind to treatment, found far more principles in the Nudge (in 52 of 61 or 85.2% of pairs) than in the Baseline (in 18 of 60 or 30.0% of pairs). This difference is highly significant ($z = 6.15$, $p < .01$). As in the Baseline, all pairs who agreed on a principle accompanied it with a specific rule for Phase 1. In addition, there were 7 pairs in the Nudge Treatment that had a rule only and 2 pairs that were coded as having neither a rule nor a principle.

Next we present the impact of the Nudge on the outcomes of interest. Our fourth result summarizes the observed impact of the Nudge on long-run performance, efficient equilibration, efficient normative consensus, and normative-behavioral consistency:

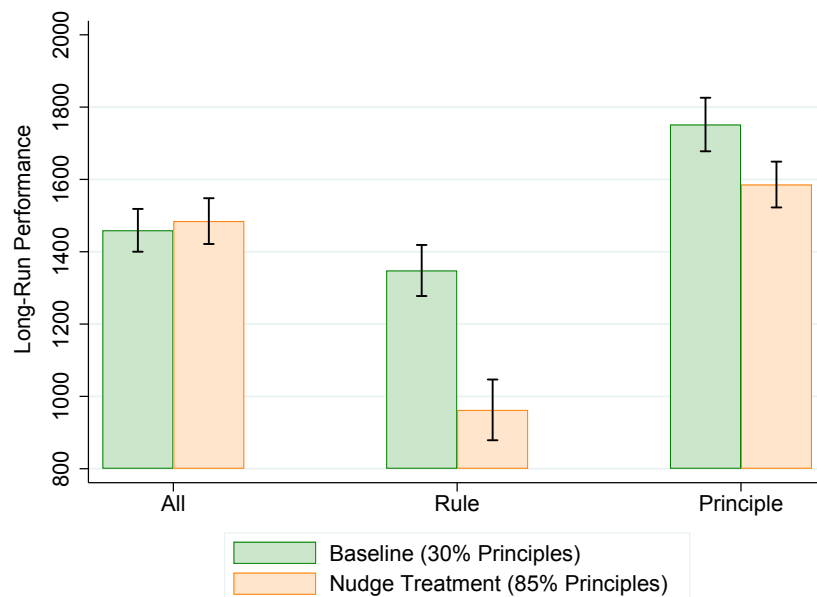
Result 4

- a) The Nudge did not significantly increase long-run performance.*
- b) The Nudge did not significantly increase efficient equilibration.*
- c) The Nudge did not significantly increase efficient normative consensus, but it did significantly*

increase the average number of scenarios where both parties exhibited normative-behavioral consistency.

The left bars in Figure 7 provide the data for Result 4a, showing the average long-run performance in the Baseline and the Nudge. The figure reveals that the Nudge Treatment insignificantly increased long-run performance ($t = 0.29$, $p = 0.77$). This finding contradicts Hypothesis 2a, and does not lend causal support to Result 1b, which described a positive correlation between endogenously emerging principles and long-run performance in the Baseline.

Figure 7: **Long-Run Performance in Baseline vs. Nudge Treatment**



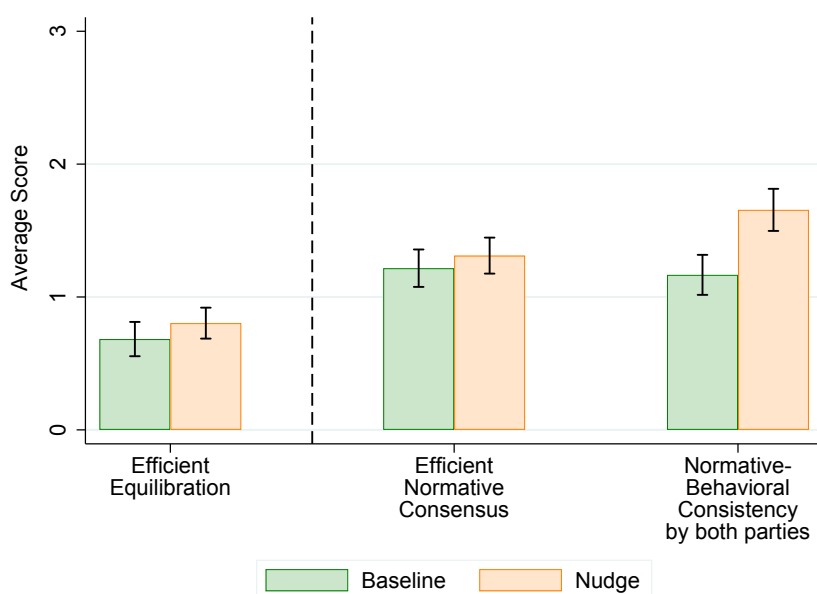
Notes: Data from the Baseline and the Nudge Treatment included. The bars represent average long-run performance in the Baseline and the Nudge Treatment, both at the overall treatment level and separately for pairs with principle-based agreements and pairs with rule-based agreements ($n = 121$ for the overall treatment effect; $n = 70$ for principles, $n = 48$ for rules; 3 pairs neither agreed on a rule nor on a principle). The error bars represent plus/minus one standard error of the mean.

The middle and right bars show the long-run performance separately for pairs with principle- versus with rule-based agreements. Both pairs with rule-based agreements and pairs with principle-based agreements exhibited lower performance in the Nudge than in the Baseline (Principles: $t = 1.70$, $p < 0.10$; Rules: $t = 3.51$, $p < 0.01$). Numerically, the fact that overall performance increased slightly, although principle- and rule-based pairs both had lower performance, is explained by the increase in the proportion of principle-based pairs (from 30% to 85%, as reported in the legend): principles do better than rules in both the Baseline and the Nudge, and there are more pairs with

principles in the latter.

Within the Nudge, the very poor performance of the very few pairs (15%) who articulated a rule-based agreement is not surprising: it may reflect negative selection. In contrast, the lower performance for pairs with principle-based agreements (in the Nudge compared to the Baseline) was unexpected. This result provides a first indication that at least some of the principles induced by the Nudge may have been less effective in inducing relational contracts, compared to the principles that emerged endogenously in the Baseline. We discuss this point in more detail in section 4.2.3.

Figure 8: **Effects of the Nudge on Efficient Equilibration, Efficient Normative Consensus and Normative-Behavioral Consistency**



Notes: The bars represent the average Efficient Equilibration Score, Efficient Normative Consensus Score, and Score of Normative-Behavioral Consistency by both parties for the Baseline ($n = 60$) and the Nudge ($n = 61$). The error bars represent plus/minus one standard error of the mean.

Figure 8 provides the data behind Results 4b and 4c. It shows the average scores for efficient equilibration, efficient normative consensus, and normative-behavioral consistency (by both parties) in the Baseline and in the Nudge. Similar to the finding for long-run performance in Figure 7, the left-hand side of Figure 8 indicates a small but insignificant increase in efficient equilibration in the Nudge ($t = 0.69$, $p = 0.49$), thus providing no support for Hypothesis 2b and no causal support for Result 2b, which showed a significant positive correlation between endogenously emerging principles

and efficient equilibration in the Baseline.

The right-hand side of the figure provides the data for Result 4c, revealing that the Nudge caused only an insignificant increase in efficient normative consensus ($t = 0.49, p = 0.63$) but a significant increase in the average number of scenarios in which both parties exhibited normative-behavioral consistency ($t = 2.24, p = 0.03$). Both the insignificant treatment effect for efficient normative consensus and the significant treatment effect for normative-behavioral consistency mimic the correlational results in the Baseline comparing these outcomes for principle- versus rule-based agreements (see Result 3b and Figure 6). The difference between the treatment effects and the analogous correlational results is that there is not a significant treatment effect for efficient equilibration. We explore the reasons for this finding in the next subsection.

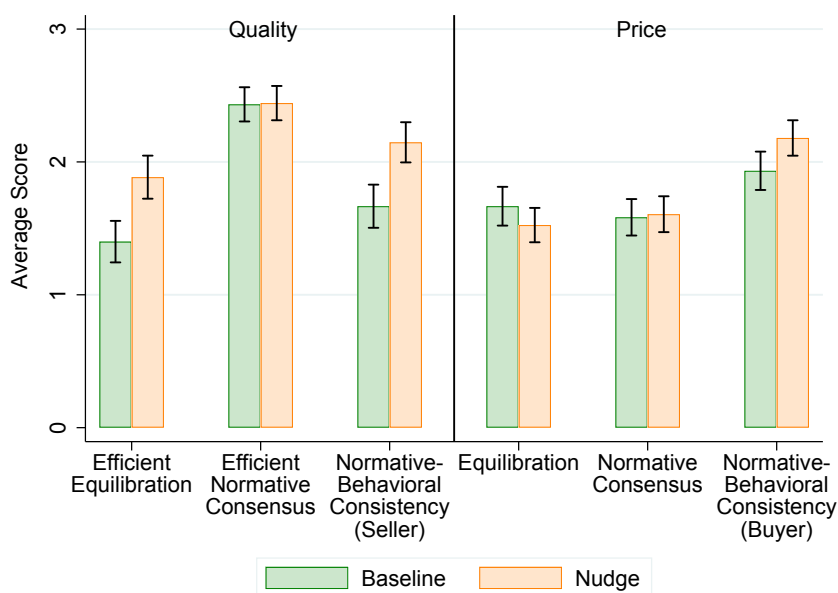
4.2.2 Understanding the Impact of the Nudge on Equilibration

To gain a better understanding of why the Nudge did not increase efficient equilibration, we now go beyond our initial hypotheses to explore potential explanations. Figure 9 decomposes efficient equilibration and efficient normative consensus into their price and quality components, and it displays normative-behavioral consistency individually for buyers and sellers. The figure provides several interesting insights.

First, the Nudge did cause pairs to achieve significantly higher efficient equilibration in quality ($t = 2.16, p = 0.03$)—meaning that the seller provided a quality at least as high as the buyer expected. Second, the Nudge also created a significant increase in normative-behavioral consistency by the sellers ($t = 2.17, p = 0.03$)—meaning that the seller chose the quality he thought the pair's agreement implied. In contrast, efficient normative consensus on quality was very high in both the Baseline and the Nudge, and there was no significant difference by treatment ($t = 0.05, p = 0.96$).

Turning to the price reveals a very different story. The Nudge failed to increase equilibration on price. In fact, we see an insignificant decrease in equilibration on price ($t = 0.73, p = 0.47$)—the buyer was not significantly more likely to pay a price as high as the seller expected. Furthermore, the Nudge caused only an insignificant increase in the normative-behavioral consistency of the buyer ($t = 1.26, p = 0.21$)—the buyer was not significantly more likely to pay the price she thought the pair's agreement implied. Finally, as with quality, normative consensus on price is essentially unchanged between Baseline and Nudge ($t = 0.12, p = 0.90$), although we note for discussion below

Figure 9: **Treatment Effect on Efficient Normative Consensus, Efficient Equilibration on Quality and Price, and Normative-Behavioral Consistency by Buyers and Sellers**



Notes: Data from the Baseline and the Nudge Treatment included ($n = 121$). The bars represent the proportion of pairs who reached a particular (Efficient) Equilibration Score, (Efficient) Normative Consensus Score, and Score of Normative-Behavioral Consistency by both parties—separately for the price and the quality dimension—in each of the two treatments. In the price dimension, since there is no “efficient” price level, we only speak of equilibration and normative consensus. The error bars represent plus/minus one standard error of the mean.

that the level of normative consensus on price is lower than that on quality.

Beyond the findings in Figure 9, we can say more about both quality and price. About quality, in addition to the positive treatment effect on efficient equilibration in quality there is also a positive treatment effect on short-run performance (i.e., the initial quality choices by the seller in the scenarios). Table 4 shows the results of a regression of average quality chosen in each of the 3 scenarios on a Nudge dummy. The data set is comprised of 363 observations (121 sellers, 3 observations per seller). Standard errors are clustered at the seller level.

Column (1) shows a marginally significant treatment effect: in the Nudge, when considering the three scenarios, sellers on average chose a quality that was 0.92 points, or 14%, higher than in the Baseline ($p = .08$). Column (2) shows the same regression, but additionally includes fixed effects for the different scenarios; the result is basically unchanged.

Turning to price, we illustrate the importance of achieving equilibration on *price* via the data

Table 4: **Treatment Effect on Sellers' Initial Quality Choices in the Three Scenarios**

	(1)	(2)
Nudge Treatment	0.92* (0.52)	0.92* (0.52)
Constant	6.78*** (0.36)	6.60*** (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 seller (pair) level are in parentheses. *** $p < .01$, ** $p < .05$, * $p < .10$

in Figure 10 on average *quality* over all the periods in the experiment. The figure considers all pairs in the Nudge that achieved efficient equilibration on quality in the scenario that was later implemented in Phase 2 ($n = 39$). The figure displays the time path of quality separately for pairs who achieved equilibration on price versus pairs who did not. Pairs 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). In contrast, pairs 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 deterioration of average quality over time (average quality was 6.6 over the 12 periods of Phase 2).

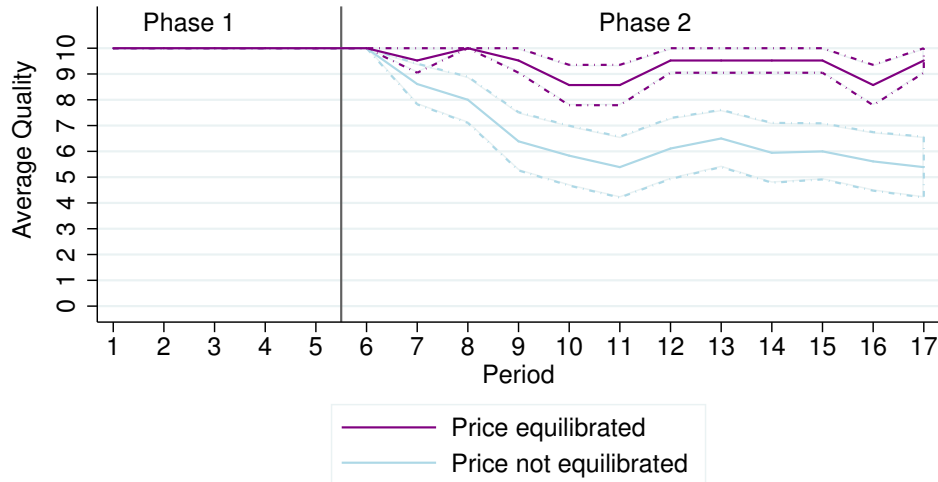
We conjecture that the deterioration of quality in pairs lacking equilibration on price resulted from conflicts about the fair distribution of payoffs in the early periods of Phase 2, which in turn led the seller to decrease the quality he was willing to deliver in later periods. Consistent with this conjecture, pairs without equilibration on price in the implemented scenario achieved significantly lower long-run performance in Phase 2 ($t = 2.90$, $p < .01$).

We summarize these exploratory findings in our next result:

Result 5

- The Nudge did significantly increase equilibration on quality, associated with marginally significantly higher quality in all three scenarios.*
- The Nudge's failure to increase long-run performance was associated with its failure to increase equilibration on price, which was the bottleneck for increasing efficient equilibration.*

Figure 10: **Average Quality Conditional on Efficient Equilibration in Quality, after Equilibration in Price**



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

Importantly, when we also include data from the Baseline in an analysis like Figure 10, we find no significant differences in the relationship between equilibration in price and long-run performance between the Baseline and the Nudge ($t = 0.71$, $p = 0.48$). Figures 9 and 10 thus illustrate two important features of our findings: (a) similarity across the Baseline and Nudge conditions, such as in the importance of achieving equilibration on price, and (b) variation among pairs within a condition, such as in whether a pair achieved equilibration on price. Both (a) and (b) motivate our next analysis, of possible heterogeneity among principles—in both the Baseline and Nudge conditions, but more importantly in the latter.

4.2.3 Digging Deeper: Heterogeneity Among Principles?

Recall from Figure 7 that there were many more principle-based initial agreements in the Nudge than in the Baseline but that on average such agreements in the Nudge had lower long-run performance than in the Baseline. In each condition, rule-based agreements performed worse than principle-based agreements, so having fewer rules helped aggregate performance in the Nudge, but this decline in the performance of principles led aggregate performance to be only insignificantly

higher in the Nudge than in the Baseline.

In this subsection we explore (1) whether there is important heterogeneity among pairs with principles (regardless of whether those pairs were in the Nudge or the Baseline) and (2) whether such heterogeneity helps explain this difference between conditions in the performance of principle-based agreements. All the analyses in this subsection therefore study the sub-sample of pairs with principles—considering data from the Baseline and the Nudge.

We now utilize three additional data sources. First, we examine personal characteristics of the parties such as age, gender, and field of study. Second, we extract further data from the chats concerning the agreement-formation process: we coded for (i) shared understanding about Phase 1, (ii) references to promises and trust, (iii) talk about Phase 2, and (iv) the mood of the chat. Third, we examine data from an exit questionnaire that asked individual parties their views on the helpfulness, clarity, and shared interpretation of their agreement. Here we report the highlights of our analyses; for a fuller report, see the Online Appendix.

We do not find any important differences in personal characteristics between pairs with principle-based agreements in the Nudge and the Baseline. Also, personal characteristics are not significantly correlated with long-run performance (see the Online Appendix for details).

In contrast, our additional codings from the chats suggest differences in the agreement-formation process across conditions. Concerning shared understanding of Phase 1, pairs with principle-based agreements in the Nudge more often had one party explain to the other the strategy combination ($p = 70, q = 10$) that yields an equal split of the maximal surplus in Phase 1 ($t = 1.96, p = 0.06$). Moreover, conditional on eventually reaching this strategy combination for Phase 1 (which 67 of 70 pairs do), pairs in the Nudge needed 38 seconds (23 percent) longer to get there, but this difference is not statistically significant ($t = 1.11, p = 0.27$).

We also explored how these two measures concerning Phase 1—having one party explain to the other, and time to reach this agreement—correlate with long-run performance in Phase 2. Having one party explain to the other was not significantly correlated with long-run performance ($\rho = -0.08, p = 0.49$), but time to reach this agreement was ($\rho = -0.28, p = 0.02$).

Turning to the exit questionnaire, pairs with principle-based agreements in the Nudge were significantly less likely to report that their agreement was “clearly formulated” ($t = 2.76, p < 0.01$) and marginally significantly more likely to report that there had been differing interpretations of the

agreement between the two parties ($t = 1.75$, $p = 0.09$) than pairs with principle-based agreements in the Baseline. In this sense, pairs in the Nudge did not reach the same level of clarity about their principles as did pairs in the Baseline.

The exit measures were correlated with long-run performance. For example, there is a significant positive correlation with the perceived clarity of the principle ($\rho = .35$, $p < .01$) and a significant negative correlation with differing interpretations of the principle ($\rho = -.72$, $p < .01$).

We interpret these correlations as consistent with the idea that the Nudge led some pairs with weaker shared understanding of (even) Phase 1 to articulate principle-based agreements (largely concerning Phase 2), but the resulting agreements were less clear and less often had a shared interpretation. In short, we conjecture that many of the additional principles in the Nudge were of low caliber, resulting in the difference between conditions in aggregate performance for principle-based agreements.

Ideally, our data would inform us about what the components of high-caliber principles are. One interesting observation is that promises were invoked in 18% of all Baseline pairs, but in only 5% of all Nudge pairs ($t = 1.98$, $p = 0.05$). Our current analysis of this issue is necessarily exploratory. We propose possible future work in the next section.

5 Discussion

5.1 Summary

This paper studied how parties build relational contracts that help them adapt after an unforeseen shock. We distinguished 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 pairs with rule-based agreements achieved a similar level of shared understanding regarding what should be done in a new situation (efficient normative consensus), they were more often unwilling to act accordingly (lack of normative-behavioral consistency). In contrast, pairs that relied on principles were more likely to be efficiently equilibrated after an exogenous shock, adapting their relational contract to the new environment.

To assess causality, we tried to stimulate principle-based agreements via a simple salience Nudge.

Our treatment indeed produced significantly more principle-based agreements. The Nudge also produced an increase in efficient equilibration on quality, but it failed to increase equilibration on price. Put differently the Nudge increased coordination on efficiency (the quality dimension) but not on distribution (the price dimension). Conflicting views about distribution may have been the cause of the deterioration of cooperation in the post-shock game. As a result, the Nudge did not have a significant effect on long-run performance.

5.2 Causal Effects of the Nudge

In this subsection we discuss two potential reasons why our Nudge had no treatment effect on long-run performance. One possibility is simple: there is no causal link between principles and performance. Instead, the correlation between principles and performance observed in the Baseline solely reflects omitted-variable bias—i.e., pairs that endogenously articulated principles may have performed better because other unobserved factors drive both the emergence of principles and performance. Exogenously increasing the frequency of principles, without affecting these other factors, would then not produce an increase in performance.

We argue against this interpretation as the sole explanation for our results, because we do observe significant treatment effects on some factors that contribute to performance, such as efficient equilibration in quality, short-run performance and normative-behavioral consistency for the seller. If the correlation between principles and performance in the Baseline is fully ascribed to omitted-variable bias, it is difficult to explain these causal effects in the data. Our findings therefore suggest that principles can play some role in adapting equilibria in unstable environments.

We believe that a second interpretation has considerable merit: our treatment may not have been as strong as it initially appeared from the coding presented in Section 4.1.1, for two reasons.

First, explicitly articulating a 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 coding of text can capture only what people actually write down, not what they may mean or think. Thus, in the Baseline, some pairs that were coded as having relied on only rules may in fact have also agreed implicitly on principles. Because the Nudge treatment pushed participants to articulate general principles, some of these pairs may have made an implicit principle explicit. They would then be coded as having a principle, but their

performance would not have changed.

Second, as discussed in Section 4.2.3, the Nudge induced some pairs with a lower shared understanding of Phase 1 to agree on a principle concerning Phase 2, but these pairs wrote lower-caliber principles than those who agreed on a principle without the Nudge. Since we only have a binary coding of agreements as containing a principle or not, we cannot account for differences in the caliber of principles in our analysis.

In sum, while we acknowledge that our data can be interpreted as there being no causal link between principles and performance, we see several reasons to think that our lack of a treatment effect is due to the Nudge being less effective than intended. An important path for future research is to explore alternative treatments that may be more powerful in getting pairs to agree on high-caliber principles.

5.3 Implications for Practice and Research

Perhaps one should not be surprised that our simple Nudge did not improve long-run performance. 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. 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 may in turn suggest challenges for researchers striving 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 (what we would call) 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.

The view that fruitful long-term collaborations require substantial work on relationship-building is also reflected in the experience of practicing lawyers and business consultants. Frydinger et al. (2019) argue that it is crucial to conduct a structured relationship-building process that specifies mutual goals and aligns expectations before a deal is signed. (Frydinger & Hart, 2019 include this idea in a formal model.) Our results are consistent with the importance of such a process and show that building a relationship is not trivial even if the setting is relatively simple.

Regarding future research, especially in organization theory and organizational economics, we hope that those contributing to the research streams described in Section 1—concerning not just relational contracts but also the related aspects of transaction-cost economics, the behavioral theory of the firm, informal organization, new organizational forms, and organizational culture—will consider not just the potential role of relational contracts once they are built but also the challenges in building and changing relational contracts as the world changes. 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 exists wide variation in this regard across pairs of parties (see also Proto et al., 2019). The experimental literature on equilibrium selection is informative on how differences across environments drive equilibrium selection (see, for example, Van Huyck et al., 1990; Goeree & Holt, 2001), but it is largely silent about differences across pairs within a fixed environment.

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. By showing that principles are associated with equilibration in uncertain settings, we provide one indication of what successful communication should be about.

Finally, 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 usefully explore the impact of shared understandings in relationships that managed to reach such understandings, but that does not imply that real parties can easily create such shared understandings. We need a theory of repeated games where different shared understandings with significant performance implications can develop even under identical initial conditions.

5.4 Limitations and Future Directions

While we believe our experimental results provide 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. One might 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 Nudge did not lead to a significant increase in long-run performance. One reason for this failure might have been, as we noted in section 4.2.3, that our Nudge was not effective in causing pairs to switch to high-caliber principles. In a certain sense, we see this ineffectiveness as encouraging, or at least consistent with prior thinking: high-performing relational contracts may be hard to build, even in laboratory settings.

The simple Nudge manipulation in our experiment did push a few behavioral levers, but failed to improve what turned out to be the bottleneck to high performance, equilibration in the price dimension. Future research should examine whether other manipulations can be more effective at causing pairs to agree on high-caliber principles that help increasing all dimensions of equilibration and, in turn, performance.

Our exploratory data analysis points to several potentially fruitful avenues for future research in this regard. First, while our Nudge made Phase 2 of the experiment more salient, it may be useful to focus subjects especially on distributional issues within this phase. Second, as we reported at

the end of Section 4.2.3, the Nudge may have crowded out discussion of promises and references to trust. Alternative treatments could emphasize these aspects. Finally, we have seen that principles induced by the Nudge were more likely to be interpreted differently by members of the pairs. Future treatments could push subjects towards creating clarity in their agreements, to avoid situations in which contending interpretations cause deteriorations in efficiency.

5.5 Conclusion

In this paper we began to explore how parties build relational contracts that achieve efficient adaptation to unforeseen change. Our results suggest that relational contracts based on general principles can improve long-run performance. However, we also find that it is difficult to induce the right kind of principles—those that parties not only understand but also follow. In retrospect, the latter finding is consistent with the logic of competitive advantage: if it were simple to build such relational contracts why is not everyone doing it? That said, our results also leave important open questions. We therefore hope this 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 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.
- Brown, M., Falk, A., & Fehr, E. (2004). Relational contracts and the nature of market interactions. *Econometrica*, 72(3), 747–780.
- Camerer, C. & Linardi, S. (2019). Robustness of relational contracts to interruptions in employer-worker gift exchange experiments. *Available at SSRN 3382951*.
- 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.
- Dore, R. (1983). Goodwill and the spirit of market capitalism. *The British journal of sociology*, 34(4), 459–482.
- Eccles, R. G. (1981). The quasifirm in the construction industry. *Journal of Economic Behavior & Organization*, 2(4), 335–357.
- 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.
- 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.
- Frydlinger, D., Hart, O., & Vitasek, K. (2019). A new approach to contracts: how to build better long-term strategic partnerships. *Harvard Business Review*, 97(5), 116–126.
- Frydlinger, D. & Hart, O. D. (2019). Overcoming contractual incompleteness: The role of guiding principles. Technical report, National Bureau of Economic Research.
- 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.
- Geertz, C. (1973). *The interpretation of cultures*, volume 5019. Basic books.
- Gibbons, R. & Henderson, R. (2012). Relational contracts and organizational capabilities. *Organization Science*, 23(5), 1350–1364.
- Gibbons, R. & Prusak, L. (2020). Knowledge, stories, and culture in organizations. *American Economic Review, Papers and Proceedings*, 110.
- 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.
- Gulati, R. (1995). Does familiarity breed trust? the implications of repeated ties for contractual choice in alliances. *Academy of management journal*, 38(1), 85–112.
- 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.
- Kogut, B. (1989). The stability of joint ventures: Reciprocity and competitive rivalry. *The journal of industrial economics*, 183–198.
- Kreps, D. M. (1990). Corporate culture and economic theory. *Perspectives on Positive Political Economy*, 90, 109–10.
- Kreps, D. M. (1996). Markets and hierarchies and (mathematical) economic theory. *Industrial and corporate change*, 5(2), 561–595.
- Lawrence, P. & Lorsch, J. (1967). *Organization and the Environment*. Boston, MA: Harvard Business School Press.
- Li, J., Matouschek, N., & Powell, M. (2017). Power dynamics in organizations. *American Economic Journal: Microeconomics*, 9(1), 217–41.
- Lumineau, F. & Malhotra, D. (2011). Shadow of the contract: How contract structure shapes interfirm dispute resolution. *Strategic Management Journal*, 32(5), 532–555.
- Macaulay, S. (1963). Non-contractual relations in business: A preliminary study. *American sociological review*, 55–67.
- Macneil, I. (1978). Contracts: Adjustments of long-term economic relations under classical, neo-classical, and relational contract law. *Northwestern University Law Review*, LCCII, 854–906.
- 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.
- Malhotra, D. & Lumineau, F. (2011). Trust and collaboration in the aftermath of conflict: The effects of contract structure. *Academy of Management Journal*, 54(5), 981–998.
- Marschak, J. & Radner, R. (1972). *Economic Theory of Teams*. New Haven, CT: Yale University Press.

- Mayer, K. J. & Argyres, N. S. (2004). Learning to contract: Evidence from the personal computer industry. *Organization Science*, 15(4), 394–410.
- Poppo, L. & Zenger, T. (2002). Do formal contracts and relational governance function as substitutes or complements? *Strategic management journal*, 23(8), 707–725.
- Powell, W. W. (1990). Neither market nor hierarchy: Network forms of organization. *Research in Organizational Behavior*, 12, 295–336.
- 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.
- Simon, H. A. (1951). A formal theory of the employment relationship. *Econometrica: Journal of the Econometric Society*, 293–305.
- Skorbiansky, S. R. (2018). Investing in communication: An experimental study of communication in a relational contract setting. *Journal of Behavioral and Experimental Economics*, 74, 85–96.
- 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. (1975). 1975 markets and hierarchies: analysis and antitrust implications new york: Free press.

- 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.