CESifo WORKING PAPERS

7871
2019

Original Version: September 2019
This Version: January 2021

Building an Equilibrium: Rules Versus Principles in Relational Contracts

Robert S. Gibbons, Manuel Grieder, Holger Herz, Christian Zehnder

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7871*
Category 13: Behavioural Economics

Building an Equilibrium: Rules Versus Principles in Relational Contracts

Abstract

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

JEL-Codes: D020, D230, L140.

Keywords: organization economics, adaptation, relational contracts.

Robert S. Gibbons
Sloan School of Management &
Department of Economics
MIT / Cambridge / MA / USA
rgibbons@mit.edu

Manuel Grieder
School of Management and Law
Zurich University of Applied Sciences (ZHAW),
Zurich / Switzerland
manuel.grieder@zhaw.ch

Holger Herz
University of Fribourg / Switzerland
holger.herz@unifr.ch

Christian Zehnder
University of Lausanne / Switzerland
christian.zehnder@unil.ch

January 4, 2021
We are very grateful for comments from Sylvain Chassang and Todd Zenger, as well as seminar participants at the Stanford CASBS (2016) workshop on organizations and their effectiveness, LMU Munich, Innsbruck, Paris School of Economics, INSEAD, SIOE (2019) in Stockholm, the Organization Science Special Issue Online Conference (2020), CEPR, the 6th Workshop on Relational Contracts in Moscow (2020), GATE Lyon-St-Etienne, and Harvard Business School. 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.

1 Introduction

In this paper we use experimental methods to study parties building "relational contracts." Following economists such as Kreps (1996), we characterize relational contracts as shared understandings of the parties' roles in and rewards from collaborating together (understandings so rooted in the details of the parties' relationship that they cannot be enforced by a court). In economic models, such relational contracts are equilibria of repeated games; that is, in our usage, relational contracts must be self-enforced, whereas formal contracts have standing in courts.

Such relational contracts are explicitly discussed in transaction-cost economics (e.g., Williamson, 1979), and we see connections with other strands of organization theory. For example, in Cyert & March's (1963) quasi-resolution of conflict, such equilibria may help address "the obvious potential for internal goal conflict inherent in a coalition of diverse individuals and groups" (1963: 31). More generally, we see such equilibria as part of informal organization (e.g., Blau & Scott, 1962: 6).

Parallel to these illustrations within organizations, we also see relational contracts between organizations (often in combination with formal contracts, as we discuss below). For example, between two entities, we see such equilibria in the supply chains of Dore (1983), the "hybrids" of Williamson (1991), and the alliances of Oliver (1990). And among more than two entities, we see such equilibria in the quasifirms of Eccles (1981), the joint ventures of Kogut (1989), and the networks of Powell (1990).

In addition to the *existence* of relational contracts within and between organizations, the large literatures on the topics above also document *heterogeneity*: in some instances, the parties benefit from a cooperative equilibrium, yet in very similar instances, other parties seem stuck in a bad equilibrium. We do not imagine parties choosing to be in a bad equilibrium, so we explore the difficulties of building a good equilibrium (and the pathways of stumbling into a bad one). More specifically, while there are many persuasive reasons why pre-existing conditions—such as historical exchange patterns (Gulati, 1995), network structure (Powell et al., 1996), or embeddedness (Granovetter, 1985)—may help parties achieve cooperative outcomes, we ask an orthogonal question: given equivalent pre-existing conditions (favorable or otherwise), why do some parties reach a cooperative equilibrium but others not?

As one way to address this question, we follow Gibbons & Henderson (2012) by identifying two problems that relational contracts must solve. First, there is the credibility problem—"Should I believe the promise you are making me?"—that has been extensively studied (see below). Second, and at least as important, there is the clarity problem—"Do we have a shared understanding of the promise you are making me?" This clarity problem disappears in formal models of repeated-game equilibria: the parties are assumed to possess a shared understanding of which actions (under which circumstances) constitute "cooperation," which "defection," and which are available as "punishment" if cooperation fails.[1] Obviously, in some settings, this is a heroic assumption.

Our experimental approach allows us to explore how parties might build a shared understanding about how adaptation should later occur if their environment changes. In particular, we create unforeseen events that require adaptation by the parties if they are to achieve consummate performance. Before such events occur, we ask the parties about their beliefs—concerning what should and what would happen following a particular event. And after such events occur, we observe the parties' actual behaviors. These features of our experiment allow us to assess not only whether an equilibrium exists but also, if one does not, why not.

In summary, our experiment explores how parties try to build a cooperative equilibrium, given fixed initial conditions. We utilize forward-looking considerations from game theory, holding constant backward-looking considerations from sociology and organization theory, to explore the heterogeneity of outcomes under fixed initial conditions; our approach thus complements large literatures asking how expected outcomes vary with different initial conditions. To connect our work to these and other existing literatures, we first provide an overview of our experiment.

1.1 Overview of Experiment

We designed a laboratory experiment in the manner of experimental economics, studying 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

[1] More precisely, in a Nash equilibrium, the parties have a shared understanding of each other's strategies.

non-binding agreement on how to play the entire game and to formulate a written statement of their 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 would be appropriate behaviors if their environment changed 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 for the coming periods and asked them to indicate for each scenario (a) what choices their agreement prescribes and (b) what choices they expect the parties to actually make. Subsequently, we implemented one of the scenarios as the game that buyer-seller pairs played for the remaining periods.

The data from the different scenarios allow us to measure a sequence of steps towards an equilibrium. In particular, we measure (1) the extent to which parties agree on what ought to be done in a particular situation according to the agreement (normative consensus), (2) whether each party is willing to do what ought to be done (normative-behavioral consistency), and (3) whether the actions that each party will take accord with the other party's expectation (equilibration). Although theoretical work on games has been completely clear about what an equilibrium *is*, our novel measures (1) and (2)—normative consensus and normative-behavioral consistency—help us assess the extent to which parties are *not* in equilibrium.

Our first set of results explores the endogenous non-binding agreements in this Baseline setting. More specifically, we investigate how different types of initial agreements correlate with subsequent performance, defined as the sum of profits earned by a buyer-seller pair in the repeated game after their environment changes.[2] We find that pairs whose initial agreement articulated a broad principle (applicable to a range of possible environments) tended to achieve higher levels of subsequent performance than did pairs whose agreement articulated only a narrow rule (specific to the original environment they had been shown). Further analysis reveals the mechanism underlying this correlation between initial agreement and subsequent performance: parties with principle-based agreements were <u>not</u> more likely to agree on what should be done according to their agreement

[2]For two parties within an organization, our definition of performance is the parties' contribution to the performance of the organization as a whole. For two parties in different organizations (such as an alliance), our definition is equivalent to the aggregate exchange value of the relationship. In both cases, our definition of performance is total economic surplus.

(normative consensus), but such parties were 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 were more likely than those with rule-based agreements to take actions consistent with their partner's expectations (equilibriation).*

We then investigated whether there is a causal link between principle-based agreements and subsequent performance. To this end, we implemented a "Nudge Treatment" designed to exogenously foster the emergence of principle-based agreements. Our Nudge indeed had a large causal impact in inducing more pairs to articulate principles in their initial agreements: whereas only a minority of pairs did so in the Baseline condition, a large majority did so after the Nudge.

Importantly, however, the Nudge failed to increase subsequent performance. More specifically, the Nudge *did* have a causal impact on the number of pairs coordinating on efficient initial quality after their environment changed, but it did *not* lead to more coordination on the initial price. Our particular Nudge was thus not successful enough in helping the parties avoid conflicts about the price that then led to a deterioration of performance in later periods. As a potential explanation for this observation, we find that some of the principle-based agreements in the Nudge were less clear and less often created a shared understanding than those in the Baseline. In short, we conjecture that some of the additional principles caused by the Nudge were of low caliber (although, to repeat, the Nudge did have a causal impact on the number of pairs coordinating on the efficient quality after their environment changed).

We see our findings from the Baseline and the Nudge as consistent with the view that relational contracts are hard to build—certainly in the world and even in the lab, especially in settings where an effective agreement must coordinate the parties' adaptation after their environment changes. Put differently, in the Baseline a minority of pairs were able to reach principle-based agreements that 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—even though our lab environment is enormously simpler than a real-world setting. More specifically, the Nudge induced more pairs to state that they had reached agreement on a principle, but many of these had not built an equilibrium: after the environment changed, at least one party departed from what the other party thought had been agreed.

1.2 Related Work

As suggested by our opening paragraphs, we build on work that began decades ago and has flourished since, so we cannot survey it here. Instead, we describe work that relates closely to ours but offer only pointers to wider literatures.

Since our notion of relational contracts has application both within and between organizations, we treat these domains in turn. As analyses of the initiation, clarification, and change of equilibria *within* organizations, research that motivates ours includes Foss's (2003) discussion of Oticon's "spaghetti organization" and Turco's (2016) "conversational firm." Notably, both Foss and Turco reference the relational-contract model of empowerment by Baker et al. (1999), which envisions empowerment as an equilibrium in a repeated game: below the top of an organization, decision rights are "loaned, not owned" (1999: 56). Our work is also motivated by Kellogg's (2009) study of how "relational spaces" affect microinstitutional change; this work does not explicitly consider equilibria of repeated games, but we interpret it in a related spirit.

Foss used detailed sources from the academic literature, the business press, and the firm itself, and Turco and Kellogg conducted ethnographic studies. While we sacrifice the rich descriptions of such work, the experimental method offers familiar benefits as a complementary approach. In our Baseline condition, we can characterize the parties' initial agreement, measure their beliefs about both appropriate and anticipated responses to different scenarios, compute the adaptation that would be efficient in their new environment and compare it to the responses actually chosen, and assess the extent to which the parties were in equilibrium after their environment changed. And by comparing the Baseline to our Nudge condition, we can assess the causal impact of our particular intervention.

Turning to analyses of the initiation, clarification, and change of relational contracts *between* organizations, we focus on alliances as a leading case. Of course, most alliance partners negotiate a formal contract to launch their relationship, even if they anticipate also building an equilibrium while they live their relationship. Our experimental setting has no formal contracts, so we begin by locating our approach within the following framework articulated by Das & Teng (1998), who:

- emphasize the "fickle and tentative nature of partner cooperation," defining such cooperation as "the willingness of a partner firm to pursue mutually compatible interests ... rather than act opportunistically" (p. 492)

- define *confidence* in partner cooperation as a firm's "perceived certainty about satisfactory partner cooperation," noting that this concept "is about a firm's expectation about its partner's behavior only" (p. 492)

- propose *trust* and *control* as "two distinct sources of confidence" (p. 508)—whereas "trust relates to expectations about the motives of the trustee" (p. 494), control mechanisms take motives as given, and

- distinguish *formal control* (via formal contracts and the like) from *social control*, stating that "Unlike in formal control, the central element of social control is organizational culture, ... [which] unifies the way organizational members process information and react to the environment ..." (p. 507).

We find this framework very helpful. In Das & Teng's terms, our relational contracts (being self-enforcing equilibria of repeated games, not formal contracts) are an attempt to build confidence through social control. In addition, we see our experimental design as highly aligned with their interest in whether parties have built a unified way to process information and react to the environment: we study whether principle-based agreements are more likely than rule-based agreements to achieve exactly this.[3]

Economists are not alone in studying relational contracts as equilibria in repeated games. As noted above, in analyses of interactions within organizations, Foss and Turco both reference a repeated-game model of empowerment. Likewise, concerning interactions both within organizations and between them, Granovetter (1985: 490) articulates the same repeated-game logic. And in a synthesis of research on alliances, Schepker et al. (2014: 202) use "relational contract" in this way.

This shared definition of relational contracts notwithstanding, the focus of our research is narrower than the full range of factors that these authors (and Das & Teng) consider. In particular, our focus on self-enforcing agreements in repeated games is narrower than a definition of *relational*

[3]To foreshadow a discussion that is beyond the scope of this paper, we note that Das & Teng's definition of organizational culture not only relates closely to this paper (by emphasizing a unified way to process information and react to the environment) but also, to us, echoes two earlier definitions: (1) Schein's (1985: 9) definition of organizational culture as "a pattern of basic assumptions—invented, discovered or developed by a given group ... " and (2) Geertz's (1973: 12) observation about culture more generally, that "Culture is public because meaning is." In short, we see the path-dependence of shared interpretation within a given group as a key aspect of organizational culture—see Gibbons & Prusak (2020) for more on this—and we foresee a rich interaction between this view and Kreps' (1990) construal of repeated-game equilibria as part of corporate culture ("how we do things around here"). The present paper is a first step.

governance that also includes concepts such as trust and norms (Zaheer & Venkatraman, 1995; Poppo & Zenger, 2002). In this sense, our use of "relational contracts" also differs from the term's original uses in sociology and law, such as Macaulay (1963) and Macneil (1978).

Of course, real alliances may involve both trust and formal control, whereas our experiment manipulates neither. Instead, our work studies social control, especially the question of whether the parties have developed a unified way to react to changes in their environment. Research on parties' behaviors after an "alliance disruption" is therefore close to our work. For example, Lumineau & Malhotra (2011) and Malhotra & Lumineau (2011) use legal files to analyze fine-grained data on parties' behaviors during and after contract disputes, studying how the relationship's formal contract influences these behaviors during and after the dispute.

Just as pre-existing conditions—such as historical exchange patterns—may help parties achieve cooperative outcomes, so too might particular formal contracts influence both the incidence and the consequences of disputes. As noted above, we ask an orthogonal question: given equivalent pre-existing conditions (or, here, equivalent formal contracts), why do some parties reach a cooperative equilibrium but others not? We thus complement research on disputes that have in fact occurred: given a fixed formal contract, our approach can be interpreted as studying which relationships are likely to experience a dispute in the first place.

Finally, longitudinal studies of particular alliances are also close to our work. For example, Doz (1996), Mayer & Argyres (2004), and Keller et al. (2021) provide rich accounts of the incubation, arrival, remediation, and consequences of "alliance disruptions" in specific alliances. The issues we study certainly can arise in such settings. For example, Mayer & Argyres (2004: 400) describe a contract clause that was added during the alliance they study, stipulating that schedule and pricing could be renegotiated if particular events occurred, but the clause "was not particularly successful because it did not stipulate any guidelines for providing a new schedule or price." Similar to the way our experimental approach complements the rich descriptions by Foss, Turco, and Kellogg described above, so too does the experimental approach complement these longitudinal studies of the evolution of cooperation in particular alliances. For example, we can characterize the parties' initial agreement, measure their beliefs about both appropriate and anticipated responses to different scenarios, and so on; and we can assess the causal impact of our particular intervention.[4]

[4]Other recent work on contracting—such as Frydlinger et al. (2019) and Grandori & Furlotti (2019)—emphasizes

Turning to related work in economics, there is a large theoretical literature on relational contracting (see Malcomson, 2013 for a survey), but such models ignore the challenges that real parties face when creating a shared understanding 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 creating shared understandings.

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).

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

2 Study Design

To study how parties build relational contracts and how they adapt after shocks, we created a laboratory experiment in which a buyer (she) and a seller (he), engaged in a repeated trading

a style of formal contracting that benefits greatly from both relationship-building processes before the contract is signed and relationship-managing processes while it is lived.

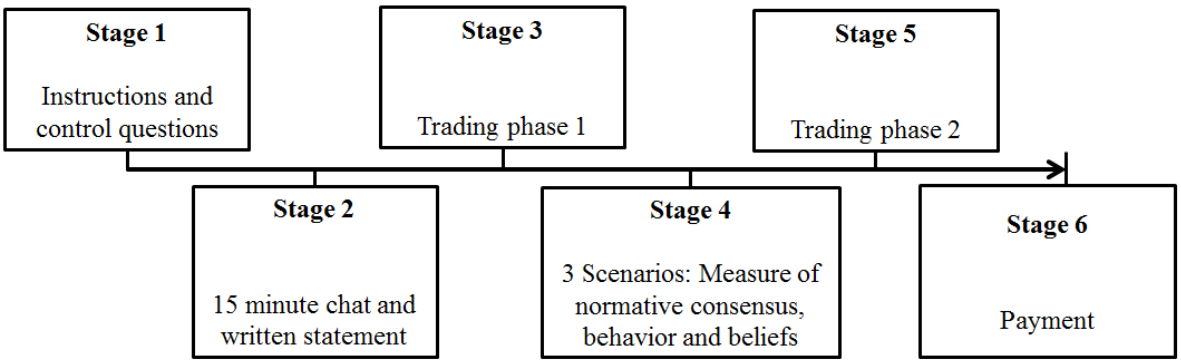
game with uncertainty about the future. Participants knew that, after a fixed period of time, the parameters of the game would change in an unforeseeable way. To assess the role of communication for building an equilibrium, before trading interactions started we gave participants the possibility to communicate via text chat in order to reach a non-binding agreement on how to play the game and to formulate a written statement capturing this agreement. After the first phase of trading interactions and before the game changed, we measured each group's shared understanding of how to behave in new environments. For this purpose we confronted participants with three possible scenarios of how the game might change. We asked them to indicate what they should do according to their initial agreement, and what they would actually do. Subsequently, one of these scenarios was implemented for the remaining interaction periods. This setup allowed us to observe what types of initial agreements endogenously emerged and how these agreements were associated with groups' shared understanding and performance after change occurred.

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

2.1 Baseline Experimental Set-up

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

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



Stage 1: Instructions and Control Questions. The instructions informed all participants of their role and of the game parameters in the first (five-period) trading phase (the full instructions are reproduced in Appendix A). Participants were made aware of the fact that the game parameters and structure would change after 5 periods, but no information was provided about the nature of the change. Thus, whereas participants were aware that something would change, they could not reasonably foresee what form that change would take exactly. Participants also knew that they would continue playing with the same partner as in the first phase and that the second phase would be an infinitely repeated game with a stopping probability of 10% after each round. At the end of the instructions, participants had to correctly answer a set of control questions to ensure understanding of the experimental set-up before they could enter the next stage.

Stage 2: Chat Communication and Statement. Participants could exchange free-form text messages with their interaction partner via computer for 15 minutes. Chats were anonymous and participants were instructed not to reveal their identities. We told participants to use the chat function to find a non-binding agreement on how they intended to play the game and to write this down in a short joint statement. This statement was drafted by the seller in an entry window that was separate from the chat function. The seller could transmit the draft of the statement to the buyer, who could demand changes in the chat window. If the buyer agreed with the statement, she could confirm the statement with a click on a button. Once a statement was confirmed, the communication phase ended. If no statement was confirmed within 15 minutes the groups continued without a statement.[5]

Stage 3: First Trading Phase. All groups played five rounds of a simultaneous buyer-seller game in which the seller decided about the quality q of a good to be delivered and the buyer determined the price p.[6] The seller chose a quality $q \in [0, 10]$ at cost $c(q)$, where higher quality was associated with higher cost ($c'(q) > 0, c''(q) > 0$). Table 1 displays the cost function $c(q)$.[7] The

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

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

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

value of the product to the buyer in the first trading phase was $v = 10q$ and the buyer picked a price $p \in [0, 100]$. In any period, the buyer's payoff was thus $\pi_B = 10q - p$ and the seller's payoff was $\pi_S = p - c(q)$. All parameters of the game were common knowledge.[8]

Table 1: **Cost Function**

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

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

For each scenario, we asked participants a set of questions aimed at measuring the *normative* and the *behavioral* implications of the pair's initial agreement. Specifically, to elicit the normative implications of the agreement, we asked participants for each scenario 1.a) "what should you do according to the agreement?" and 1.b) "what should your partner do according to the agreement?". We incentivized these two questions (1.a and 1.b) by paying participants a bonus of 20 experimental points for consensus with their partner (for each question in each scenario). For the questions targeting behavioral implications, i.e., actual choices, we asked 2.a) "what will you do in the first period?" and, to elicit the belief about the expected choice of the partner, 2.b) "what do you think that your partner will do in the first period?". Question 2.a) was incentivized by making the answer to that question the participant's binding first period choice in case that scenario was implemented. We decided not to incentivize question 2.b).

[8]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 ($v' = 10$).

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

Increasing the value of the product to the buyer means that the payoff equalizing price was not the same anymore as in the first phase.[9] Adding the outside option means that the situation became riskier for the seller if he was unsure about whether the buyer would choose the outside option or whether she would choose to buy from him. For the buyer, the outside option was attractive (compared to buying from the seller) because there was no uncertainty about q^o and p^o, whereas when buying from the seller, uncertainty remained about the quality q the seller decided to produce. These changes represent a tough test for the stability of relational contracts, because the sellers had to be quite certain about the buyers' actions in order to be willing to incur the cost of production in case an attractive outside option was present for the buyers.

Stage 6: Payments. Once the second trading phase was finished, participants were provided with a summary of their earnings in the experiment and received their payments, corresponding to the profits made in the experiment, in cash and in private.

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

2.2 The Nudge Treatment

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

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

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

2.3 Participants and Procedure

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

We conducted seven experimental sessions with a minimum of 32 and a maximum of 36 participants per session. Participants were randomly assigned to a computer cubicle upon arrival and were randomly assigned the role of buyer or seller. Before each trading phase, participants received detailed written instructions and experimenters read aloud a summary of the instructions before the start of each phase. Interactions were anonymous, i.e., buyers and sellers in a pair could not identify each other, but they knew that their interaction partner was another participant present in

the room during the same experimental session. The experiment was programmed in z-Tree (Fischbacher, 2007). The experimental manipulation was randomly assigned within each experimental session; there were 60 groups in the Baseline and 61 in the Nudge Treatment.

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

2.4 Coding of Joint Statements and Chats

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

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

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

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

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

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

3 Predictions

3.1 Framework and Terminology

Our study investigates how parties build relational contracts that allow them to adapt to unforeseen change in an efficient manner. To this end, we study a repeatedly interacting buyer-seller pair that experiences an exogenous change in their trading environment. The pair starts their relationship in a setting with known parameters and they are aware that there will be a need for adaptation, but the nature of the upcoming change is unknown to them.

Our buyer-seller pairs faced a game in which efficient cooperation could be sustained as an equilibrium of the infinitely repeated game.[10] From a standard game-theoretic perspective one would therefore expect that the parties not only cooperate from the outset but also immediately adapt their strategies to keep cooperating after the shock.[11] 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.

[10] 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.

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

They argue that successful adaptation requires not only that collaborative equilibria exist (the "credibility problem"), but also that the partners succeed in building a shared understanding of the equilibrium strategies (the "clarity problem").

Because the post-shock phase of our experiment is a repeated game (not a one-shot interaction), we believe that successful adaptation requires that the parties align their beliefs and actions about both the quality to be provided and the price to be paid after the shock.[12] We use the term *equilibrated* to describe a buyer-seller pair that coordinates on a new price-quality combination that simultaneously meets the expectations of both parties. Specifically, a pair is equilibrated if the buyer's chosen price (p) is equal or larger than the seller's price expectation ($\hat{p}_S$) and the seller's chosen quality (q) is equal or larger than the buyer's quality expectation ($\hat{q}_B$). Intuitively, a pair is therefore equilibrated if neither party takes an action that disappoints the partner.

We would like to point out that it is possible to equilibrate on an inefficient outcome. If a pair is not only equilibrated but also coordinates on actions that maximize the total surplus (the sum of profits in the pair, i.e., our measure of performance), we use the term *efficiently equilibrated*. In our experiment efficiency is entirely determined by the seller's quality choice. Thus, an equilibrated pair is efficiently equilibrated if the seller chooses a quality of 10.

Our first goal is to understand the process that produces equilibration after change occurs. Put differently, we are interested in how relational contracts that facilitate adaptation are built. All the game-theoretic models of cooperation that we know are silent about this issue.[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 equilibriation 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

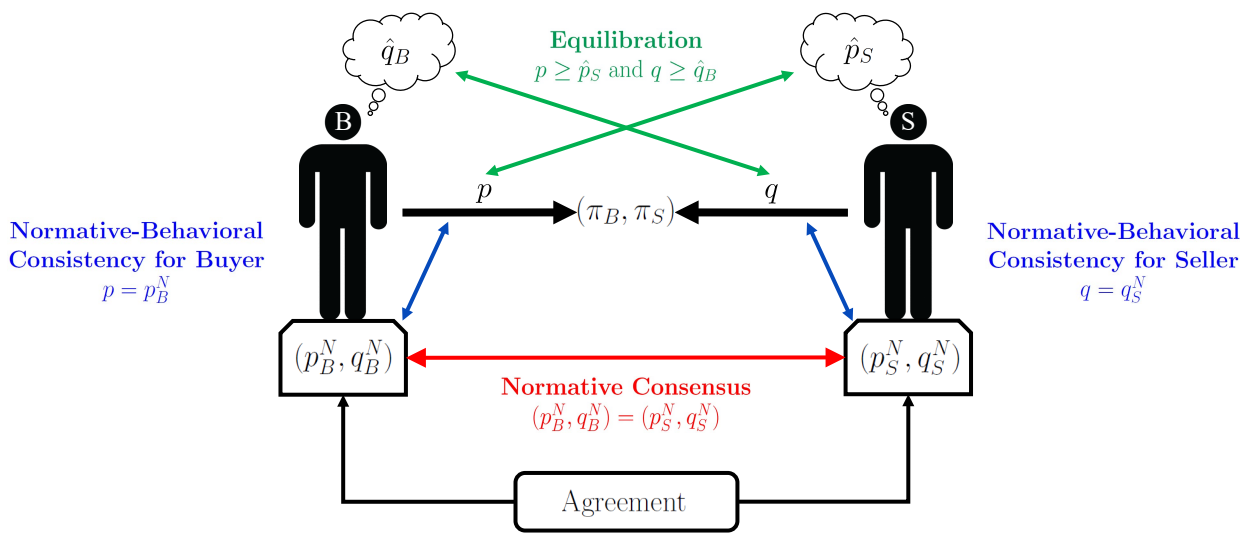
[12] That is, in a one-shot interaction, it *might* be possible for parties to stumble into high performance without having their beliefs and actions aligned, but abundant theoretical, experimental, and real-life evidence about repeated games strongly suggests that collaboration withers once parties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

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

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 simply means that each party does what he or she thinks the agreement implies. See Figure 2 for a summary of this terminology.

Figure 2: **Framework and Terminology**



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

We note that, strictly speaking, efficient normative consensus and normative-behavioral consistency are neither sufficient nor necessary conditions for pairs to be equilibrated.[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) be efficiently equilibrated after a change and (b) reach efficient normative consensus and exhibit normative-behavioral consistency.

3.2 Testable Hypotheses

Our first hypothesis describes our predictions regarding the expected 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 be efficiently equilibrated 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 summarizes this chain of arguments:

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

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

b) ***Mechanisms:*** *Buyer-seller pairs with principle-based (rather than rule-based) agreements are more likely to exhibit both normative consensus as a pair and normative-behavioral consistency*

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

as individuals, and such pairs are more likely to be efficiently equilibrated after an exogenous change in the environment.

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

Hypothesis 2 (Causal Effects of the Nudge Treatment)

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

b) ***Mechanisms:*** *Buyer-seller pairs in the Nudge Treatment are more likely to exhibit both normative consensus as a pair and normative-behavioral consistency for both individuals, and they are more likely to be efficiently equilibrated after an exogenous change in the environment than pairs in the Baseline setting.*

4 Results

In the first part of this section we present analyses of the effects of endogenously emerged principles in the Baseline condition (Hypothesis 1). In the second part we analyze the results of the Nudge Treatment (Hypothesis 2).

4.1 Baseline Outcomes

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

We then explore the mechanisms underlying the observed correlation between initial agreement and performance after the shock in two steps. First, we consider the association between equilibration and performance (and the association between principle-based agreements and equilibration).

Second, we consider the association between efficient normative consensus and normative-behavioral consistency, on the one hand, and equilibration, on the other (as well as the association between principle-based agreements and the consensus and consistency concepts).

As there was no exogenous variation in the Baseline, agreements were established endogenously and all results in this subsection are purely correlational.

4.1.1 Rules vs. Principles: Emergence and Performance

Our first result establishes the relative frequency with which different types of initial agreements emerged and how agreement type correlates with the post-shock performance:

Result 1

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

b) Pairs who formulated principle-based agreements tended to achieve significantly higher post-shock performance than those who relied on rule-based agreements.

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

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

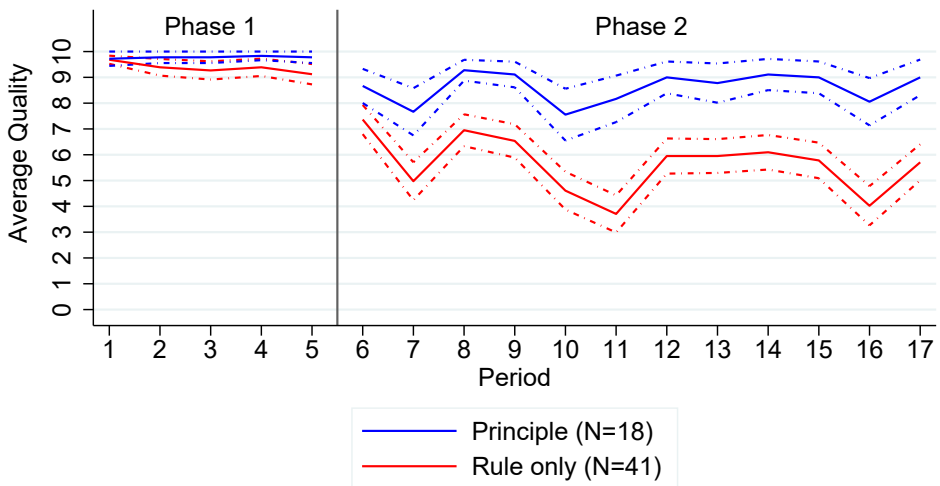
	Statements	*Chats*	*Combined*
Principle	.133	.250	.300
Rule	.883	.950	.983
N		60	

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 and chat messages.

We consider it likely that both the statements and the chats could have an impact on the initial agreement. Consequently, in our analyses we rely on the combined coding from both statements

and chats. This means that, in what follows, we say that a pair had a principle whenever a principle was agreed upon in the statement, the chat, or in both (see the Online Appendix[15] for analyses of principles that were coded as present in the chats only versus in the statements). When considering the combined data, we observe that 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 combination with a principle, in contrast, the rule specifies the implication of the principle for the current setting in Phase 1 (which was known at the time that the parties could communicate). We therefore think that the latter case should 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.

[15] https://osf.io/ne7rd/?view_only=ee19b21676cc4366aed8c0729dae141b

Figure 3 illustrates the second part of Result 1. The Figure shows the average quality provided by sellers over time and distinguishes between pairs with an initial agreement that relied on a rule alone versus pairs who also formulated a principle. Note that in our setting the seller's quality choice fully determines overall performance of the relationship (i.e., joint profits). The figure reveals that pairs with principle-based agreements clearly outperformed pairs that relied solely on rules. Indeed, and in line with Hypothesis 1a, when using joint payoffs in Phase 2 as an overall measure of performance, we find that pairs with principles had average earnings of 1,752 points, whereas pairs with rules earned 1,348 points on average. This difference is statistically significant ($t = 3.95$, $p < .01$).[16] In contrast, when comparing average earnings in Phase 1, we find no significant difference ($t = 1.21$, $p = .23$).[17]

4.1.2 The Role of Efficient Equilibration for Long-Run Performance

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

In Section 3, we hypothesized that a fruitful buyer-seller collaboration after the shock is facilitated if beliefs and actions about chosen quality and price are equilibrated on an efficient outcome.[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 was efficiently equilibrated. Recall that decisions in these scenarios were incentivized by making the answer the first-period choice in Phase 2 if that scenario was implemented. We then explore whether efficient equilibration in the scenarios is correlated with high performance throughout Phase 2 (i.e., during the full repeated game after the shock) and whether

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

[17] To cleanly 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 significantly larger profits in Phase 2 among those pairs with principle-based agreements ($t = 3.4$, $p < 0.01$). 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.

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

pairs with principle-based initial agreements were more likely to achieve efficient equilibration in the scenarios. The findings are summarized in our second result:

Result 2

a) Pairs that were efficiently equilibrated in the scenarios reached significantly higher performance levels in the subsequent trading phase 2.

b) Pairs with principle-based initial agreements were significantly more likely to be efficiently equilibrated in the scenarios.

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

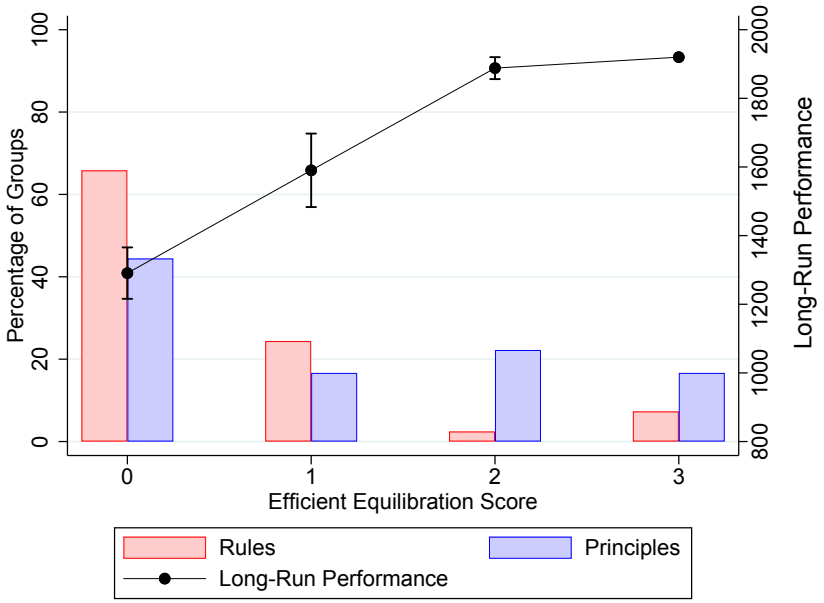
Of the three scenarios discussed before Phase 2 began, one was implemented as the trading game in Phase 2, so a more conservative approach is to study the correlation between long-run performance and the Efficient Equilibration Score based on only the two non-implemented scenarios. The result is essentially unchanged. The correlation then is $r = 0.46$ ($p < .01$) between long-run performance and efficient equilibration.

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 pairs who relied on rules only. The average Efficient Equilibration Score is more than twice as large for pairs with principle-based agreements (1.11 vs. 0.51, $t = 1.93$, $p = 0.06$).

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

Result 1 established that pairs with principle-based initial agreements had better post-shock performance than did pairs relying on rules alone, and Result 2 established that post-shock performance is correlated with efficient equilibration in the scenarios. As discussed in Section 3, two concepts seem important in achieving such equilibration. First, trading partners need to have a shared understanding of *what should be done* to reach an efficient outcome in a particular situation (*efficient*

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



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

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).[19] Note that all the measures in this analysis are drawn from the parties' responses to the scenarios, before the shock occurs.

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

[19] Again, section 3 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.

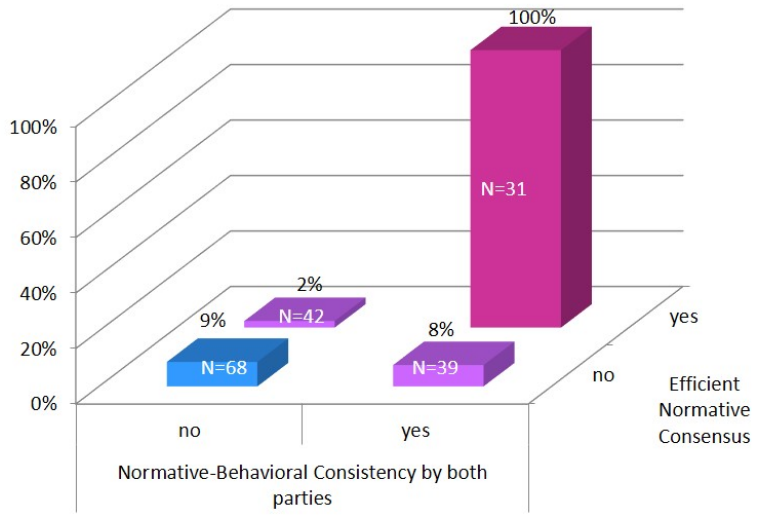
Result 3

a) *Efficient equilibration is unlikely without the joint presence of efficient normative consensus and normative-behavioral consistency by both parties.*

b) *Pairs with principle-based agreements are not significantly more likely to develop efficient normative consensus, but they do display significantly stronger normative-behavioral consistency (for both parties).*

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

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.

Evidence for Result 3b stems from Figure 6. The figure shows in how many scenarios, on average, pairs with principle- versus pairs with rule-based initial agreements were (a) efficiently equilibrated; (b) reached 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 are more likely to be efficiently equilibrated (see also Result 2b and Figure 4). New data is presented to the right of the dashed line. The bars in the middle of the figure show that pairs with principle-based agreements reached, on average, a 34% higher score for efficient normative consensus, but this effect is not statistically significant ($t = 1.19$, $p = 0.24$). 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, a highly significant effect ($t = 2.71$, $p = 0.01$). This higher degree of normative-behavioral consistency (i.e., actually doing what a party thinks should be done) by both parties in pairs with principle-based agreements is strongly correlated with high performance in our Baseline data ($r = 0.44$, $p < 0.01$).

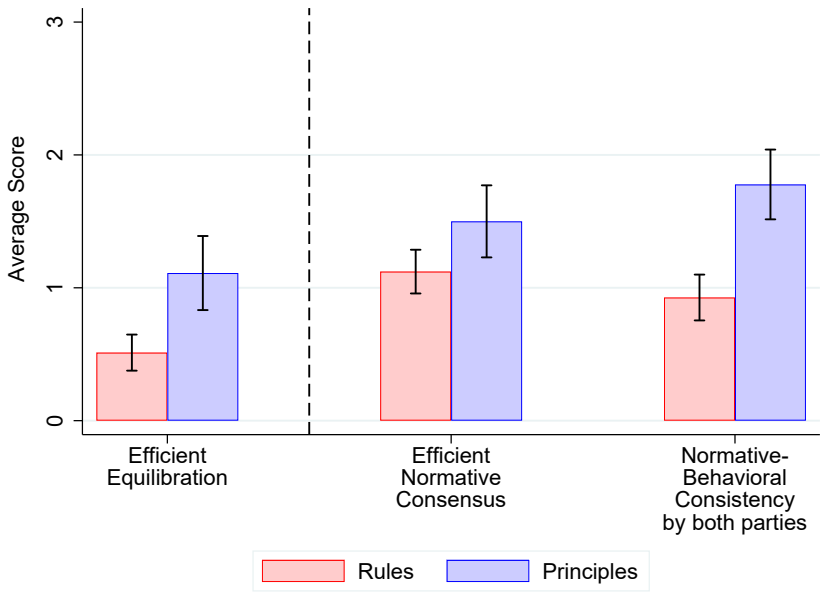
4.2 Experimental Manipulation: Nudging the Emergence of Principles

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

We now analyze whether our Nudge Treatment can provide causal evidence about any of these effects. Recall that the Nudge manipulation consisted of only a short reminder in the experimental instructions, making more salient the fact that the game was going to change in the second trading phase and telling participants it could be beneficial to take this into account when communicating, for instance by trying to include a general principle about how to act in their written statement.

Subsection 4.2.1 parallels the analyses from subsection 4.1. While we find that the Nudge significantly increased the emergence of principle-based agreements, we then find a treatment effect only on normative-behavioral consistency; we find no statistically significant increase in long-run performance, efficient equilibration, or efficient normative consensus.

Figure 6: **The 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.

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

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

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

Before discussing the impact of the Nudge on the outcomes of interest, we first report a manipulation check to establish that the treatment had the intended effect on observed agreements: inducing more pairs to establish principle-based initial agreements. We find that this is the case. Coders, who were blind to treatment, found far more principles in the Nudge (in 52 of 61 or 85.2% of pairs) than in the Baseline (in 18 of 60 or 30.0% of pairs). This difference is highly significant ($z = 6.15$, $p < .01$). As in the Baseline, all pairs who agreed on a 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.

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

Result 4

a) The Nudge did not significantly increase long-run performance.

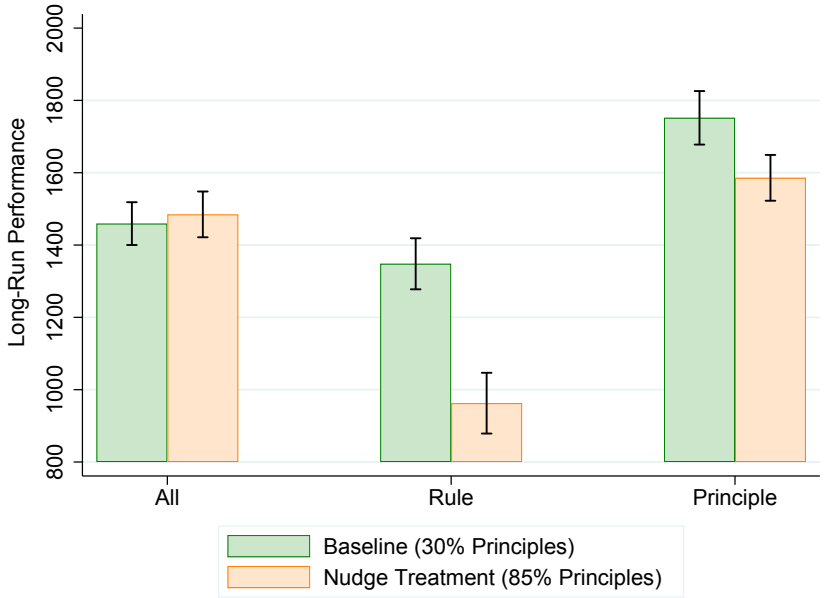
b) The Nudge did not significantly increase efficient equilibration.

c) The Nudge did not significantly increase efficient normative consensus, but it did significantly increase the average number of scenarios where both parties exhibited normative-behavioral consistency.

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

The middle and right bars show the long-run performance separately for pairs with principle- versus pairs with rule-based agreements. Both pairs with rule-based agreements and pairs with

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

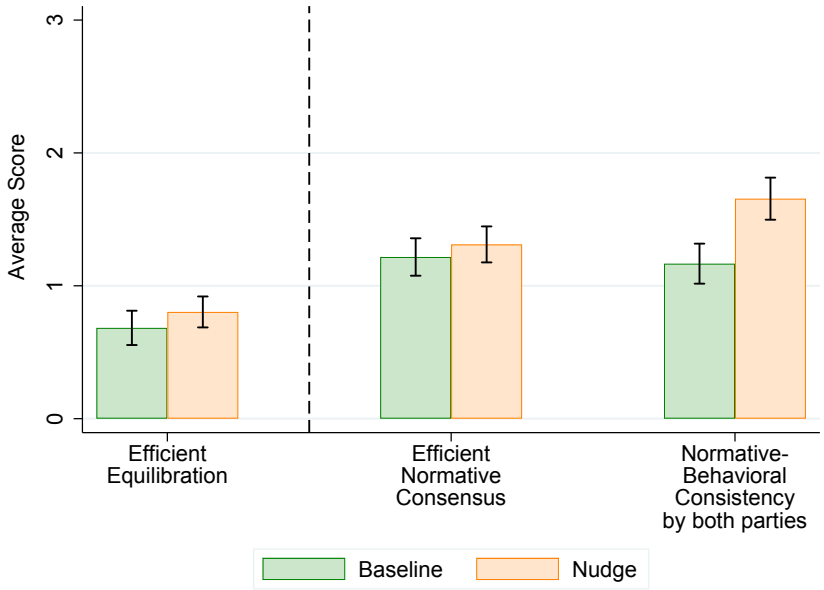


Notes: Data from the Baseline and the Nudge Treatment included. The bars represent average long-run performance (average profits in Phase 2) in the Baseline and the Nudge Treatment, both at the overall treatment level and separately for pairs with principle-based agreements and pairs with rule-based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

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 very poor performance of the very few pairs (15%) who articulated a rule-based agreement is not surprising: it may reflect negative selection. In contrast, the lower performance for pairs with principle-based agreements (in the Nudge compared to the Baseline) was unexpected. This result provides a first indication that at least some of the principles triggered by the Nudge may 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 4.2.3.

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.

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

The right-hand side of the figure provides the data for Result 4c. The Nudge caused only an insignificant increase in efficient normative consensus ($t = 0.49$, $p = 0.63$), but a significant increase in the average number of scenarios in which both parties exhibited normative-behavioral consistency ($t = 2.24$, $p = 0.03$). Both 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). However, contrary to the within-Baseline comparison, there is no significant treatment

effect for efficient equilibration when we compare the Nudge to the Baseline. We explore the reasons for this finding in the next subsection.

4.2.2 Understanding the Impact of the Nudge on Equilibration

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

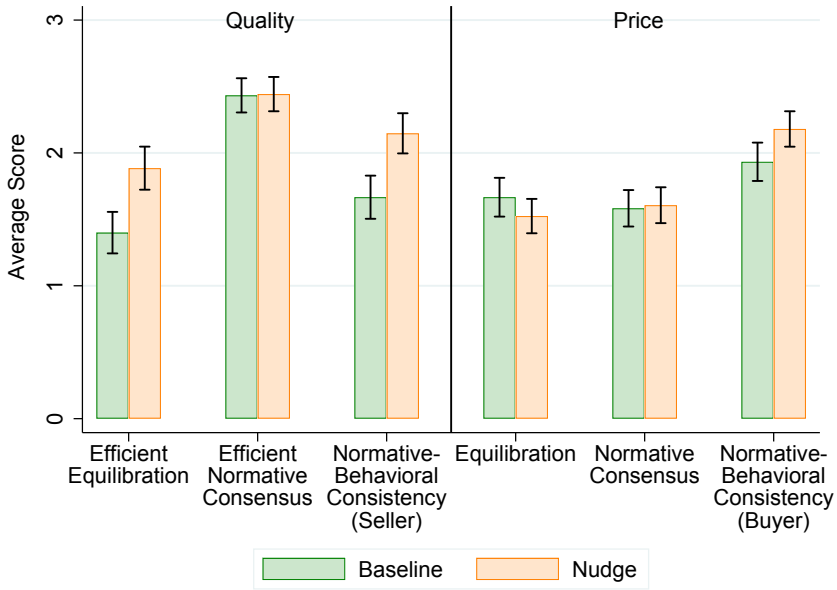
First, the Nudge did make it more likely for pairs to achieve efficient equilibration on quality ($t = 2.16$, $p = 0.03$)—meaning that the seller provided the efficient quality level. Second, the Nudge also created a significant increase in normative-behavioral consistency by the sellers ($t = 2.17$,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

Turning to the price reveals a very different story. The Nudge failed to increase equilibration on price. In fact, we see an insignificant decrease in equilibration on price ($t = 0.73$, $p = 0.47$)—the buyer was not significantly more likely to pay a price as high as the seller expected. Furthermore, the Nudge caused only an insignificant increase in the normative-behavioral consistency of the buyer ($t = 1.26$, $p = 0.21$)—the buyer was not significantly more likely to pay the price she thought the pair's agreement implied. Finally, as with quality, normative consensus on price is essentially unchanged between Baseline and Nudge ($t = 0.12$, $p = 0.90$), although we note for discussion below 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 short-run performance (i.e., the initial quality choices by the seller in the scenarios). Table 4 shows the results of a regression of average quality chosen in each of the 3 scenarios on a Nudge dummy. The data set is comprised of 363 observations (121 sellers, 3 observations per seller). Standard errors are clustered at the seller level.

Column (1) shows a marginally significant treatment effect: in the Nudge, when considering all

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.

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. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .10$

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 very close to the efficient quality level throughout (average quality was 9.4 over the 12 periods of Phase 2). In contrast, pairs who were not equilibrated on price also started out at the efficient quality level of 10 in period 6 (by construction of the sub-sample) but then experienced a deterioration of average quality over time (average quality was 6.6 over the 12 periods of Phase 2).

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. In the latter case, self-serving biases and the like may therefore 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 significantly lower long-run performance in Phase 2 ($t = 2.90$, $p < .01$).

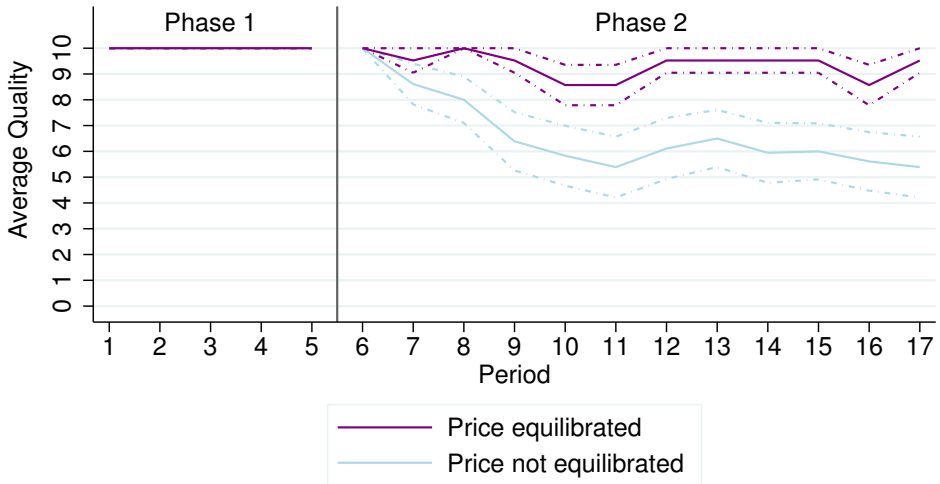
We summarize these exploratory findings in our next result:

Result 5

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

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

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



Notes: All pairs that achieved efficient equilibration on quality in the Nudge are included. The sample is then split into pairs who also achieved equilibration 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.

Importantly, including data from the Baseline in an analysis like Figure 10 leaves the outcome largely unchanged. The relation between equilibration on price and long-run 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.

4.2.3 Digging Deeper: Heterogeneity Among Principles?

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

In this subsection we explore (1) whether there is important heterogeneity among pairs with principles (regardless of whether those pairs were in the Nudge or the Baseline) and (2) whether

such heterogeneity helps explain this difference between conditions in the performance of principle-based agreements. All the analyses in this subsection therefore study the sub-sample of pairs with principles—considering data from the Baseline and the Nudge.

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

We do not find any important differences in personal characteristics between pairs with principle-based agreements in the Nudge and the Baseline. Also, personal characteristics are not significantly correlated with long-run performance (see the Online 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 pairs did), pairs in the Nudge needed 38 seconds (23 percent) longer to get there, but this difference is not statistically significant ($t = 1.11$, $p = 0.27$).

We also explored how these two measures concerning Phase 1—having one party explain to the other, and time to reach this agreement—correlate with long-run performance in Phase 2. Having one party explain to the other was not significantly correlated with long-run performance ($r = -0.08, p = 0.49$), 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 significantly less likely to report that their agreement was "clearly formulated" ($t = 2.76$, $p < 0.01$) and marginally significantly more likely to report that there had been differing interpretations of the agreement between the two parties ($t = 1.75$, $p = 0.09$) than pairs with principle-based agreements in the Baseline. In this sense, pairs in the Nudge did not reach the same level of clarity about their principles as did pairs in the Baseline.

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

5 Discussion

5.1 Summary

This paper studied how parties build relational contracts that help them adapt after an unforeseen shock. We distinguished between rule-based agreements that define specific actions versus principle-based agreements that provide general guidelines. In our Baseline data, agreements that include a principle rather than just a rule achieved higher performance after the environment changed. Although pairs with rule-based agreements achieved a similar level of shared understanding regarding what should be done in a new situation (normative consensus), they were more often unwilling to act accordingly (lack of normative-behavioral consistency). As a result, pairs that relied on rules were less likely to achieve efficient equilibration in their new environment.

One of our goals was to assess causality, so we tried to stimulate principle-based agreements via a simple salience Nudge. Our treatment indeed produced significantly more principle-based agreements. The Nudge also produced an increase in efficient equilibration on quality, but it failed to increase equilibration on price. Put differently, the Nudge increased coordination on efficiency

(the quality dimension) but not on distribution (the price dimension). Conflicting views about distribution may have been the cause of the deterioration of cooperation after the environment changed. As a result, the Nudge did not have a significant effect on long-run performance.

5.2 Causal Effects of the Nudge

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

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 significant treatment effects on some factors that contribute to performance, such as efficient equilibration on quality, short-run performance and normative-behavioral consistency for the seller. If the correlation between principles and performance in the Baseline is fully ascribed to omitted-variable bias, it is difficult to explain these causal effects in the data.

Our findings thus suggest that principles can play *some* role in achieving 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 4.1.1, for two reasons.

First, explicitly articulating a principle may be hard, so some pairs in the Baseline may have concentrated on discussing a rule, while implicitly understanding that their rule is a manifestation of a broader principle. Even the most careful coding of text can capture only what people actually write down, not what they may mean or think. Thus, in the Baseline, some pairs that were coded as having relied on only 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 4.2.3, the Nudge induced some pairs with a lower shared understanding of Phase 1 to agree on a principle concerning Phase 2, but these pairs wrote lower-caliber principles than those who agreed on a principle without the Nudge. 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 long-run performance is due to the Nudge being less effective than intended. An important path for future research is to explore alternative treatments that may be more powerful in getting pairs to agree on high-caliber principles.

Put differently, perhaps one should not be surprised that our simple Nudge did not improve long-run performance. For example, Barney (1986) argued long ago that if organizational culture is to create sustained competitive advantage then the culture must be hard to imitate. The same is true for relational contracts: if the success stories described by Gibbons & Henderson (2012) were easy to imitate, presumably they would be run-of-the-mill stories rather than success stories.

5.3 Limitations and Future Research

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 had significant effects in the short-run, it did not cause a significant increase in long-run 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. Relatedly, as we reported at the end of Section 4.2.3, the Nudge may have crowded out discussion of promises and references to trust. Rather than nudging participants to formulate principles, alternative treatments could attempt to stimulate social preferences in order for pairs to build strong partnerships (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, experiments on communication, 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.

Similarly, we advance the study of communication effects in social-dilemma settings by shedding light on what kind of non-binding agreement helps sustain cooperation in the face of future uncertainty. By showing that principles are associated with equilibration in uncertain settings, we provide one indication of what successful communication should be about.

Finally, these observations about experimental 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.

5.4 Conclusion

In this paper we began to explore how parties build relational contracts that achieve efficient adaptation to unforeseen change. Our results suggest that agreements based on general principles can improve long-run performance. However, we also find that it is difficult to induce the right kind of principles—those that parties not only understand but also follow. In retrospect, the latter finding is consistent with the logic of competitive advantage: if it were simple to build such relational contracts, why is not everyone doing it? That said, our results also leave important open questions. We therefore hope this paper is the starting point of an empirical and theoretical literature that will explore how parties within and between organizations develop shared understanding to achieve adaptation in unstable environments.

References

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

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

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

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

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

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

Camerer, C. & Linardi, S. (2019). Robustness of relational contracts to interruptions in employer-worker gift exchange experiments. *Available at SSRN 3382951*.

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

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

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

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

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

Das, T. K. & Teng, B.-S. (1998). Between trust and control: Developing confidence in partner cooperation in alliances. *Academy of Management Review*, *23*(3), 491–512.

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.

Geertz, C. (1973). *The interpretation of cultures*, volume 5019. Basic books.

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

Gibbons, R. & Prusak, L. (2020). Knowledge, stories, and culture in organizations. *American Economic Review, Papers and Proceedings*, *110*.

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

Grandori, A. & Furlotti, M. (2019). Contracting for the unknown and the logic of innovation. *European Management Review*, *16*(2), 413–426.

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*.

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.

Lumineau, F. & Malhotra, D. (2011). Shadow of the contract: How contract structure shapes interfirm dispute resolution. *Strategic Management Journal*, *32*(5), 532–555.

Macaulay, S. (1963). Non-contractual relations in business: A preliminary study. *American Sociological Review*, 55–67.

Macneil, I. (1978). Contracts: Adjustments of long-term economic relations under classical, neoclassical, and relational contract law. *Northwestern University Law Review*, *LCCII*, 854–906.

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

Malhotra, D. & Lumineau, F. (2011). Trust and collaboration in the aftermath of conflict: The effects of contract structure. *Academy of Management Journal*, *54*(5), 981–998.

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.

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

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.

Zaheer, A. & Venkatraman, N. (1995). Relational governance as an interorganizational strategy: An empirical test of the role of trust in economic exchange. *Strategic Management Journal*, *16*(5), 373–392.

Appendix

A Instructions First Trading Phase

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

General Instructions for Sellers

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

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

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

1 Point = 0.04 Dollars.

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

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

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

Course of the first part of the study

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

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

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

Course of a round

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

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

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

Stage 1: Determination of quality and price

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

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

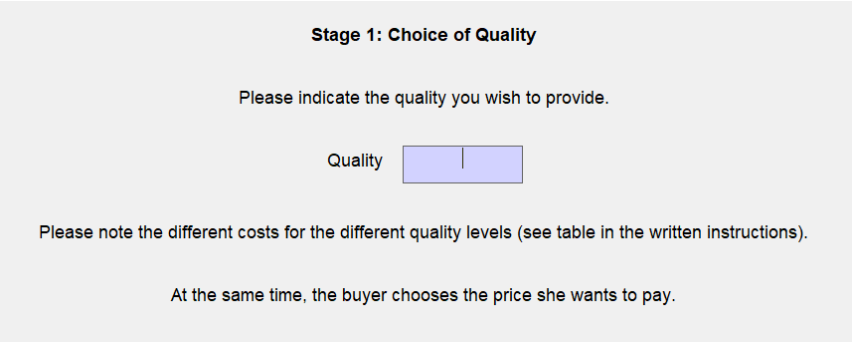
Quality and production cost

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

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

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

You can select the quality using the following screen display:

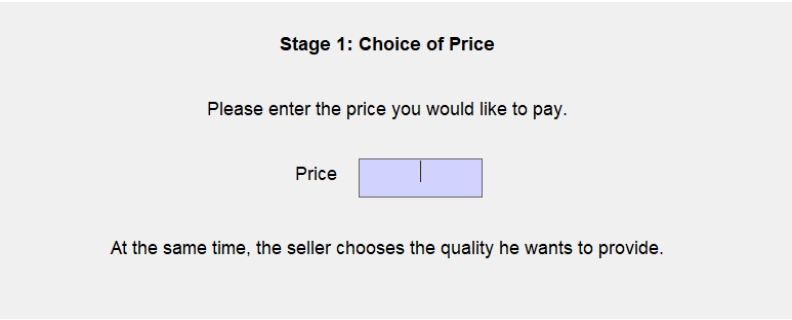


Price

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

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

The buyer enters the price via the following screen display:



Stage 2: Calculation of incomes

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

Your income in a round is calculated as follows:

Your Income = Price – Production Cost of the Provided Quality

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

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

Value of the Product = 10 * Provided Quality.

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

The income of the buyer is thus as follows:

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

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

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

Examples

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

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

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

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

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

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

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

The Second Part of the Study

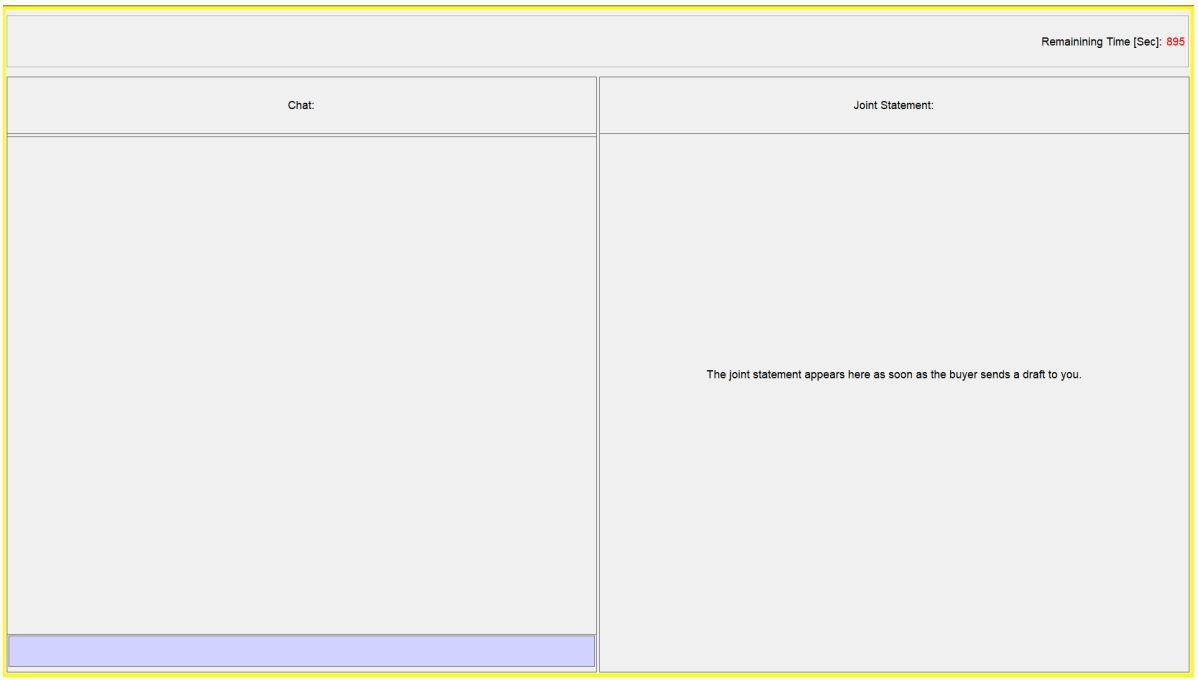
In the second part of the study you are still assigned to the same buyer. You therefore continue to interact with the same person as in the first part. The second part of the study consists of an unknown number of rounds. In the second part of the study, some elements concerning the course of a round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

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

The Agreement

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

You can draft the written statement on the following screen:



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

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

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

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

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

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

B Instructions Measure of Shared Understanding and Second Trading Phase

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

Instructions for the Second Part of the Study for Sellers

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Value of the Product = 20 * Quality.

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

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

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

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

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

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

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

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

Your income = Price – Production cost of the provided quality

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

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

Purchase of the product from you:

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

Alternative source of supply:

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

Examples

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

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

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

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

Scenario 2: Fluctuating value of the product

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

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

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

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

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

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

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

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

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

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

Your income = Price - Production cost of the provided quality

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

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

Examples

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

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

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

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

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

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

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

Value of the product = 20 * Quality.

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

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

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

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

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

  In this situation, your income is calculated as follows:

  Your income = 0 - production cost of the provided quality

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

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

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

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

Your income = price - production cost of the provided quality

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

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

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

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

Examples

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

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

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

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

C Instructions to Coders

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

Coding of Statements and Chats

Background: Structure of the Experiment

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

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

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

Definitions of Coding Categories

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

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

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

Flexible rule:

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

Examples:

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

Suboptimal rule:

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

Example:

- All cases with quality < 10.

Unequal rule:

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

Examples:

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

Principle:

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

Examples:

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

Mention of Fairness:
Participants mention fairness considerations in the agreement / chat. This includes avoiding inequality or striving for equality or discussing what is unfair or fair behaviour.
Examples:
- We try to be fair to each other.
- We want to balance the profits.
- If someone acts unfairly, he will be punished.
- Not paying the agreed price would be unfair.

Mention of efficiency
The participants mention efficiency considerations in the agreement / chat. This is particularly about welfare maximization discussions, i.e. when participants discuss implicitly or explicitly that they want to maximize the sum of payoffs.
Examples:
- We optimize the sum of the profits.
- Optimal quality where marginal costs = marginal benefit

Mention of the second phase
The still unknown second phase of the experiment is mentioned or discussed.
Examples:
- In the second phase we will do xy.
- What happens later in the experiment?

No agreement
"No agreement concluded' (to be coded only for agreements).
Examples:
- You have not made an agreement
- Empty field

Coding Units and Data Structure

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

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

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

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