DISCUSSION PAPER SERIES

DP16802

Building an Equilibrium: Rules versus Principles in Relational Contracts

Robert Gibbons, Manuel Grieder, Holger Herz and Christian Zehnder

ORGANIZATIONAL ECONOMICS

CEPR

ISSN 0265-8003

Building an Equilibrium: Rules versus Principles in Relational Contracts

Robert Gibbons, Manuel Grieder, Holger Herz and Christian Zehnder

Discussion Paper DP16802
Published 10 December 2021
Submitted 09 December 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Building an Equilibrium: Rules versus Principles in Relational Contracts

Abstract

Effective collaboration within and between organizations requires efficient adaptation to unforeseen change. We study how parties build relational contracts that achieve this goal. We focus on the "clarity problem"—whether parties have a shared understanding of the promises they make each other. Specifically, (a) a buyer and seller play a trading game in several periods, (b) they know their environment will change but do not know how, and (c) before any trading occurs they can reach a non-binding agreement about how to play the entire game. We hypothesize that pairs whose initial agreement defines a broad principle rather than a narrow rule will be more successful in solving the clarity problem and in achieving efficient adaptation after unforeseen change. In our Baseline condition, we indeed observe that pairs who articulated 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 agreement prescribed. To investigate a causal link between principle-based agreements and performance, we implemented a "Nudge" intervention that induced more pairs to articulate principles. The intervention succeeded in co- ordinating more pairs on efficient quality immediately after the unforeseen change, but it failed to coordinate expectations on price, ultimately leading to conflicts and preventing an increase in long-run performance after the shock. Our results suggest that (1) principle-based agreements may improve organizational performance, but that (2) high-performing relational contracts may be difficult to build.

JEL Classification: N/A

Keywords: Relational Contracts, clarity, Adaptation, Cooperation, principles

Robert Gibbons - rgibbons@mit.edu
MIT, Sloan School of Management and CEPR

Manuel Grieder - Manuel.grieder@zhaw.ch
Faculty of Economics, UniDistance Suisse

Holger Herz - holger.herz@unifr.ch
Department of Economics, University of Fribourg

Christian Zehnder - christian.zehnder@unil.ch
Faculty of Business and Economics, University of Lausanne

Acknowledgements
We are very grateful for comments from Sylvain Chassang and Todd Zenger, as well as seminar participants at the Stanford CASBS Summer Institute (2016), LMU Munich, Innsbruck, Paris School of Economics, INSEAD, SIOE (2019) in Stockholm, the Organization Science Special Issue Online Conference (2020), CEPR, the 6th Workshop on Relational Contracts in Moscow (2020), GATE Lyon-St-Etienne, Harvard Business School, and Tilburg University. We also thank Lino Eden, Eleonora Egli, and Elise Moret, as well as Stefan Wehrli and the staff at the ETH Zurich 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. Prior to data collection, the experiment reported in this paper received IRB approval from ETH Zurich (EK 2017-N-04) and the MIT (COUHES #: 1704912215).

Building an Equilibrium:

Rules versus Principles in Relational Contracts

Robert S. Gibbons[*†] Manuel Grieder[‡§] Holger Herz[¶] Christian Zehnder[‖]

June 11, 2021

Abstract

Effective collaboration within and between organizations requires efficient adaptation to unforeseen change. We study how parties build relational contracts that achieve this goal. We focus on the "clarity problem"—whether parties have a shared understanding of the promises they make each other. Specifically, (a) a buyer and seller play a trading game in several periods, (b) they know their environment will change but do not know how, and (c) before any trading occurs they can reach a non-binding agreement about how to play the entire game. We hypothesize that pairs whose initial agreement defines a broad principle rather than a narrow rule will be more successful in solving the clarity problem and in achieving efficient adaptation after unforeseen change. In our Baseline condition, we indeed observe that pairs who articulated 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 agreement prescribed. To investigate a causal link between principle-based agreements and performance, we implemented a "Nudge" intervention that induced more pairs to articulate principles. The intervention succeeded in coordinating more pairs on efficient quality immediately after the unforeseen change, but it failed to coordinate expectations on price, ultimately leading to conflicts and preventing an increase in long-run performance after the shock. Our results suggest that (1) principle-based agreements may improve organizational performance, but that (2) high-performing relational contracts may be difficult to build.

Keywords: Relational contracts, clarity, adaptation, cooperation, principles

[*]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 are very grateful for comments from Sylvain Chassang and Todd Zenger, as well as seminar participants at the Stanford CASBS Summer Institute (2016), LMU Munich, Innsbruck, Paris School of Economics, INSEAD, SIOE (2019) in Stockholm, the Organization Science Special Issue Online Conference (2020), CEPR, the 6th Workshop on Relational Contracts in Moscow (2020), GATE Lyon-St-Etienne, Harvard Business School, and Tilburg University. We also thank Lino Eden, Eleonora Egli, and Elise Moret, as well as Stefan Wehrli and the staff at the ETH Zurich 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. Prior to data collection, the experiment reported in this paper received IRB approval from ETH Zurich (EK 2017-N-04) and the MIT (COUHES #: 1704912215).

1 Introduction

In this paper we use experimental methods to study parties building "relational contracts." Following economists such as Kreps (1996), we define relational contracts as having two features. First, relational contracts are informal: shared understandings so rooted in the details of the parties' relationship that they cannot be enforced by a court. Second, and importantly different from some other informal understandings, relational contracts are equilibria of repeated games: each party's strategy must be that party's best response to the other party's strategy.

We see such relational contracts as germane to several strands of organization theory—some studying activities within organizations and others studying activities between them. Within organizations, we see relational contracts in discretionary bonus schemes and in promises about empowerment or promotions, to name just a few examples. Between organizations, we see relational contracts in the way partners promise to react to events not covered in their formal contract, such as in supply chains or alliances or dealings among more than two parties. For ease of exposition, in this Introduction we refer to relational contracts within organizations; in Section 2 we discuss applications both within and between organizations.

Our interest in relational contracts within organizations flows from the possibility that organizations that reach different equilibria may achieve different levels of economic performance (Kreps, 1990; Gibbons & Henderson, 2012). Many celebrated accounts suggest that such differences in the relational contracts within organizations may arise from differences in pre-existing conditions—for example, see Hofstede (1980) on culture's consequences or Granovetter (1985) on embeddedness. This paper complements such work on differences in pre-existing conditions: we explore how organizations with identical pre-existing conditions may reach different relational contracts.

More specifically, we consider two problems that relational contracts must solve (Gibbons & Henderson, 2012). First, there is the *credibility* problem—"Should I believe the promise you are making me?" Second, there is the *clarity* problem—"Do we have a shared understanding of the promise you are making me?" Of the two features of relational contracts defined above—namely, relational contracts are informal and they are equilibria—the credibility problem is associated with the latter: it must be in the interest of each party to keep their promise given their belief about the other party's behavior. In contrast, the clarity problem is associated with the former: promises

that are so rooted in the details of the parties' relationship that they cannot be enforced by a court may also be difficult for the parties to state clearly to each other.

To study the clarity problem, we designed a laboratory experiment in the manner of experimental economics, exploring both how parties build a relational contract early in their relationship and how well they later adapt when their environment changes. Specifically, we implemented a repeated buyer-seller game with a random stopping rule, where participants knew the initial structure and parameters of the game but they 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 (including after the unspecified future change) 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 buyer-seller pairs reached a shared understanding about what behaviors would be appropriate if their environment were to change in particular ways. To collect the latter data, we interrupted interactions before the environment changed. We confronted participants with different scenarios of how the game *might* change and for each scenario asked them (a) what choices their agreement prescribes and (b) what choices they expect the parties to actually make. We then implemented one of the scenarios as the game that buyer-seller pairs played for the remaining periods.

These data from the scenarios allow us to define and assess steps we expect to be important for reaching game-theoretic equilibrium. In particular, we measure (1) the extent to which parties have a shared understanding of what ought to be done in a particular situation according to their agreement (*normative consensus* for the pair), (2) whether each party is willing to do what ought to be done (*normative-behavioral consistency* for each party), and (3) whether each party's actions would fulfill the other party's expectation (*equilibration* for the pair). Although game theory has been completely clear about what it means to have reached an equilibrium, these novel measures allow us to explore variations in the extent to which parties have *not* done so.

Motivated by the clarity problem, our first results explore the association between these measures and economic performance after the shock (defined as the sum of profits earned by a buyer-seller pair in the repeated game after the environment changed). Using data from the Baseline setting sketched above, we find that pairs who lacked either normative consensus or normative-behavioral consistency were unlikely to achieve equilibration. And we then find that achieving

equilibration was strongly correlated with higher performance after the shock. These findings suggest that low performance may result not only from reaching a bad equilibrium but also from failing to reach an equilibrium at all.

In our setting, where adaptation to a future shock is crucial for performance, the clarity problem may be an important reason why parties do not reach an equilibrium. The second step in our analysis therefore explores how pairs might avoid the clarity problem: we hypothesize that pairs who agree on a broad principle (applicable to a range of possible environments) rather than a narrow rule (specific to the known environment) when discussing their initial agreement have a better chance after the shock to produce normative consensus, normative-behavioral consistency, equilibration, and high performance.[1]

In our Baseline setting we indeed find that pairs whose initial agreement articulated a principle tended to achieve higher post-shock performance than did pairs whose agreement articulated only a rule. Further analysis reveals the mechanism underlying this correlation: parties with principle-based agreements were <u>not</u> more likely to agree on what should be done according to their agreement (normative consensus), but such parties <u>were</u> more often willing to take the actions that were prescribed by their initial agreement (normative-behavioral consistency). As a consequence *parties with principle-based agreements before the shock were more likely than those with rule-based agreements to achieve equilibration after the shock.*

The results described thus far refer to correlations between endogenous variables within our Baseline setting and hence do not allow us to make causal claims regarding the effects of principle-based agreements on outcomes. To investigate whether there is a causal link between principle-based agreements and post-shock performance, we implemented a "Nudge Treatment" designed to exogenously foster the emergence of principle-based agreements. Our Nudge indeed had a large causal impact on inducing more pairs to articulate principles. Moreover, in our buyer-seller setting involving quality and price, the Nudge also had a causal impact on the number of pairs coordinating on efficient quality after their environment changed. However, the Nudge failed to improve coordination on a new price after the shock; to the contrary, some buyers paid less than sellers expected, creating conflicts and ultimately preventing an increase in long-run performance after

[1]Our distinction between broad principles versus narrow rules echoes the role that Gibbons & Prusak (2020) describe for stories in organizations—namely, stories can convey broad messages beyond the story's narrow details. For example, a story may convey such messages even if its details are known to be false.

the shock.

We see our findings from the Baseline and the Nudge as consistent with the view that high-performance relational contracts are hard to build—especially in settings where an effective agreement must coordinate the parties' adaptation after their environment changes, so solving the clarity problem is difficult. Our argument about relational contracts thus parallels Barney's (1986) argument about organizational culture: if relational contracts are important for competitive advantage then they cannot be easy to copy.

More specifically, in our Baseline setting a minority of pairs were able to reach principle-based agreements that improved clarity (as assessed by our novel measures) and supported efficient adaptation, but most pairs were not. The goal of the Nudge was to induce the latter kind of pairs to behave like the former. Apparently our particular Nudge was too weak to achieve this result: the Nudge induced more pairs to state that they had reached agreement on a principle, but many of these measured principles may have been of lower caliber (less useful in allowing the parties to achieve equilibration) than those that emerged endogenously in the Baseline. Accordingly, in many Nudge pairs that stated a principle, after the environment changed, at least one party departed from what the other party thought had been agreed.

We hope this Introduction has conveyed our paper's specific contributions about how different relational contracts (with different performance consequences) may emerge, holding pre-existing conditions constant. We discuss more general contributions in Section 2 below, where we consider related literatures. Section 3 then describes the experimental design, Section 4 summarizes our hypotheses, and Section 5 presents the results. In Section 6 we discuss implications of our results and conclude.

2 Related Work

As noted above, we see relational contracts as germane to several strands of organization theory (OT)—some studying activities within organizations and others studying activities between them. In the three sub-sections below we first discuss our connections to each of these two topics and then our connections to organizational economics (OE) and experimental economics.

2.1 OT and Relational Contracts Within Organizations

Within organizations, we see relational contracts as part of the "informal relations and the unofficial norms" that go beyond the "formal hierarchy of authority and the official body of rules" (Blau & Scott, 1962: 6). More specifically, in Cyert & March's (1963) quasi-resolution of conflict within organizations, we see relational contracts as helping to manage "the obvious potential for internal goal conflict inherent in a coalition of diverse individuals and groups" (1963: 31). In a similar vein, Williamson (1979) explicitly considered relational contracts within organizations.

Building on these classics, more recent (and more grounded) research in OT is even more closely connected to this paper. For example, consider Foss's (2003) discussion of Oticon's "spaghetti organization" and Turco's (2016) "conversational firm." Both Foss and Turco explicitly reference the relational-contract model of empowerment by Baker et al. (1999), which envisions empowerment as an equilibrium in a repeated game: below the top of an organization, decision rights are "loaned, not owned" (1999: 56). For the purposes of this paper, however, it is more important that both Foss and Turco also describe the dynamics of building and changing relationships—something that steady-state models like Baker et al. entirely omit.

In addition, our paper also connects with Kellogg's (2009) study of how "relational spaces" affect microinstitutional change within organizations. Unlike Foss and Turco, Kellogg does not explicitly consider equilibria of repeated games, but like Foss and Turco, Kellogg does describe dynamics of ongoing relationships that are beyond steady-state models like Baker et al. (1999).

We believe our work makes two contributions to this Foss-Kellogg-Turco strand of OT within organizations. First, we propose that the dynamics they describe could be building or changing a game-theoretic equilibrium—i.e., each party's strategy may come to be that party's best response to the other party's strategy. More specifically, we are not aware of work in OT that measures (1) whether parties agree on what should be done, (2) whether they then do what they think should be done, and (3) whether their resulting actions accord with the other party's expectations. And second, we propose that such an equilibrium, if and when it is reached, could be associated with economic performance. We are not aware of work in OT that connects measures such as (1)-(3) to the parties' subsequent economic performance, especially controlling for the possibly important effects of pre-existing conditions.

2.2 OT and Relational Contracts Between Organizations

Following Granovetter (1985: 490) we see the repeated-game logic of relational contracts as an important aspect of economic exchange both within and between organizations. Between two firms, we see such equilibria in the supply chains of Dore (1983), the alliances of Oliver (1990), and the "hybrids" of Williamson (1991); among more than two firms, we see such equilibria in the quasifirms of Eccles (1981), the joint ventures of Kogut (1989), and the networks of Powell (1990).

Parallel to our discussion of relational contracts within organizations, when we now consider relational contracts between organizations, we are again interested in the possibility that differences in such equilibria may be associated with differences in economic performance. For relational contracts between organizations, there are again celebrated accounts suggesting the possible importance of pre-existing conditions—for example, see Gulati (1995) on historical exchange patterns or Powell et al. (1996) on network structure.

Like the rich accounts by Foss, Kellogg, and Turco of the development and change of relationships within organizations, longitudinal studies of the development and change of relationships between organizations relate to our work. As one example from the many between-organization relationships noted above (supply chains, alliances, hybrids, quasifirms, joint ventures, and networks), there are rich accounts of "disruptions" during specific alliances—see Doz (1996), Mayer & Argyres (2004), and Keller et al. (2021).

We believe our work makes the same two contributions to longitudinal studies of relationships between organizations as it does to the Foss-Kellogg-Turco strand of OT within organizations. Continuing to take alliances as one example of the many between-organization relationships, we are not aware of work on alliances that measures (1)-(3) or that connects such measures to the parties' subsequent economic performance, especially controlling for the possibly important effects of pre-existing conditions.

There is of course one important difference that often arises in considering relationships within organizations, on the one hand, versus relationships between organizations, on the other: the latter are much more likely to involve not only relational contracts but also formal contracts that have standing in courts. Our experiment does not involve formal contracts. Similarly, our experiment does not involve systematic variation in pre-existing conditions such as those explored by Hofstede

(1980), Granovetter (1985), Gulati (1995) or Powell et al. (1996). In this sense, we hold both pre-existing conditions and formal contracts constant and then study how different relational contracts may nonetheless emerge.

Finally, there is a growing literature on interactions between formal contracts and "relational governance," where the latter includes the effects of pre-existing conditions—see Poppo & Zenger (2002) and Reuer & Ariño (2007) for early contributions and Schepker et al. (2014) and Cao & Lumineau (2015) for surveys. As just described, we hold both formal contracts and pre-existing conditions constant, so our work is different from and complementary to this literature. In addition, we are also different from and complementary to this literature in the ways we analyze relationships developing over time (given fixed formal contracts and pre-relationship conditions). Specifically, in terms of the credibility and clarity problems defined in Section 1, this literature has explored the former, such as by studying the development of trust over time, whereas we focus on the latter.

In sum, since influential work such as Poppo & Zenger (2002) discusses economic models of relational contracting as part of relational governance, we hope our results concerning relational contracting are seen as advancing the study of relational governance. That said, there is of course much work to be done, including the introduction of formal contracts to settings like ours, where adaptation to unforeseen shocks is crucial for economic performance.

2.3 OE and Experimental Economics

Turning to related work in economics, there is a large theoretical literature on relational contracting, but the many models like Baker et al. study the steady-state properties and consequences of relational contracts that are assumed to exist. This steady-state literature has made important contributions (see Malcomson, 2013 for a survey of 20 years of work), but such models ignore the challenges that real parties face before their relationship has begun. Some of the most interesting recent papers—such as Chassang (2010) or Andrews & Barron (2016) or Li et al. (2017)—analyze settings with learning or adaptation, but it is exactly in such settings where real parties may have most difficulty building an equilibrium.

Methodologically, our approach follows the traditions of experimental economics. More specifically, our study is related to experiments on cooperation in repeated social dilemmas (see, e.g., Dal Bó, 2005; Dal Bó & Fréchette, 2011 for infinitely repeated games, and e.g., Brown et al., 2004;

Camerer & Linardi, 2019 for finitely repeated games), and to experiments studying how asymmetric information impedes shared understanding (Herz et al., 2019). 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.

Finally, our experiment also builds on previous experimental studies of how communication between partners facilitates cooperation (see, e.g., 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).

3 Study Design

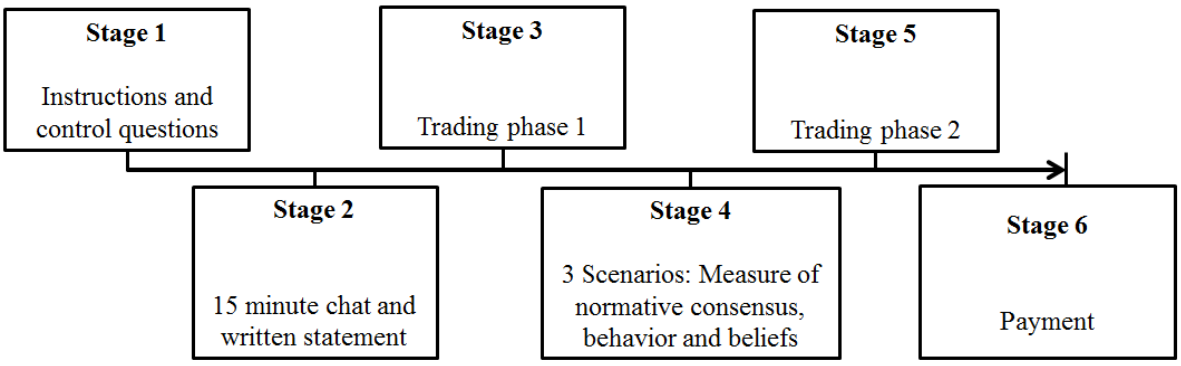
To study how parties build relational contracts, how they adapt after shocks, and to assess the clarity problem we created a laboratory experiment in which a buyer (she) and a seller (he), engaged in a repeated trading game with uncertainty about the future.

3.1 Baseline Experimental Set-up

Participants were randomly assigned to a role (buyer or seller) and matched in pairs. Subjects were aware that partners and roles would remain fixed for the entire experiment. Figure 1 represents the timeline of the experiment. The full instructions for participants are shown in the Online Appendix.[2]

[2]The Online Appendix, the experimental software, and the data are available on OSF (`https://osf.io/ne7rd/?view_only=dfeb89ecd8a64dfaa80ecc5294fb941b`).

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



Stage 1: Instructions and Control Questions. The instructions provided detailed information on the first (five-period) trading phase (see Appendix C). Participants were made aware of the fact that the game parameters and structure would change after 5 periods, but no information was provided about the nature of the change. Participants also knew that 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.

Stage 2: Chat Communication and Statement. Participants exchanged anonymous, free-form text messages with their interaction partner via computer for 15 minutes (they were instructed not to reveal their identities). We told participants to come up with a non-binding agreement on how they intended to play the game and to write down a short joint statement. This statement was drafted by the seller in a separate entry window. The buyer could demand changes through the chat function. Once agreement was reached, the buyer could confirm the statement with a mouse click. If no statement was confirmed within 15 minutes the groups continued without a statement.[3]

Stage 3: First Trading Phase. All pairs played five rounds of a simultaneous buyer-seller game in which the seller chose a quality $q \in [0, 10]$ at cost $c(q)$ and the buyer determined the price $p \in [0, 100]$.[4] Table 1 displays the cost function $c(q)$.[5] The value of the product to the buyer in the first trading phase was $v = 10q$. In any period, the buyer's payoff was thus $\pi_B = 10q - p$ and the

[3]It is possible that a pair reached an agreement without having the time to write it down. For our analyses, we therefore coded agreements not only in the statements but also in the chats (see Section 3.4).

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

[5]All parameters are expressed in experimental points. The points had real value and were paid out in cash at the end of the experiment (exchange rate: 25 Points = US$ 1).

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 details, see Appendix D). To avoid order effects, we randomized the order of presentation. One of the three scenarios was implemented in Trading Phase 2 (see Stage 5).

For each scenario, we measured the *normative* and *behavioral* implications of the pair's initial agreement. We elicited the normative implications by asking 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 paid participants a bonus of 20 points for consensus with their partner (for each question in each scenario). To measure the behavioral implications (actual choices), we asked 2.a) "what will you do in the first period?" (of the second trading phase) and, to elicit the belief about the partner's behavior, 2.b) "what do you think that your partner will do in the first period?". Question 2.a) was incentivized by making the answer the participant's binding first period choice in case that scenario was implemented. We did not incentivize question 2.b).

Stage 5: Second Trading Phase. The implemented scenario changed two elements of the game. First, the buyer's value doubled for every quality level: $v = 20q$. Second, the buyer had access to an attractive outside option in some periods. With probability 2/3 the outside option was worthless ($v_o = 0$), but with probability 1/3 the outside option was valuable ($v_o = 160$; delivering a quality of $q^o_{hi} = 10$ at a price of $p^o_{hi} = 40$). If the buyer picked the outside option, the seller received a price of $p = 0$ for his product, but still incurred the production cost. In 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 making their choices. Participants knew that game would stop with probability 10% after each round.[7] These changes represent a tough test for the stability of relational contracts. The

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

[7]To ensure better comparability of results across sessions, 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. The random device

new parameters implied a different payoff-equalizing price and the presence of the outside option created a temptation for the buyer and a risk for the seller.

Stage 6: Payments. After the second trading phase, participants were provided with a summary of their earnings and received their payments, in cash and in private.

3.2 The Nudge Treatment

To assess the causal effect of principles on shared understanding and performance, we conducted a second treatment in which we exogenously stimulate the emergence of principles using a simple nudge which was visible in the instructions and on the screen of 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."*

The nudge did not add new information, but increased the salience of the upcoming change and suggested to think about general principles.

3.3 Participants and Procedure

We recruited 242 participants (52% women, $M_{age} = 21.8$ years, $SD_{age} = 2.6$ years) from a university subject pool. Participants received the equivalent of US$ 10 as a show-up fee. Further payment depended on performance in Stages 3-5. Average earnings amounted to US$ 49 per participant.

We conducted seven experimental sessions with 32 to 36 participants. Interactions were anonymous, but participants knew that their partner was another participant present in the room. The experiment was programmed in z-Tree (Fischbacher, 2007). The experimental manipulation was randomly assigned at the pair level within session (60 pairs in Baseline and 61 in Nudge).

set the number of periods to be played in the second phase to 12.

3.4 Coding of Joint Statements and Chats

Three research assistants independently coded the joint statements and the chat protocols of each pair for the presence of principles and rules. In the coder instructions (Appendix E) a principle was defined as "... not based solely on a numerical definition of the quality to be delivered or the price to be paid ..., but [it] 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. If a rule is combined with a principle, we interpret it as an example specifying the implications of the principle in the first trading phase. Importantly, both rules and principles were only to be coded if there was clear agreement within the pair.[8]

Table 2 shows that agreement among coders about principles and rules in statements and chats was generally very high (80-98%). Gwet's AC score (Gwet, 2008) indicates that inter-rater reliability is very good for all categories except principles in chats, where it is slightly lower.

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

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

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

4 Predictions

4.1 Framework and Terminology

Our study investigates how parties build relational contracts that solve the clarity problem and allow for efficient adaption to unforeseen change. We study repeatedly interacting buyer-seller pairs that experience an exogenous change in their trading environment. Relationships start in a setting with known parameters. The pairs are aware that adaptation will be needed, but the nature of the upcoming change is unknown.

[8]The coders also coded for some sub-categories that we do not use in our main analyses (see Appendix E).

In our buyer-seller interactions efficient cooperation could be sustained as an equilibrium of the infinitely repeated game.[9] The standard game-theoretic perspective suggests that the parties should not only cooperate from the outset but also immediately adapt their strategies to keep cooperating after the shock.[10] In contrast, we believe that such 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"). We focus on the latter.

Because the post-shock phase of our experiment is a repeated game (not a one-shot interaction), successful adaptation requires aligned beliefs and actions about both the quality to be provided and the price to be paid after the shock.[11] We use the term *equilibration* to describe a situation where a buyer-seller pair coordinates on a new price-quality combination that simultaneously meets the expectations of both parties. Specifically, a pair reaches equilibration 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 ($\hat{q}_B$). Intuitively, equilibration means neither party takes an action that disappoints the partner.[12]

Because the conditions of equilibration can be met even if the outcome is inefficient, we use the term *efficient equilibration*, if equilibration is combined with coordination on actions that maximize the sum of profits in the pair (i.e., our measure of performance). In our experiment efficiency is fully determined by quality: efficient equilibration therefore requires a quality of 10.

Our first goal is to understand the process that produces equilibration after change occurs.

[9] We 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.

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

[11] 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 efficiency-reducing conflicts.

[12] Note that equilibration is *not* equilibrium. In a game-theoretic equilibrium each party's strategy is a best response to the other party's strategy and both parties have correct beliefs. In a repeated game a party's strategy defines the party's choice of action for each possible history in the game on and off the equilibrium path. Our concept of equilibration, in contrast, is much less demanding. It only requires that each party's chosen action in a given period and for a given history matches or exceeds the expectations of the other party.

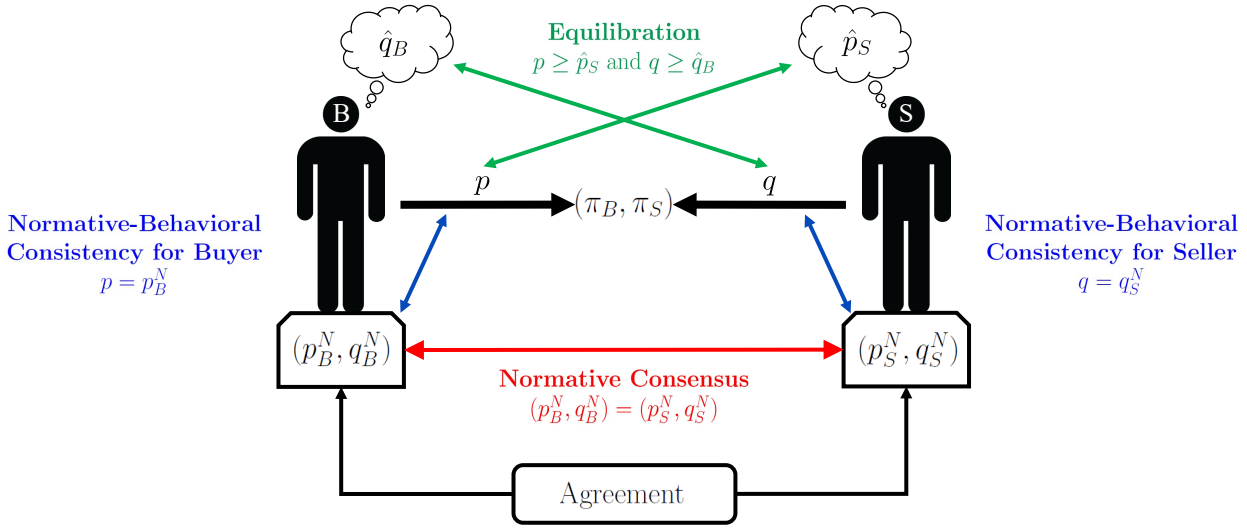
All the game-theoretic models of cooperation that we know are silent about this issue.[13] In these models equilibration is taken to be so simple that it does not appear in the model.

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. Intuitively, the term normative consensus thus refers to the fact that both partners interpret their agreement in the same way in a given 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*.

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 ($q = q_S^N$). Intuitively, normative-behavioral consistency thus means that each party does what he or she thinks the agreement implies. See Figure 2 for a summary of this terminology.

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

Figure 2: **Framework and Terminology**



Notes: 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 reaches *equilibration* 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 $q \geq \hat{q}_B$. *Efficient equilibration* is attained if equilibration is combined with 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 equilibration.[14] However, we 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) reach efficient equilibration after a change and (b) achieve efficient normative consensus and exhibit normative-behavioral consistency.

4.2 Testable Hypotheses

Our first hypothesis describes our predictions regarding the expected post-shock 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 reach efficient equilibration and will therefore outperform the pairs who govern their relationship with a rule (in Trading Phase 2, i.e., after the exogenous shock has occurred).

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.*

[14] That is, pairs *could* reach equilibration (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 implications. 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).

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 reach efficient equilibration after an exogenous change in the environment.*

Hypothesis 1 refers to correlations between endogenous variables within our Baseline setting. Accordingly, evidence supporting this hypothesis will not allow us to make any causal claims regarding the effects of principle-based agreements on outcomes. Our Nudge Treatment therefore aims at exogenously triggering the emergence of 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 condition.*

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 reach efficient equilibration after an exogenous change in the environment than pairs in the Baseline condition.*

5 Results

In this section, we present the results of our experiment. We first present results from the Baseline Condition in subsection 5.1 and then from the Nudge Treatment in subsection 5.2. For expositional reasons, we first present differences in performance between principle-based and rule-based agreements. We then relate these differences to efficient equilibration, efficient normative consensus, and normative-behavioral consistency.

5.1 Baseline Outcomes

5.1.1 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 post-shock performance:

Result 1

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

b) Pairs who formulated principle-based agreements achieved higher post-shock performance than those who relied on rule-based agreements.

Table 3 is based on our coding data and provides support for Result 1a. We show the relative frequency (in percent of all pairs) with which principle-based and rule-based agreements endogenously emerged—separately for the statement, the chat, and combining both of them.

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

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

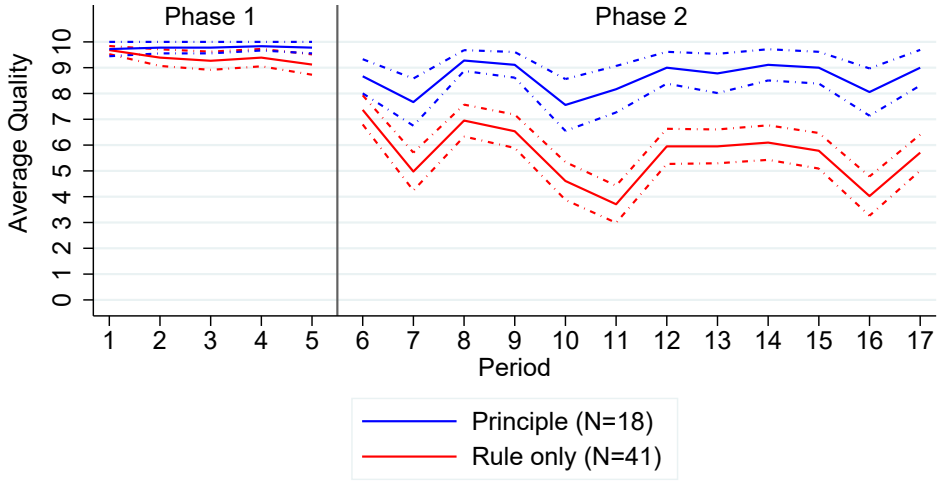
Notes: The table shows the coded frequency of pairs (in fraction of all pairs) who included principles or rules in their statements in the baseline. The first column contains only codings for the final written statements. The second column contains only codings for chat messages. The third column combines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

Because both the statements and the chats could have an impact on the initial agreement, our analyses rely on the combined coding of both statements and chats. We therefore say that a pair had a principle whenever a principle was agreed upon in the statement, the chat, or in both.[15] When considering the combined data, we observe that only 30% of all pairs articulated a principle, whereas (with the exception of one pair) all pairs wrote down a rule. However, it is important to take into account that the role of a rule changes depending on whether it is combined with a principle or not. Standalone rules prescribe price-quality choices irrespective of the state of nature and do not provide guidelines on how to adapt in case of a change in the environment. When in

[15] see Appendix B.1 for analyses of principles that were coded as present in the chats only versus in the statements.

combination with a principle, in contrast, the rule specifies the implication of the principle for the current setting in Phase 1 (which was known when the parties communicated). The latter case should therefore be seen as a principle with example.

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 Result 1b. It shows the average quality provided by sellers over time and distinguishes between pairs with only a rules-based agreement 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. In line with Hypothesis 1a, pairs with principles had average earnings of 1,752 points compared to 1,348 points for pairs with rules only ($t = 3.95$, $p < .01$).[16] In contrast, average Phase 1 earnings were very similar ($t = 1.21$, $p = .23$).[17]

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

[17] To disentangle the impact of the shock on Phase 2 performance from pre-existing time trends, we also performed an analysis in which we restrict our sample to those pairs that consistently achieved maximal profits in all periods of Phase 1 (36 out of 41 pairs with rule-based agreements and 17 out of 18 pairs with principle-based agreements). When comparing Phase 2 profits between these subgroups, we again find larger profits in Phase 2 for pairs with principles ($t = 3.40$, $p < 0.01$). Consequently, the difference between pairs with principle-based and rule-based agreements in Phase 2 cannot solely be a continuation of a pre-existing difference in trends across pairs.

5.1.2 The Role of Efficient Equilibration for Post-Shock Performance

We next explore the mechanisms underlying the observed correlation between initial agreement and performance after the shock. In Section 4, we hypothesized that a fruitful buyer-seller collaboration after the shock is facilitated if parties achieve equilibration of beliefs and actions on an efficient outcome.[18] 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 achieved efficient equilibration. 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 achieved efficient equilibration in the scenarios reached higher post-shock performance in the subsequent Trading Phase 2.

b) Pairs with principle-based initial agreements were more likely to achieve efficient equilibration in the scenarios.

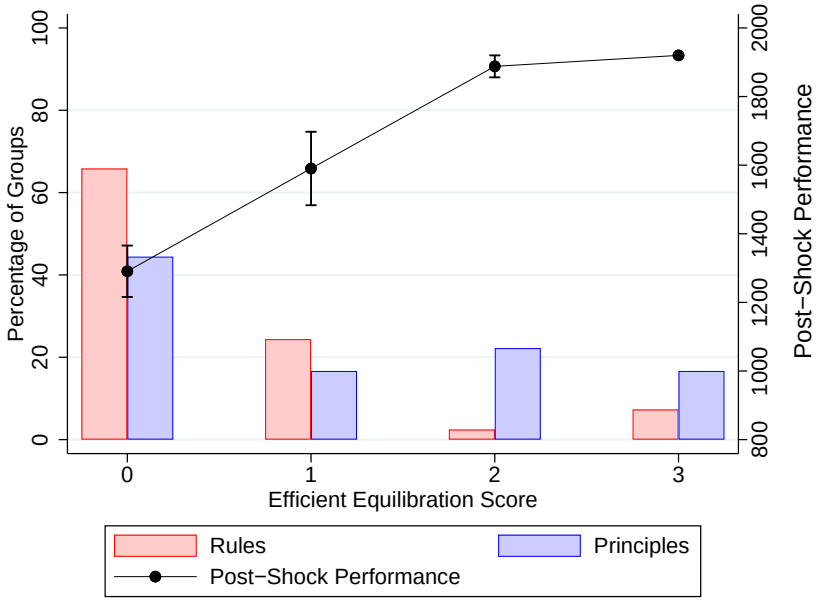
Figure 4 provides support Result 2a. As indicated by the black line, there is a clear correlation between post-shock performance, measured in terms of aggregate profits throughout Phase 2 (right-hand axis), and the Efficient Equilibration Score in the scenarios ($r = 0.53$, $p < .01$).[19]

Figure 4 also provides support for Result 2b. The bars show the proportion of pairs that reached 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

[18] Section 4 notes that it is possible for pairs to reach equilibration 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 reached equilibration on an efficient outcome. We therefore focus on efficient equilibration. Qualitatively, all our results hold when replicating the analysis for equilibration instead of efficient equilibration.

[19] Of the three scenarios presented 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 post-shock performance and the Efficient Equilibration Score based on only the two non-implemented scenarios. The result is essentially unchanged. The correlation then is $r = 0.46$ ($p < .01$) between post-shock performance and efficient equilibration.

Figure 4: **Efficient Equilibration, Post-Shock Performance, and Agreement Type**



Notes: Only data from the Baseline included. The line shows average profits in Phase 2 (post-shock 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).

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, $t = 1.93$, $p = 0.06$).

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

Result 2 established that post-shock performance is correlated with efficient equilibration in the scenarios. As discussed in Section 4, 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).[20] Note that all the measures in this analysis are drawn from the parties' responses to the scenarios, before the shock was implemented.

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 principles 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 more likely to develop efficient normative consensus, but display stronger normative-behavioral consistency (by both parties).

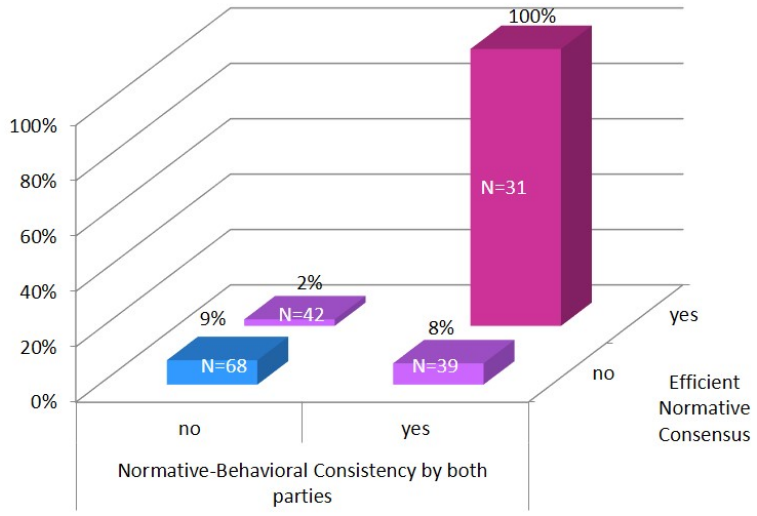
Recall from Section 4 that, strictly speaking, efficient normative consensus and normative-behavioral consistency are neither sufficient nor necessary conditions for pairs to reach efficient equilibration. Figure 5 illustrates that, in our data, the likelihood of efficient equilibration is extremely high (actually, in these data, certain) if efficient normative consensus and normative-behavioral consistency are both present, and extremely low otherwise. This observation supports Result 3a.

Figure 6 provides evidence for Result 3b. It shows in how many scenarios, on average, pairs with principle- versus pairs with rule-based agreements reached (a) efficient equilibration; (b) efficient normative consensus; and (c) had both parties display normative-behavioral consistency.

As a benchmark, the data to the left of the dashed line repeats the finding discussed in subsection 4.1.2 that pairs with principle-based agreements were more likely to reach efficient equilibration. New data is presented to the right of the dashed line. The bars in the middle of the figure show

[20] Again, section 4 notes that it is possible for pairs to have normative consensus on an inefficient outcome. However, our data reveal that this phenomenon plays no important role in our setting. in 91.3% of the cases in which a pair reached normative consensus, the consensus was on the efficient quality of 10. We therefore restrict our analysis to efficient normative consensus. All our results hold qualitatively when replicating the analysis for normative consensus instead of efficient normative consensus.

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



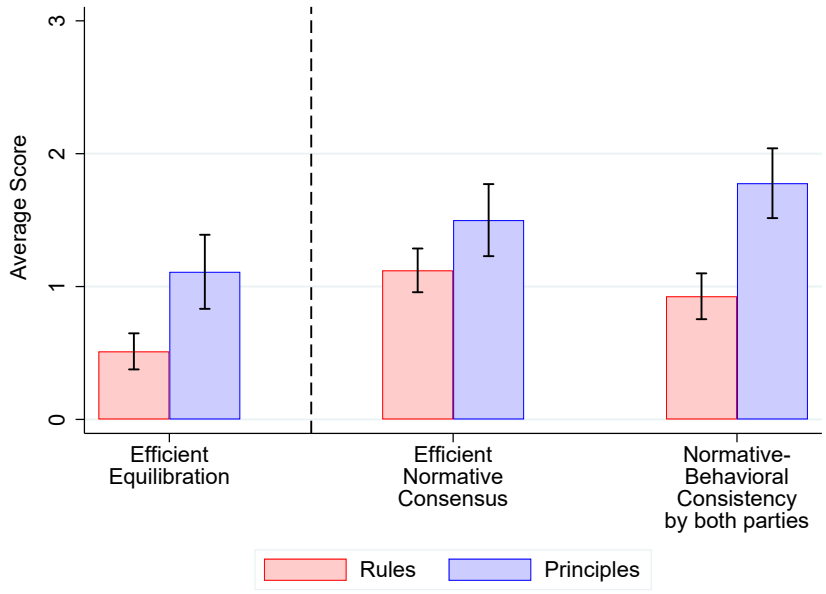
Notes: Only data from the Baseline included ($n = 180$ from 60 pairs in 3 scenarios). The horizontal axes categorize pairs based on whether both parties behaved normative-behaviorally consistent and whether the pair had efficient normative consensus at the scenario level. The vertical axis plots the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

that pairs with principle-based agreements reached, on average, a 34% higher score for efficient normative consensus. Even though the point estimate of a 34% difference indicates that principle-based pairs may have been more successful in establishing efficient normative consensus than rule-based pairs, given the variance and the sample size, this effect does not reach conventional levels of statistical significance ($t = 1.19$, $p = 0.24$). The right side of the figure shows, however, that pairs with principle-based agreements were 92% more likely to have both parties display normative-behavioral consistency ($t = 2.71$, $p = 0.01$). 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 strongly correlated with high performance in the Baseline ($r = 0.44$, $p < 0.01$).

5.2 Experimental Manipulation: Nudging the Emergence of Principles

So far we have shown that pairs who endogenously reached principle-based initial agreements had higher post-shock performance, were more likely to achieve efficient equilibration in the scenarios considered before the shock, and were more likely to have both parties display normative-behavioral

Figure 6: **Relationship between Principles and 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.

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.[21]

5.2.1 Treatment Effects on Post-Shock 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 of inducing more pairs to establish principle-based 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, $z = 6.15$, $p < .01$). As in the Baseline, all pairs who agreed on a general principle accompanied it with a rule stating the meaning of the principle for the already known parameters of 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.

[21]For an overview of summary statistics and correlations of key variables across both experimental conditions (Baseline and Nudge), see Appendix A.

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

Result 4

a) The Nudge did not increase overall post-shock performance.

b) The Nudge did not increase efficient equilibration.

c) The Nudge did not increase efficient normative consensus, but it did increase normative-behavioral consistency by both parties.

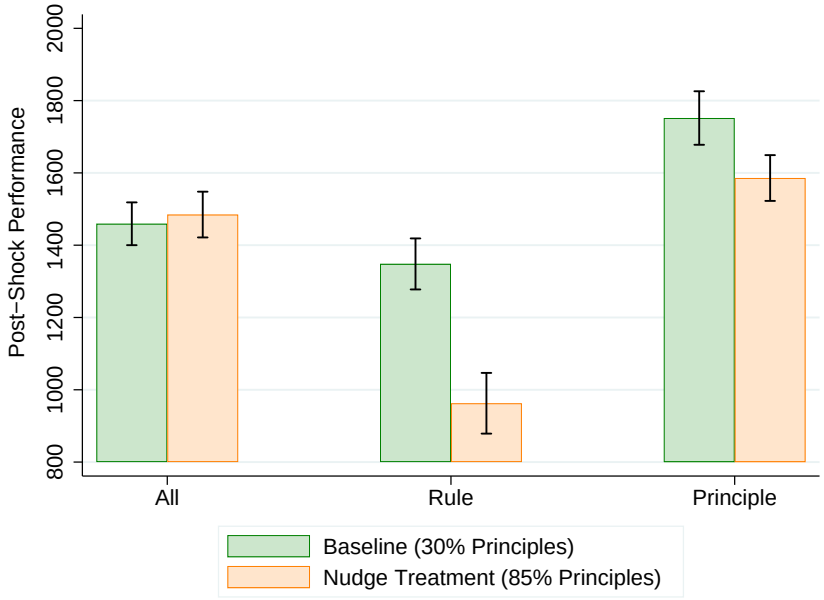
The left bars in Figure 7 provide the data for Result 4a, showing the average post-shock performance in the Baseline and the Nudge. The figure reveals that the Nudge Treatment increased post-shock performance only little ($t = 0.29$, $p = 0.77$). This finding contradicts Hypothesis 2a, and does not lend causal support to Result 1b, which described a positive correlation between endogenously emerging principles and post-shock performance in the Baseline.[22]

The middle and right bars show the post-shock performance separately for pairs with principle- versus pairs 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 in the Nudge treatment increased slightly, although pairs with principle- and rule-based agreements both had lower performance, is explained by the increase in the proportion of pairs with principle-based agreements (from 30% to 85%, as reported in the legend): principles did better than rules in both the Baseline and the Nudge, and there were more pairs with principles in the latter.

Within the Nudge treatment, the poor performance of the 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) is unexpected: it may indicate that at least some of the principles triggered by the Nudge may

[22] If we had observed the same difference in post-shock performance between Nudge and Baseline as between principle-based and rule-based pairs in the Baseline, an ex-post power analysis indicates that a total sample size of $N = 30$ pairs would have been sufficient to detect a statistically significant difference between the two conditions (at the 5% level and with a statistical power of 80%). However, given the small observed difference in overall post-shock performance between Baseline and Nudge, the sample size would have had to be very large to reach statistical significance (again at the 5% level and with a statistical power of 80%): $N = 10,958$ pairs.

Figure 7: **Post-Shock Performance in Baseline vs. Nudge Treatment**



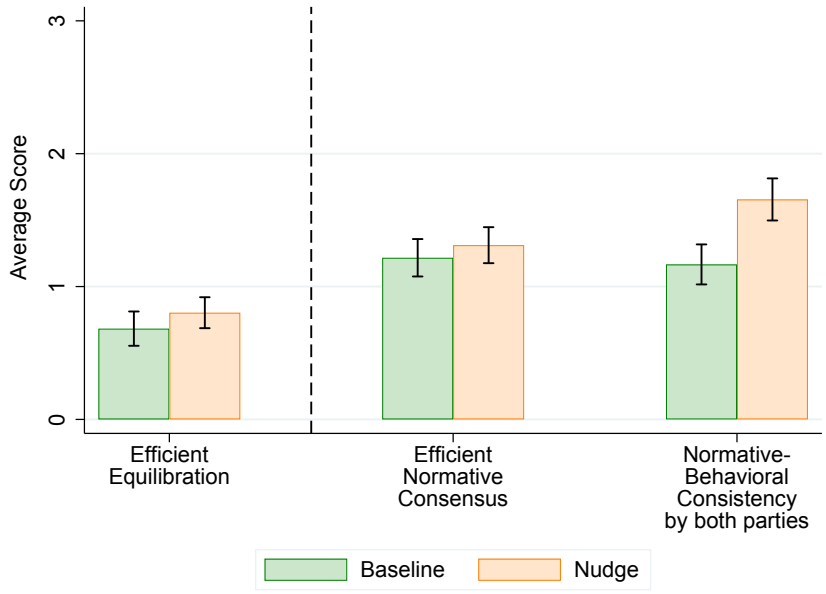
Notes: Data from the Baseline and the Nudge Treatment included. The bars represent average post-shock performance (average profits in Phase 2) in the Baseline and the Nudge Treatment, both at the overall treatment level and separately for pairs with principle-based agreements and pairs with rule-based 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.

have been less effective in inducing shared understanding, compared to the principles that emerged endogenously in the Baseline. We discuss this point in more detail in section 5.2.3.

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 post-shock performance in Figure 7, the left-hand side of Figure 8 indicates only a small 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 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. The Nudge caused only a small increase in efficient normative consensus ($t = 0.49$, $p = 0.63$), but a clear increase in the average number of scenarios in which both parties exhibited normative-behavioral consistency ($t = 2.24$, $p = 0.03$). Both these findings are in line with the corresponding comparison among pairs with principle-based and rule-based agreements within the Baseline (see Result 3b and Figure 6).

Figure 8: **Treatment Effect 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.

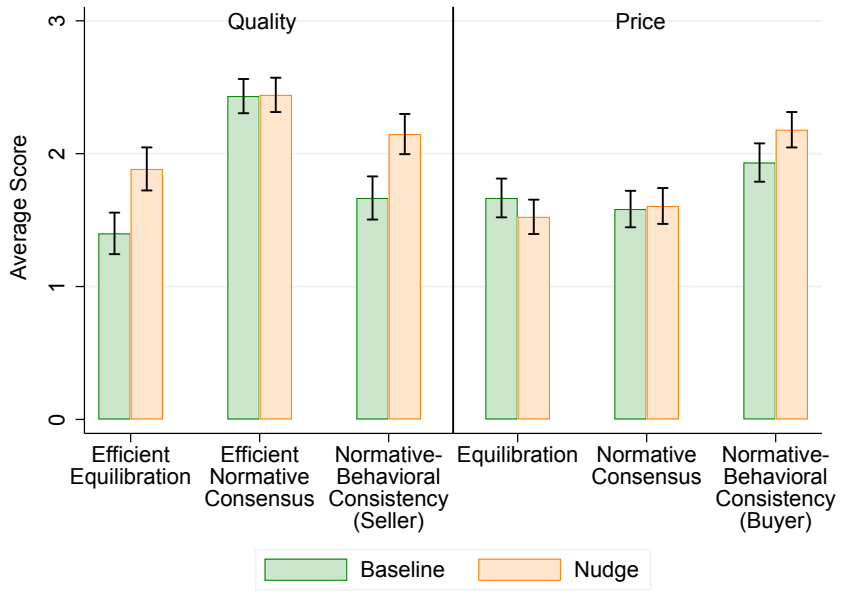
However, contrary to the within-Baseline comparison, there is no treatment effect for efficient equilibration when comparing Nudge to Baseline. We explore the reasons for this finding next.

5.2.2 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 make it more likely for pairs to achieve efficient equilibration on quality ($t = 2.16$, $p = 0.03$)—meaning that the seller provided the efficient quality level in the scenarios. Second, the Nudge also created an 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 thus no difference by treatment ($t = 0.05$, $p = 0.96$).

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.

Turning to the price reveals a very different story. The Nudge failed to increase equilibration on price. In fact, we even see a small decrease in equilibration on price ($t = 0.73$, $p = 0.47$)—the buyer was not more likely to pay a price as high as the seller expected. Furthermore, the Nudge produced only a small increase in the normative-behavioral consistency of the buyer ($t = 1.26$, $p = 0.21$)—the buyer was not 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 that the level of normative consensus on price is lower than that on quality.

Beyond the findings illustrated in Figure 9, we can say more about both quality and price. About quality, in addition to the positive treatment effect on efficient equilibration on quality, we also find a positive treatment effect on 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.

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

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

Notes: OLS regressions based on the quality decisions of 121 sellers in the 3 scenarios. Standard errors clustered at the seller (pair) level are in parentheses. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .10$

Column (1) shows that in the Nudge, when considering all 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.

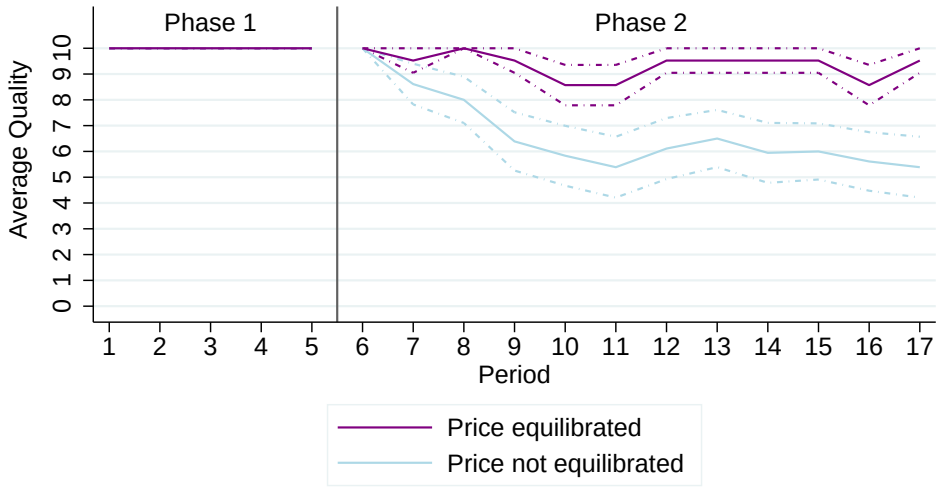
Turning to price, we illustrate the importance of achieving equilibration on *price* via the data in Figure 10 on average *quality* over all the periods in the experiment. The figure considers all pairs in the Nudge that achieved efficient equilibration on quality in the scenario that was later implemented 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 close to efficient quality level throughout (average quality was 9.4 over the 12 periods of Phase 2). In contrast, pairs who did not reach equilibration 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).

Why is reaching equilibration on price more challenging than reaching equilibration on quality? We see two reasons. First, whereas the surplus-maximizing quality is relatively easy to determine and remains constant after the shock, the payoff-equalizing price changes and is more complicated to compute. Second, and probably more importantly, cooperating pairs in a repeated game have a common interest in making the pie as large as possible so that conflicts about quality make little sense. When it comes to the determination of the price, in contrast, the two parties have competing interests, because the price determines how the pie is split between the two parties. Or put differently, while the efficient quality is defined objectively, the fair price needs to be determined

subjectively, which may considerably complicate coordination.

We therefore 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 lower post-shock performance in Phase 2 ($t = 2.90$, $p < .01$).

Figure 10: **Average Post-Shock Quality for Pairs Having Reached Efficient Equilibration on Quality in the Scenarios**



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 on price vs. pairs who did not. 39 pairs are included in this subsample, 21 of which achieved equilibration on price. The dashed lines represent plus/minus one standard error of the mean.

We summarize these exploratory findings in our next result:

Result 5

a) The Nudge increased efficient equilibration on quality and led to a marginally significantly higher quality in all three scenarios.

b) The Nudge failed to increase equilibration on price. Pairs who failed to reach equilibration on price experienced a deterioration of quality over time in Phase 2.

Importantly, including data from the Baseline in an analysis like Figure 10 leaves the outcome largely unchanged. The relation between equilibration on price and post-shock performance is statistically indistinguishable between Baseline and Nudge ($t = 0.71$, $p = 0.48$). Figures 9 and

10 thus illustrate an important feature of our findings: there is substantial variation among pairs within a condition in whether a pair achieved equilibration on price. This motivates our next analysis, of possible heterogeneity among principles.

5.2.3 Digging Deeper: Heterogeneity Among Principles?

Recall from Figure 7 that there were many more principle-based agreements in the Nudge than in the Baseline but that on average such agreements in the Nudge had lower post-shock performance. In each condition, rule-based agreements performed worse than principle-based agreements, so having fewer rules helped aggregate performance in the Nudge, but the decline in the performance of principles led aggregate performance to be only slightly 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 (section B).

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 meaningfully correlated with post-shock performance (see the Online Apendix 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

principle-based pairs did), pairs in the Nudge needed 38 seconds (23 percent) longer to get there, but this difference does not reach conventional levels of statistical significance ($t = 1.11$, $p = 0.27$).

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 post-shock performance in Phase 2. Having one party explain to the other was not strongly correlated with post-shock performance ($r = -0.08, p = 0.49$), but time to reach this agreement was ($r = -0.28, p = 0.02$).

Turning to the exit questionnaire, pairs with principle-based agreements in the Nudge were less likely to report that their agreement was "clearly formulated" ($t = 2.76$, $p < 0.01$) and somewhat 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 post-shock performance. For example, there is a positive correlation with the perceived clarity of the principle ($r = .35$, $p < .01$) and a negative correlation with differing interpretations of the principle ($r = -.72$, $p < .01$).

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 with principle-based agreements, but in only 5% of all Nudge pairs with such agreements ($t = 1.98$, $p = 0.05$). Our current analysis of this issue is necessarily exploratory. We propose possible future work in the next section.

6 Discussion

This paper studied how parties solve the clarity problem and build relational contracts that help them adapt after an unforeseen shock. In our baseline data, we found that parties basing their

agreements on general principles performed better than parties formulating specific rules. To assess causality, we stimulated principle-based agreements via a salience Nudge that succeeded in triggering more principles, but did not improve post-shock performance overall. Below we discuss the reasons for this absence of a performance effect of the Nudge in more detail. We conclude this discussion section by summarizing again the contributions of this paper and by pointing to limitations and areas for future research.

6.1 Causal Effects of the Nudge

In this subsection we discuss two potential reasons why our Nudge had no treatment effect on overall post-shock 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.

While omitted variables may certainly explain a substantial part of the correlation between principles and performance in the Baseline, we argue against omitted variables as the *sole* explanation for this finding, as we do observe treatment effects on some factors that contribute to performance, such as efficient equilibration on quality in the different scenarios, short-run performance immediately after the shock, 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.

Our findings thus suggest that principles can play *some* role in achieving clarity and efficient adaptation after a change in the environment. We therefore believe that a second interpretation also has merit: our treatment may not have been as strong as it initially appeared from the coding presented in Section 5.1.1 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 a rule may in fact have also agreed implicitly on a principle. As the Nudge treatment pushed participants to articulate general principles, some pairs may have made an implicit principle explicit. They would then be coded as having a principle, but stating that principle would have no impact on their performance.

Second, as discussed in Section 5.2.3, 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. Results from the exit questionnaire support this interpretation, as pairs with principle-based agreements in the Nudge treatment reported that their agreements were less clearly formulated and led to more misunderstandings than pairs with principle-based agreements in the Baseline.

In sum, while our findings 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 on overall post-shock performance is due to the Nudge being less effective than intended. As we discuss below, 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.

Perhaps one should not be surprised that our simple Nudge did not improve post-shock performance. 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 may be 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.

6.2 Contributions, Limitations, and Future Research

Our results help to better understand what it takes to build a relational contract and why parties often fail to reach an equilibrium. Neither existing work in organizational economics (OE) nor in organization theory (OT) addresses this point. OE takes it for granted that different repeated-game equilibria exist and that these equilibria are associated with differences in economic performance. However, the fact that equilibria exist in theory does not mean that parties will be able to reach them in practice. We provide empirical evidence documenting the difficulties that building an equilibrium may entail and we show that low performance may result not only from reaching a bad equilibrium but also from failing to reach an equilibrium at all.

Regarding OT we complement the large literature suggesting that differences in relational contracts within organizations, or as a part of relational governance in relationships between organizations, may arise from differences in pre-existing conditions. In particular, we show that parties with identical pre-existing conditions may reach different relational contracts because some may fail to overcome the clarity problem.

While we believe our experimental results provide novel insights about a 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 from laboratory experiments conducted with student participants. It would of course be desirable to conduct similar studies with different and more diverse subject pools, as well as in field settings. In addition, 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.

Turning from our methods to our results, as noted above, although our Nudge produced a number of causal effects in the short run immediately after the shock, it did not lead to an increase in overall post-shock performance. Our exploratory data analysis suggests several possible avenues for future research on this point.

First, we have seen that principles induced by the Nudge were less likely to produce a shared understanding between the parties than were principles in the Baseline. Future treatments could attempt to push subjects towards creating clarity in their agreements, to avoid situations in which contending interpretations cause deterioration in efficiency. For example, parties could be nudged to think in scenarios or to engage in role play (taking each other's perspective) when building their agreements. Also, because our manual text codings of chats and statements have provided some interesting first insights, it could be fruitful to analyze the content of parties' shared understandings more thoroughly by applying machine learning methods such as topic modeling to data on participants' communication.

Second, while our Nudge made Phase 2 of the experiment more salient, it failed to create equilibration on price—the distributional aspect of the relationship. It may be useful for a new Nudge to focus subjects on distributional issues that could arise after the environment changes. Rather than nudging participants to formulate principles, alternative treatments could attempt to stimulate social preferences (in line with the perspective of Frydlinger et al., 2019; Frydlinger &

Hart, 2019). In addition, the two parties in our setup were in a relatively equal situation, which makes equality a natural focal point. It would be interesting to conduct similar studies in an environment in which the focal point is different.

In addition to the specific next steps proposed above, we also hope to inspire future experiments on repeated games and game theory itself. For example, our results illustrate that it is of course not guaranteed that parties in a repeated game will 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 average behaviors (see, for example, Van Huyck et al., 1990; Goeree & Holt, 2001), but it is largely silent about differences across pairs within a fixed environment.

Finally, these observations about experimental work have analogs in game theory. Repeated-game models that assume the parties to be in equilibrium from the beginning of their relationship 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.

6.3 Conclusion

In this paper we began to explore how parties solve the clarity problem and build relational contracts that achieve efficient adaptation to unforeseen change. Our results suggest that agreements based on general principles can help solve the clarity problem and foster efficient adaptation after unforeseen change. 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 overcome the clarity problem and to achieve adaptation in unstable environments.

References

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*.

Cao, Z. & Lumineau, F. (2015). Revisiting the interplay between contractual and relational governance: A qualitative and meta-analytic investigation. *Journal of Operations Management*, *33*, 15–42.

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.

Doz, Y. L. (1996). The evolution of cooperation in strategic alliances: initial conditions or learning processes? *Strategic Management Journal*, *17*(S1), 55–83.

Eccles, R. G. (1981). The quasifirm in the construction industry. *Journal of Economic Behavior & Organization*, *2*(4), 335–357.

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.

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*.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage Publications.

Keller, A., Lumineau, F., Mellewigt, T., & Ariño, A. (2021). Alliance governance mechanisms in the face of disruption. *Organization Science*.

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

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.

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

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

Mayer, K. J. & Argyres, N. S. (2004). Learning to contract: Evidence from the personal computer industry. *Organization Science*, *15*(4), 394–410.

Oliver, C. (1990). Determinants of interorganizational relationships: Integration and future directions. *Academy of Management Review*, *15*(2), 241–265.

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.

Powell, W. W., Koput, K. W., & Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly*, 116–145.

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

Reuer, J. J. & Ariño, A. (2007). Strategic alliance contracts: Dimensions and determinants of contractual complexity. *Strategic Management Journal*, *28*(3), 313–330.

Schepker, D. J., Oh, W.-Y., Martynov, A., & Poppo, L. (2014). The many futures of contracts: Moving beyond structure and safeguarding to coordination and adaptation. *Journal of Management*, *40*(1), 193–225.

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.

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. *American Economic Review*, *80*(1), 234–248.

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

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

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