Flexibility in Long-Term Relationships: An Experimental Study

Kyle Hyndman
Naveen Jindal School of Management, University of Texas at Dallas, KyleB.Hyndman@utdallas.edu,
http://www.hyndman-honhon.com/Main/Kyle_Hyndman.html

Dorothée Honhon
Naveen Jindal School of Management, University of Texas at Dallas, Dorothee.Honhon@utdallas.edu,
http://www.hyndman-honhon.com/Main/Dorothee_Honhon.html

Problem definition We compare two types of contracts which can govern business partnerships: *indefinitely binding agreements* (IBA) wherein partners are bound together in a relationship of indefinite duration and *temporarily binding agreements* (TBA) wherein they are able to dissolve the relationship at any moment.

Academic/Practical relevance Managers need to determine what type of contract they want to engage in with their business partners. While Toyota famously argues for building strong, long-lasting relationships with its suppliers, such a commitment makes it difficult to escape from profit-losing relationships.

Methodology We conduct lab experiments in which subjects are matched in groups of two and play a two-person newsvendor game. In the IBA treatment, subjects are bound together; while in the TBA treatment, they have the ability to dissolve their relationship and be subsequently rematched. We also consider a treatment where the subjects can choose between IBA and TBA, as well as treatments where dissolutions are made more difficult or lead to a fixed outside option rather than rematching.

Results We find that average earnings are higher in the IBA treatment. However, subjects in the TBA treatment with long-lasting partnerships tend to perform better than the average pair in IBA. We argue that TBA facilitates sorting between subjects while IBA provides a greater incentive for forward-looking teaching behavior among subjects. When given the choice, most subjects choose TBA and do better than when the type of contract is imposed on them, irrespective of their choice. We also see evidence of better performance when dissolutions are no longer automatic upon request or lead to the subjects exiting the rematching pool.

Managerial implications Having the flexibility to dissolve a relationship can be beneficial – but this option must be used wisely. When engaging into partnerships with suppliers and customers, managers should balance out the teaching benefits of long-term contracts, with the potential drawbacks of unpunished disingenuous behavior. Short-term contracts may be preferable if the parties can refrain from breaking up the partnership in dark times which are beyond the control of the business partners.

Key words: Long-term relationships; joint ventures; institution choice; coordination
JEL codes: C72, C91, D83
History: **June 11, 2018**

1. Introduction

Researchers in economics and operations management usually assume that economic agents such as firms or individuals interact for an exogenously pre-determined length of time (which can be finite or infinite). However, in reality, when undertaking a venture, agents typically choose the duration and scope of their relationship. Thus, the endogenous formation and dissolution of both business partnerships and employer/employee relations, and the terms governing such relationships is an important – and neglected – area of research.

Long-term relationships have a number of advantages. For example, they may be able to support more efficient equilibria than when relationships are short-lived. In the business world, long-term relationships can support investment in information systems, communication systems and other relationship-specific investments. They may also signify trust and commitment, which could be efficiency-enhancing. Indeed, long-term contracts in supplier/buyer relationships have been linked to improved supply chain performance such as cost reductions, quality improvements and higher profitability (see, e.g., Han et al. (1993), Kalwani and Narayandas (1995)), as well as improved information sharing (Özer et al. 2011). Long-term relationships can also increase efficiency because relational incentives may substitute for costly/inefficient monetary incentives (Davis and Hyndman 2018). As another example, Uzzi (1996) analyzes data from the International Ladies Garment Workers' Union in 1991 studying how the mix of long-term vs. short-term business relationships affected firm survival. He shows that a firm that has only short-term contracts was nearly twice as likely to exit the market as a firm that has long-term contracts with only one other firm. One other role for long-term relationships that has been pointed out is their ability to promote learning. One well-known example is that of Toyota and their building of supplier *keiretsu*: "close-knit networks of vendors that continuously learn, improve, and prosper along with their parent companies" (Liker and Choi 2004). Camerer et al. (2002) and others in economics have highlighted another benefit of long term relationships, which is that they create stronger incentives for players to engage in forward-looking behavior such as "teaching" their partner to choose better, more efficient, actions.

However, long-term contracts are not without their drawbacks. They may increase dependency and make it difficult to escape from profit-losing relationships, especially when relationship-specific investments are present. Long-term partners can also have higher expectations and be more difficult to satisfy (Kalwani and Narayandas 1995). The reality is that long term partnerships and joint-ventures fail very often, with studies suggesting that the success rate of business joint ventures is only about 50% (Bamford et al. 2004, KPMG 2009, Center for Digital Strategies 2006).

There are several reasons why partnerships may fail. First, long-term relationships may create incentives for opportunistic behavior. For example, Neuville (1997) documents opportunistic behavior by automobile parts supplier that was found to be cutting corners on quality. Second, a partnership may fail because the incentives of the parties are misaligned, as appeared to be the case for Microsoft and Nokia (Grundberg and Stoll 2012, Kovach 2014). Third, partnerships may simply fail to live up to expectations because of problems coordinating effort. Finally, problems may occur because of bad luck despite good intentions of the parties involved. These last two factors appear to be at the heart of Larsen et al.'s (2010) case study of the LEGO/Flextronics joint venture.

In this paper, we focus on how flexibility impacts the success or failure of relationships. The main question we seek to answer is: to what extent is the option to dissolve a relationship a desirable feature of a business relationship? Our answer to this question is: despite some positive aspects to flexibility, the overall average effect is detrimental. Therefore, we devote a substantial portion of our analysis to understanding why and we consider several changes to our basic experimental design to highlight the key drivers of our result.

We conduct lab experiments in which subjects are matched in groups of two and play a version of a two-person newsvendor game. In particular, two decision makers simultaneously choose capacities; revenues are determined by the minimum of their chosen capacities and a random demand realization, and each player pays the cost of his/her capacity. This game can represent a team production environment, similar to the minimum effort game. The main difference is the additional random component that may also determine the minimum. Arguably, there is a stochastic component to most such environments – for example, obstacles may arise which limit the team's production, despite hard work by all members. In a project management setting, Kwon et al. (2010) incorporate randomness about when each firm's task will be finished. In their setting, the randomness is correlated with the firm's effort, while in ours, it is independent of the players' actions. The randomness of revenues is an important feature of our setting because it allows us to test whether luck plays a role in the success or failure of relationships as suggested by Larsen et al. (2010).

Our environment can be viewed as implementing a "risk sharing" contract in which suppliers are paid based on total sales rather than based on the number of units provided. Thus it is similar to the type of contract Boeing implemented with its tier 1 suppliers for the Boeing 787 (Tang et al. 2009). Other studies in project management which are related include Kwon et al. (2010) and Xu and Zhao (2013). Alternatively, our game can also represent an assembly system in which players provide crucial components to a final product assembler and are paid based on total sales, which are given by the minimum of the capacity of each player and demand for the final product. As such, our game shares many similarities with the "assembler as leader" model from Wang and Gerchak (2003) as well as Bernstein and DeCroix (2004). Coordination of capacity in such systems appears to be difficult to achieve as noted by Cachon and Lariviere (2001), Tomlin (2003), Hendricks and Singhal (2005), Lunsford (2007), Hyndman et al. (2013, 2014), among others.

To answer our main question, we compare subjects' behavior under two different, exogenously imposed, institutions: *indefinitely binding agreements* (IBA) and *temporarily binding agreements* (TBA). In both cases, subjects are matched in pairs and play the two-player newsvendor game for an indefinite number of periods. Under IBA, subjects know that they will play with the same player for the entire duration of their interaction, while under TBA, subjects can dissolve their pairing at the end of every period and be rematched with another subject whose pairing also dissolved. This is the distinguishing feature of our paper from the existing behavioral OM literature.

We find that subjects in the IBA treatment earn significantly more – about 10% – than subjects in the TBA treatment. The difference appears to be due to higher capacity choices, rather than better alignment of decisions. However, these averages mask some interesting findings. First, subjects in the IBA treatment can descend into a near complete coordination failure for several periods. This effect, although rare, is the dark side of long-term relationships. Such extreme outcomes never occur in the TBA treatment. Second, in the TBA treatment, nearly 30% of initial pairings survive until the random termination of the repeated game. Subjects in these long-lasting pairings choose higher actions, are better-aligned and earn substantially more than either the subjects in the TBA treatment whose initial pairings dissolve or subjects in the IBA treatment. They also gradually increase their capacities over time, which increases profits, while the subjects who frequently break up their relationships choose lower and lower capacities over time. More generally, in the TBA treatment, we find that subjects who chose to have a limited number of partners perform the best and subjects who frequently break up relationships perform the worst. This suggests that having the flexibility to dissolve a relationship can be beneficial – but that this option must be used wisely.

We argue that there are four inter-connected reasons for the relatively poor performance of TBA, which we investigate using three additional experiments. First, we show that relationships in the TBA treatment dissolve quite frequently. While subjects dissolve relationships for sensible reasons – e.g., their partner's capacity was below their own – they also do so for arguably irrational reasons. For example, controlling for capacity choices, subjects are more likely to dissolve the lower were their realized profits, which is a function of the random demand component. In other words, bad luck also plays a role in whether or not relationships are maintained. Second, since the rematching pool consists of subjects whose pairings dissolved, its quality may deteriorate over time. If subjects do not anticipate this, they may dissolve too frequently. We address these two issues with two treatments designed to either limit the flexibility to dissolve or to replace the rematching pool with an exogenous termination payoff. Third, we argue that the TBA institution reduces the incentives for sophisticated teaching, which has been shown to promote efficient coordination (e.g., Brandts and Cooper 2006, Hyndman et al. 2009). Finally, we argue that subjects differ in their willingness to try to make long-term relationships work. By making the institution choice endogenous, subjects can signal their type/intentions, which may reduce strategic uncertainty. Therefore, we consider a treatment which endogenizes the choice between TBA and IBA.

Our results show that each of the explanations have merit. Limiting the flexibility to dissolve relationships appears to make subjects more focused on the long term, which leads to higher capacity choices, longer lasting relationships and higher earnings. Similarly, by replacing the rematching pool with a fixed, exogenous payoff upon terminating a relationship, very few relationships actually dissolve and subjects also make better, more profitable decisions. Finally, when the institution – TBA or IBA – is endogenous, our results are consistent with the signaling hypothesis: subjects who chose either institution do better than when the same institution was exogenously imposed and the gap between the TBA and IBA institution is eliminated. Despite this benefit, some drawbacks of TBA remain. Specifically, subjects who chose into TBA still appear to dissolve too much and because of bad luck. Similar to the exogenously imposed TBA institution, subjects who had fewer pairings tend to earn significantly more than those who had many pairings. Thus, a key insight of our research is that, by itself, flexibility is disadvantageous but that when combined with something to promote a long-run focus, as two of our follow-up studies appear to do, it may be advantageous.

2. Related Experimental Literature

Games With Multiple Pareto Rankable Equilibria. The game that we study shares many features of the minimum effort game, which is well-studied in experimental economics, starting with van Huyck et al. (1990). Coordination failures are quite common as soon as the group size is three or more (Knez and Camerer 1994, Brandts and Cooper 2006), even under fixed matching. However, for fixed groups of size two, van Huyck et al. (1990) show that nearly all groups eventually converge to the efficient equilibrium, while under random matching, no stable pattern emerges.

In OM, there is a small but growing literature on coordination games. The three most closely related studies are Hyndman et al. (2014), Shokoohyar et al. (2016) and Fan et al. (Forthcoming). Using a similar game, Hyndman et al. (2014) studied how behavior differed when subjects played a series of one-shot games (i.e., random matching) or played a finitely repeated game with the same partner. They showed that, in later rounds, profits were often higher and less variable under one-shot interactions. The apparent cause for this difference was a *first-impressions bias*, which was present only under fixed matching. This suggests that the flexibility of TBA may be able to reduce this negative path dependence. Indeed, our results find some support for this, but show that flexibility is over-used. In a minimum-effort game, Shokoohyar et al. (2016) show how coordination can be improved by allowing effort to accumulate (and information flows to all players) over several stages. Fan et al. (Forthcoming) study the minimum effort game with endogenous group selection. Specifically, players can choose to enter a group with an entrance fee or a group with no entrance fee. Subjects choosing to pay the entrance fee generally make more efficient choices. Thus there is some signaling value to choosing to join the exclusive group. This is similar to our result that endogenous institution selection signals one's type/intentions and promotes more efficient coordination.

Endogenous Matching/Termination. The only papers that we are aware of in OM that study endogenous matching are the aforementioned Fan et al. (Forthcoming) and Beer et al. (2017) who study a two-period contracting problem in a market with two suppliers and two buyers. They study the role that supplier awards have on promoting quality. In experimental economics, Riedl et al. (2016) also study minimum effort games in which players can choose with whom to interact. They show that the ability to endogenously choose one's group (and group size) is efficiency enhancing.

There is also a literature in experimental economics which studies endogenous matching in social dilemmas such as public goods games or prisoner's dilemma-like games. In a public goods game, Page et al. (2005) show that periodically allowing subjects to choose with whom they interact leads to higher contributions through a process of assortative matching by contribution level. Other examples include Bernard et al. (2014) who study gift exchange games and D'Evelyn (2013) who considers a stylized marriage/dating market. Bolton et al. (2008) examine buyer-seller relationships and, in some treatments, allow buyers to either maintain a long-term relationship with the same seller or to interact with a new seller. They show that the benefit of long term relationships disappears in a competitive market where buyers and sellers can build reputations.

Wilson and Wu (2017) study cooperation when partnerships – in a prisoner's dilemma-type setting – can be dissolved. They impose exogenous termination payoffs in the event that a partnership is dissolved and find that symmetric termination payoffs increase cooperation, while asymmetric termination payoffs lead to inefficient outcomes. Finally, Honhon and Hyndman (2016) study the role of the flexibility to dissolve relationships in the repeated prisoners' dilemma, where the possibility for opportunistic behavior emerges. Like our results, they show that flexibility by itself is clearly disadvantageous, but flexibility combined something to promote a long-term orientation – in their case, a reputation mechanism – can lead to higher earnings that even the IBA institution.

3. Experimental Design

3.1. The Stage Game

In all of our experiments, we use a two-player newsvendor game as our stage game. In this game, subjects are matched in groups of two and they each simultaneously decide on a *capacity* level before observing the realization of a random *demand*. Sales are equal to the minimum of the two chosen capacities and the realization of demand. The profits of each player are increasing in sales but, holding sales constant, decreasing in their own chosen capacity. Formally, let k_i denote the capacity chosen by player i. The profits π_i of player i are given by:

$$\pi_i(k_1, k_2, D) = 5\min\{k_1, k_2, D\} - 2k_i \text{ for } i = 1, 2 \tag{1}$$

where 5 is the selling price, 2 is the unit-cost of capacity and D is the realized demand, which is a random draw from the discrete uniform distribution over $\{1, 2, \ldots, 100\}$. Subjects know the demand distribution but do not observe the realization until capacity decisions, k_i, have been made.

This game captures operationally relevant features of several settings. It is similar to the "assembler-as-leader" model of Wang and Gerchak (2003) as well as Bernstein and DeCroix (2004). The game shares many features of Shokoohyar et al. (2016) and Fan et al. (Forthcoming), and is also similar to the model of Kwon et al. (2010), who study project management. The common thread between our game and these papers is the presence of multiple, Pareto rankable equilibria.

To characterize the set of Nash equilibria, we study player 1's optimal capacity choice given a capacity choice of k_2 for player 2. Player 1's expected profit is equal to:

$$\begin{aligned}\mathbb{E}_D[\pi_1(k_1, k_2, D)] &= \begin{cases} \frac{1}{100} 5\left(\sum_{D=1}^{k_1} D + \sum_{D=k_1+1}^{100} k_1\right) - 2k_1 & \text{if } k_1 \leq k_2 \\ \frac{1}{100} 5\left(\sum_{D=1}^{k_2} D + \sum_{D=k_2+1}^{100} k_2\right) - 2k_1 & \text{if } k_1 > k_2 \end{cases} \\ &= \begin{cases} \frac{121k_1 - k_1^2}{40} & \text{if } k_1 \leq k_2 \\ \frac{201k_2 - k_2^2}{40} - 2k_1 & \text{if } k_1 > k_2 \end{cases}\end{aligned} \tag{2}$$

The first part of the expression (for $k_1 \leq k_2$) is maximized at $k_1 \in \{60, 61\}$ and the second part (for $k_1 > k_2$) is decreasing in k_1. Therefore, given a capacity of k_2 for player 2, player 1 should pick $k_1 \in \{60, 61\}$ if $k_2 \geq 61$ and k_2 otherwise. Thus, we have the following result:

Result 1 *There are 61 pure strategy Nash equilibria:* (k, k) *with* $k \in \{1, 2, \ldots, 60, 61\}$*. Moreover, the equilibria are Pareto rankable:* $\mathbb{E}_D[\pi_i(k, k, D)] < \mathbb{E}_D[\pi_i(k+1, k+1, D)]$ *for* $i = 1, 2$ *and* $k = 1, \ldots, 59$*, and* $\mathbb{E}_D[\pi_i(60, 60, D)] = \mathbb{E}_D[\pi_i(61, 61, D)]$*.*

In every equilibrium, both players choose the same capacity, k, and higher capacity choices lead to higher expected payoffs up to $k \in \{60, 61\}$. These latter two equilibria are the *Pareto efficient* equilibria, and each player has an expected profit of 91.5. Further, the variance of equilibrium payoffs, equal to $\mathbb{V}_D[\pi_i(k,k,D)] = {(k(200-603k+406k^2-3k^3))}/{4800}$, is incrasing in k for $k \in \{1, ..., 61\}$. In other words, higher capacities lead to higher expected payoffs (up to 61) but also higher risk.

In what follows, we refer to the absolute difference between the two players' chosen capacities, i.e., $|k_1 - k_2|$, as the degree of *misalignment* between the two players. We also say that two players are closer to a *coordination failure* the lower is the smallest chosen capacity; i.e., $\min\{k_1, k_2\}$.

3.2. The Treatments

In all treatments, the subjects' interactions were divided into 8 *cycles*, with each cycle consisting of an a-priori unknown number of *periods*. The length of each cycle was determined by simulating the roll of a fair 10-sided die. The cycle would end if a 10 was rolled, otherwise it would continue for at least one more period. This is the standard method, first introduced by Roth and Murnighan (1978), used to mimic infinitely repeated games in a laboratory setting.

In each period, subjects played the stage game in groups of two, referred to as a *pairing*. We refer to a subject's partner in a group as his/her *match*. At the beginning of each period, subjects made their capacity decisions simultaneously. Subsequently, they were given the following feedback: (i) realized demand, (ii) their capacity choice, (iii) their match's capacity and (iv) their own payoff for that period. Furthermore, in each period, subjects could see a history table with information on chosen capacities, demand realizations and profits for every previous period of the current pairing.

The experiments differed based on the institution that was in force governing the relationship of each pair. We have two main treatments and three additional follow-up/robustness treatments.

Treatment 1: Indefinitely Binding Agreements (IBA). Pairings were formed randomly at the start of each new cycle and subjects played the stage game with the same match for the entire cycle.

Treatment 2: Temporarily Binding Agreements (TBA). After receiving feedback at the end of a period, subjects were asked whether they wanted to remain matched to the same subject for the next period or be rematched to another subject. It was enough for one player in the group to request to be rematched to trigger a break-up of the pairing. Assuming the cycle continued for another period, at the beginning of the next period, all subjects whose pairings broke up were randomly rematched from within this subset of subjects. In this case, subjects were explicitly informed that they had been rematched at either their or their previous match's request.

In order to better understand the reason for the differences we observe between the IBA and TBA treatments, we consider three more treatments.

Treatment 3: Exogenous Termination Payoff (TBA-FP). At the end of each period in which they are *active*, subjects could choose to maintain the pairing or to terminate it. Terminating the pairing placed subjects in an *inactive* state and gave them a fixed payoff of 50ECU per period until the end of the cycle. In comparison, playing one of the Pareto-efficient Nash equilibria of the game (Result 1) generates an expected profit of 91.5ECU and the worst equilibrium generates an expected profit of 3. 50 is also midway between the average per period earnings of subjects who dissolved frequently and those who dissolved infrequently in the TBA treatment.

Treatment 4: Disallowed Rematching Requests (TBA-D). Each time a new pairing is formed, there was a 10% chance that the subjects in the pairing would be forced to remain matched until the end of the cycle. Subjects were infromed of this possibility in the instructions but did not know whether they would be able to dissolve a particular pairing or not until they actually requested to do so. In those instances, the subject who requested a dissolution was told that the rematching request was disallowed and that he/she would remain paired with the same subject until the end of the cycle. The subject's match was not informed that a rematching request was made and denied.

Treatment 5: Institution Choice (IC). At the beginning of each cycle, subjects chose which institution they would like to govern their interaction for the current cycle: either IBA and TBA. When an even number of subjects chose each institution, then initial pairings would be formed by randomly matching subjects from within their chosen institution. Once the institutions were determined and initial pairings were formed, everything proceeded according to the rules of their respective institution, as described above. If an odd number of subjects chose each institution, then one subject who chose TBA was randomly assigned to the IBA institution. The subject and his/her match were informed that one of them had requested TBA but was actually placed in the IBA institution. This was done to minimize any potential negative effects of granting one subject her preferred institution because, had we placed the subject in TBA, she could potentially interact with all other subjects who chose TBA. It was rare – occurring in 5.92% of initial pairings – to place one subject in the IBA institution when they chose the TBA institution.

3.3. Details of the Experiment

For each treatment, we conducted four sessions at the experimental laboratory of a public U.S. university. To facilitate data analysis, we used the same realization of random numbers across treatments for the two sources of randomness in our experiment: the length of each cycle and the demand realizations. For the cycle length, we pre-drew random numbers to determine the length of the eight cycles and created four random permutations, one for each session. Therefore, in all sessions, subjects experienced the same eight values of cycle length, but in a different order. For the demand realizations we also pre-drew four random sequences – one for each session – that subjects would see in each period. Therefore, in each session $\#i \in \{1,2,3,4\}$, subjects in each of the five treatments experienced the same sequence of cycle lengths, in the same order and with the same demand realizations.

After reading the instructions, but before starting the experiment, subjects had three minutes to get familiar with the stage game: they could enter hypothetical values for their and their match's decisions and see a graph with the payoff consequences for each possible value of demand.

Table 1 Information About Sessions

		Number of Subjects					Average Earnings (USD)				
	Cycle Lengths	IBA	TBA	TBA-D	TBA-FP	IC	IBA	TBA	TBA-D	TBA-FP	IC
S1	$(3,5,4,6,12,7,26,16)$	10	20	22	22	20	21.34	16.18	17.27†	23.54	17.72
S2	$(5,16,6,7,3,4,26,12)$	10	20	20	20	16	14.76	14.03	21.92	20.40	17.72
S3	$(4,26,16,12,6,3,5,7)$	10	18	20	20	18	17.61	17.03	22.00	20.56	21.86
S4	$(4,3,26,7,6,12,5,16)$	12	20	22	14	20	22.78	18.50	19.44	22.63	19.86
Overall		42	78	82	76	76	19.30	16.42	20.09	21.76	19.32

Note: The average earnings include both earnings from the main part of the experiment as well as from the incentivized risk elicitation task that took place after the completion of the main experiment.
† In this session, due to a malfunction, we could only complete 55 of 79 periods.

After playing all eight cycles, subjects completed both an incentivized and non-incentivized risk elicitation task, and a final survey, which contained: (i) demographic questions, (ii) questions specific to the experiment (such as about their strategy), (iii) the Cognitive Reflection Test (CRT) (Frederick 2005) and (iv) a 10-item set of questions to measure each subject's "Need For Cognition" (NFC) (Cacioppo and Petty 1982). Each session lasted approximately 90 minutes. Due to time constraints, we were unable to complete the final survey in three TBA sessions and one TBA-FP session. For each measure, we conducted the Kruskall-Wallis test for differences across treatments. There is weak evidence that the incentivized risk measure differs across treatments ($p = 0.073$); the other three measures all have $p > 0.10$, suggesting that our treatment randomization was successful. Table 1 summarizes the details of the experiment.

Remark 1. The difference in risk preferences is mainly driven the TBA treatment where subjects are significantly more risk averse than subjects in the IC treatment. We are not concerned with this because (i) when doing so many pairwise tests, it is likely that some comparisons will be significant by chance and (ii) the risk elicitation was done after the main experiment. Since subjects performed worst in the TBA treatment, it may have induced more risk averse behavior in the incentivized elicitation. This interpretation is further supported by the fact that there are no differences in the non-incentivized risk elicitation, which asked of "general willingness to take risks" and where we would expect less contamination from the main experiment because it is more general and has no monetary consequences.

3.4. Remarks on the Experimental Design

The rematching process of subjects in the TBA and TBA-D treatments was done randomly at the end of each period from the set of participants whose pairing broke up. If more than one pairing broke up, our software made sure that subjects were not matched with the same participant in the next period. However, if only one pairing broke up, the same two subjects were matched again in the next period. In this case, subjects were not informed that their match was the same person as in the previous period and could not infer that this was the case. In order to limit this forced rematching due to only one pairing dissolving in any period, the TBA sessions had about twice as many subjects per session as the IBA sessions. Indeed, in the TBA treatment, this forced re-matching only occurred in 0.58% of the cases in which subjects needed to be rematched.

Since our stage game has multiple Pareto-rankable equilibria, it is not clear what role indefinite repetition plays. Unlike a prisoner's dilemma game, there are no equilibria of the indefinitely repeated game that dominate the Pareto-efficient stage game Nash equilibrium. Nevertheless, we chose to run the experiment as a series of indefinitely repeated games because we wanted to create a stationary environment in which to study the decision to dissolve relationships in the TBA treatment. With a fixed, finite number of periods, the decision to dissolve would be influenced by the number of periods remaining which we wanted to avoid. It also facilitates a study of far-sighted teaching behavior, as we will demonstrate later. In our setting, with a continuation probability of 90%, the end of each cycle is indefinite and in particular, subjects should always expect the cycle to continue for 10 additional periods. Hence, a subject contemplating teaching can expect 10 additional periods in which to potentially enjoy the fruits of successful teaching. In contrast, under a fixed, finite number of periods, the incentives to teach would be non-stationary.

4. Results: Temporarily and Indefinitely Binding Agreements

We first analyze behavior in the two treatments where the institution is exogenously imposed: TBA and IBA. Our main goal is to determine which institution leads to more efficient outcomes and to gain insights, in the case of TBA, about the events that trigger the dissolution of a pairing.

4.1. Performance Analysis and Determinants of Capacity Choice

In Table 2 we provide some summary statistics on average capacity choices, misalignment in a pairing (defined as the absolute difference of the capacity choices, i.e., $|k_1 - k_2|$) and profits for

Table 2 Summary Statistics (Average Capacity, Misalignment and Profits)

	Average Capacity	$p-$value (vs IBA)	Average Misalignment	$p-$value (vs IBA)	Average Profit	$p-$value (vs IBA)
IBA	37.05		9.39		56.48	
TBA pooled	33.72	0.019	9.59	0.841	51.38	0.001
TBA 1 pairing (28.8%)	42.21	0.448	5.18	0.000	72.09	0.000
TBA 2 pairings (13.5%)	36.73	0.073	7.84	0.328	60.20	0.780
TBA 3 pairings (17.3%)	35.73	0.818	12.47	0.030	44.10	0.000
TBA 4+ pairings (40.4%)	28.77	0.000	11.50	0.023	41.08	0.000

Note: $p-$values are generated from random-effects regressions (with standard errors corrected for clustering at the treatment-session level), where we include indicators for the explanatory variables. In all cases, we also include indicator variables for the different sessions. The percentages in parentheses indicate the frequency of observations across all subject-cycle combinations.

each of the two treatments. For the TBA treatment, we break down these statistics between those subjects who had 1, 2, 3 or 4+ pairings during each cycle.

Consider first the comparison between IBA and the pooled average for TBA. Subjects in the TBA treatment are approximately as well-aligned as their IBA counterparts but choose significantly lower capacities. Because of this, subjects in the IBA treatment earn significantly more – approximately 10% more – than subjects in the TBA treatment. Thus, on average, the flexibility to dissolve relationships is disadvantageous. However, this aggregate comparison masks interesting differences which emerge once we differentiate based on the number of pairings experienced by a subject during a cycle. Subjects in the TBA treatment whose initial pairings never dissolved were significantly better aligned and earned significantly higher profits – approximately 33% more – than subjects in the IBA treatment. Subjects in the TBA treatment who had two pairings during the cycle (i.e., their initial pairing dissolved, but the subsequently formed pairing lasted for the remainder of the cycle) are also better aligned and earn more profits than subjects in the IBA treatment, but the difference is not significant. However, subjects in TBA who had 3 or 4+ pairings during the cycle chose lower capacities, suffered from greater misalignment and earned significantly less than subjects in the IBA treatment. This is our first piece of evidence that performance in TBA appears to be strongly related to the frequency with which subjects choose to dissolve relationships.

Table 3 takes a regression approach to better understand the determinants of capacity choices and profits. Consider columns (1) and (2), which focus on capacity. There are four results to highlight. First, across cycles, capacities decline in TBA, but are relatively stable in IBA. Second, within a cycle, the coefficient on Cycle Period is significantly negative for both TBA and IBA. This decreasing trend indicates negative reinforcement of capacities across periods. However, and this is our third result, longer-lived pairings in TBA choose higher capacities, as evidenced by the fact that the sum of the coefficients on TBA $\times$ Pairing Duration and Cycle Period is significantly

Table 3 Random Effects Regression: Factors Influencing Capacity Choice and Profits

	Capacity				Profits			
	(1)		(2)		(3)		(4)	
TBA	−2.746	(3.681)	−2.266	(3.351)	−12.756	(12.995)	−12.954	(12.332)
TBA × Cycle Index	−0.257*	(0.135)	−0.259*	(0.136)	1.764**	(0.845)	1.757**	(0.846)
IBA × Cycle Index	0.371	(0.289)	0.370	(0.288)	2.243*	(1.244)	2.271*	(1.248)
TBA × Pairing Duration	0.611***	(0.136)	0.611***	(0.135)	2.453***	(0.391)	2.458***	(0.401)
Cycle Period	−0.328***	(0.057)	−0.328***	(0.057)	0.079	(0.308)	0.078	(0.308)
Willingness to Take Risk (Incent.)			0.247	(0.380)			−0.512**	(0.255)
Willingness to Take Risk (Non-Incent.)			1.278***	(0.382)			1.377*	(0.761)
Constant	37.926***	(2.444)	27.743***	(2.544)	44.963***	(9.402)	38.519***	(10.415)
Observations	9480		9480		9480		9480	
R^2	0.082		0.114		0.039		0.041	

Note 1: ***1%, **5%, *10% significance using standard errors clustered at the treatment-session level.
Note 2: Since we are interested in treatment differences and also the role of risk preferences, both of which are fixed for each subject, we must estimate the model with random-effects.

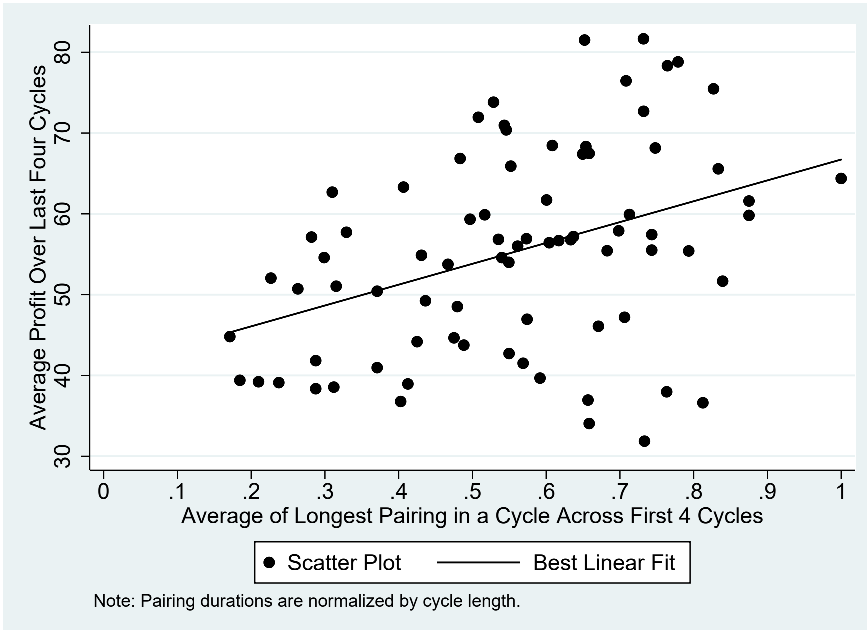
positive ($p = 0.012$), so much so that the longest-lived pairings are choosing significantly *higher* capacities in later periods. Finally, capacity choice is increasing subjects' willingness to take risk. This is consistent with Result 1, which showed that the variance of payoffs is increasing in capacity.

Turn to columns (3) and (4) of Table 3, which focus on profits. First, notice that subjects earn more in later cycles in the experiment. Second, subjects in longer-lasting relationships in TBA earn significantly more. Lastly, we get a mixed result with respect to risk preferences. According to our non-incentivized measure, subjects who are more willing to take risks earn (marginally) significantly more; however, the result is flipped when we consider the incentivized measure. Therefore, while willingness to take risk has an consistent effect on capacity choices, its effect on profits is ambiguous.

Tables 2 and 3 show that the number of dissolutions per cycle affects profits. It is also the case that subjects who have fewer, longer-lasting relationships in the first four cycles earn significantly higher profits over the last four cycles. This can be seen in Figure 1, which shows a scatter plot of subjects' average per-period profit over the last four cycles against the average duration (as a fraction of cycle length) of their longest pairing in each of the first four cycles (remember each subject saw 8 cycles in total). The figure also shows the best linear regression fit, which clearly shows a positive relationship ($p = 0.016$). This suggests that subjects who adopt a more "patient" strategy in terms of dissolving relationships are rewarded with higher profits.

Remark 2. Some caution is needed in making a direct causal link between our measure of patience and average profits over the last four cycles. There could be another variable which affects pairing duration and profits, but is unrelated to patience. The most likely variable is initial misalignment. Below, we show that subjects who had only one pairing started the cycle much better-aligned. However, while there is a negative relationship between patience and average initial

Figure 1 **Average Earnings Over Last Four Cycles Against Average Duration of Longest Pairing in a Cycle Across First Four Cycles (TBA Treatment)**



misalignment in the first four cycles, initial misalignment does not affect average profits in the last four cycles. Moreover, controlling for initial misalignment, subjects with longer average pairing durations over the first four cycles had longer average pairing durations over the last four cycles. This suggests that we are capturing patience and that the relationship in Figure 1 is likely causal.

4.2. Evolution of Capacity Choices

As we have mentioned, outcomes in coordination games are often highly history-dependent. In Table 4, we report the average capacity choice, misalignment and profits over the first period and last period of each cycle. For the TBA treatment, we differentiate between whether the initial pairing was maintained over the entire cycle (“1 pairing”) or not (“2+ pairings”). Since all treatments consisted of four sessions and all sessions had the same number of cycles (8) with identical lengths (though in different orders), we can compare behavior between the first and last period of a cycle without worrying about differences in cycles and cycle lengths between the treatments.

All groups of subjects are substantially and significantly better-aligned in the last period of a cycle than in the first period ($p < 0.01$). However, the same cannot be said for capacities. In IBA, the average capacity declined by 4.54 points over the course of a cycle, and this difference is statistically significant ($p < 0.001$). In contrast, in TBA, the overall average capacity declined by only 1.97, and this is not significant ($p = 0.205$). However, this masks differences between the two subgroups in TBA: those subjects with 2+ pairings, saw capacities decline by a statistically

Table 4 Comparing the First and Last Periods of Cycles

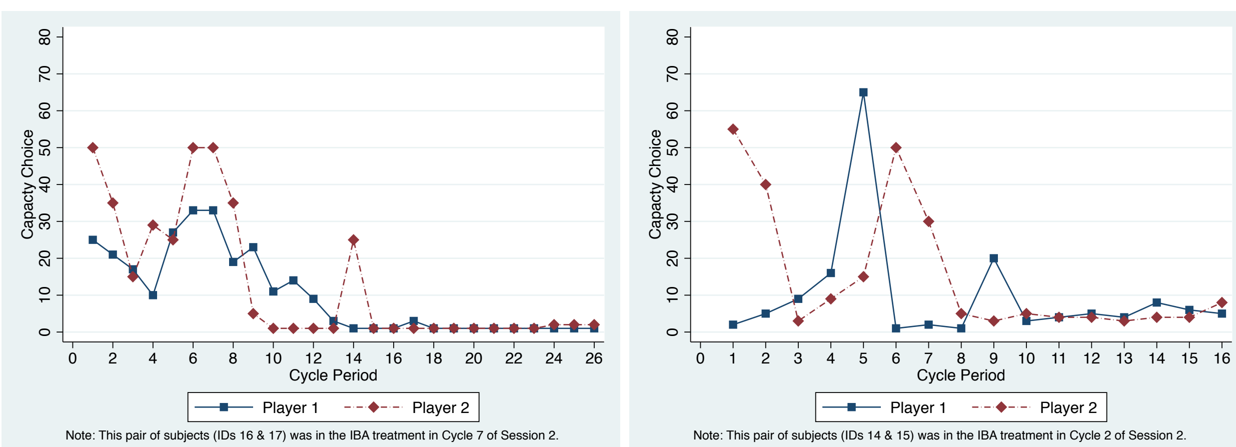
	Average Capacity			Average Misalignment			Average Profits		
	First	Last	$p-$value	First	Last	$p-$value	First	Last	$p-$value
IBA	41.01	36.47	0.000	16.27	9.08	0.000	46.21	60.66	0.231
TBA (pooled)	36.04	34.07	0.205	15.00	8.73	0.000	44.73	56.32	0.237
TBA (1 pairing)	39.55	44.04	0.043	9.68	5.14	0.000	59.79	84.39	0.080
TBA (2+ pairings)	34.73	30.33	0.000	17.01	10.08	0.000	39.09	45.81	0.477

Note: $p-$values are generated from random-effects regressions (with standard errors corrected for clustering at the session level), where we include indicator variables as the relevant explanatory variables. In all cases, we also include indicator variables for the different sessions.

significant 4.40 ($p < 0.001$), while those with only 1 pairing, saw their capacities *increase* by a statistically significant 4.49 ($p = 0.043$). Thus pairings which do not dissolve in TBA not only started off well, but also ended the cycle well. In the online appendix, we show that the results in Tables 2–4 are robust to different partitions of the data according to cycle length.

Taken to the extreme, the negative reinforcement, which seem to happen more frequently in IBA, can lead to a near complete coordination failure. Figure 2 depicts two such examples where players' capacities become stuck at very low levels for several periods. While this extreme coordination failure was rare in IBA (for cycles with 6 or more periods, 5 out of 105 pairings had instances in which both players chose a capacity of 10 or less for four or more consecutive periods), it did not occur in TBA and likely never would because a subject would dissolve before things got so bad.

Figure 2 Descent to Coordination Failure in IBA



4.3. Determinants of Capacity Choices: Lagged Choices and Outcomes

Traditional newsvendor experiments have identified a number of common biases (Schweitzer and Cachon 2000), with the two most prevalent being *mean anchoring* and *demand chasing*. We cannot test mean anchoring because all of our experiments had the same critical fractile (equal to 0.6). Yet

Table 5 Fixed Effects Regression: Effect of Lagged Variables on Capacity Choice

		IBA		TBA	
Lag Own Choice	New Cycle$\times k_{i,t-1}$	0.287^{**}	(0.059)	0.283^{*}	(0.090)
	Pairing Dissolved$\times k_{i,t-1}$			0.337^{***}	(0.047)
	Pairing Maintained$\times k_{i,t-1}$	0.313^{***}	(0.050)	0.330^{***}	(0.047)
Lag Match Choice	New Cycle$\times k_{3-i,t-1}$	0.094^{*}	(0.032)	0.149^{**}	(0.028)
	Pairing Dissolved$\times k_{3-i,t-1}$			0.190^{***}	(0.020)
	Pairing Maintained$\times k_{3-i,t-1}$	0.333^{***}	(0.026)	0.353^{***}	(0.009)
Lag Demand	New Cycle$\times D_{t-1}$	-0.034	(0.022)	-0.024	(0.024)
	Pairing Dissolved$\times D_{t-1}$			0.037^{**}	(0.012)
	Pairing Maintained$\times D_{t-1}$	0.017	(0.027)	0.015	(0.020)
	Cycle Index	0.321	(0.172)	0.098	(0.051)
	Cycle Period	-0.039	(0.059)	0.073	(0.032)
	Number of Pairings			-0.211^{*}	(0.080)
	New Cycle	15.753^{***}	(2.644)	12.334^{***}	(1.323)
	Pairing Dissolved			5.786^{***}	(0.763)
	Constant	10.569^{*}	(3.846)	8.473^{**}	(2.444)
	Observations	3276		6084	
	R^2	0.523		0.633	

Note 1: ***1%, **5%, *10% significance using standard errors clustered at the treatment-session level. Since this is a fixed-effects model with clustering at the treatment-session level, the $p-$values are very conservative.

we believe that it is unlikely to occur as the strategic uncertainty (due to our's being a two-player game), is likely to dominate any mean anchoring bias. Indeed, anchoring would predict average capacities between 50 and 60, while we find average capacities less than 40. On the other hand, subjects may be influenced by demand chasing (Bostian et al. 2008), which posits that a subject's choice in period t is positively correlated with demand in period $t-1$. In Table 5, we investigate possible lag effects in the subject's capacity choices using a fixed-effects model which is generally preferred to the random effects specification in panel data models with lagged dependent variables and long panels (Hyndman and Embrey Forthcoming). The variables Lag Own Choice ($k_{i,t-1}$), Lag Match Choice ($k_{3-i,t-1}$) and Lag Demand (D_{t-1}) are used to measure a one-period lag effect of the relevant variable. Because lagged actions should be less informative of current behavior at the start of each new cycle and after a pairing has dissolved, we interact the variables of interest with indicators for (1) the first period of a new cycle (*New Cycle*), (2) whether the pairing was dissolved (*Pairing Dissolved*) and (3) whether the pairing was maintained (*Pairing Maintained*). Table 5 clearly shows that lagged actions have a strong influence on behavior. Subjects' own capacity choices are very persistent, but the persistence does not vary much depending whether it was a new cycle, the pairing was maintained or the pairing was dissolved.

We also see that players' current capacity is positively associated with their match's previous capacity. As could be expected, the magnitude of the effect is stronger when the pairing has been maintained than at the start of a new cycle (IBA: $p = 0.02$; TBA: $p < 0.01$) or following the dissolution of a pairing in TBA ($p < 0.01$). Table 5 also allows us to determine whether subjects chase demand. As can be seen, there is no evidence for demand chasing in the IBA treatment and very weak evidence for it in the TBA treatment, having a significant effect only after a pairing dissolved. Finally, consistent with our finding that lagged outcomes matter less at the start of a new cycle or following a dissolution, we see that the coefficients on "New Cycle" and "Pairing Dissolved" are positive and significant. This suggests that subjects reset their capacities following each rematching. Interestingly, in TBA the reset is significantly larger ($p < 0.01$) at the start of a new cycle rather than following a dissolution. This suggests that subjects perceive the rematching pool *within a cycle* to be worse than the pool of all subjects, an issue we will return to below.

4.4. The Decision to Dissolve a Relationship

We now turn our attention to why relationships break up in the TBA treatment. The most logical reason to dissolve a pairing is that the players' capacity choices are misaligned since this negatively affects the profits of the subject choosing the higher capacity. In Figure 3, we plot the frequency that subjects dissolved for each range of one-sided misalignment defined as Own Choice − Match's Choice. We treat misalignment equal to 0 (i.e., perfectly aligned capacities) as a distinct bin because it is the modal level of misalignment and also because there appears to be a discrete increase in the likelihood of dissolution when misalignment is anything other than 0: When the subjects are perfectly aligned, the likelihood of dissolution is 4.4%. However, for misalignment in the range $[-5, 0) \cup (0, 5]$, the likelihood of dissolution jumps to approximately 20%, and conditional on *some* misalignment, the average frequency of dissolution is 28%.

The figure depicts an asymmetric relationship between misalignment and the frequency of requesting a dissolution depending on whether misalignment is positive or negative. Remember that, all else equal, a subject's profit is non-decreasing in her match's capacity choice. When misalignment is positive – meaning that the match chose a lower capacity than the subject – the average frequency of dissolution requests is 34.2% and it increases sharply as misalignment increases. On the other hand, when misalignment is negative – meaning that the match chose a higher capacity than the subject – the average frequency of dissolution is significantly lower at 21.8% ($p < 0.01$), and the rate of increase as misalignment becomes more negative is also smaller.

The sharp rate of increase in the frequency of dissolution requests when misalignment is positive is intuitive because the subject suffers negative payoff consequences from their match's lower capacity choice. In contrast, it is surprising that subjects request to dissolve over 20% of the time when they chose a lower capacity than their match, since there are no adverse payoff consequences in the current round. It is possible that the subject expects her match to react to the low profits she

Figure 3 **Misalignment and the Likelihood of Requesting to Dissolve**

likely received in the current period by dropping her capacity in the next one and therefore the subject prefers to pre-emptively dissolve to avoid this.

Table 6 takes a regression approach to learn more about the motivations of a subject to dissolve a relationship. Not surprisingly, subjects are significantly more likely to dissolve the higher was their own capacity and are significantly less likely to dissolve the higher was their match's capacity. Note also that the effect of the match's capacity is more than twice as large as the subject's own capacity. We can also see that the probability of dissolving a relationship declines significantly the longer the relationship has been intact — there is a large drop if the pairing lasted two periods ($\mathbf{1}[\text{Pairing Duration} = 2]$), followed by a more continuous decline for pairings lasting 3 or more periods ($\mathbf{1}[\text{Pairing Duration} \geq 3] \cdot \text{Pairing Dur.}$).

Table 6 also allows us to investigate the role of luck (i.e., the random nature of demand in the stage game) in the decision to dissolve a relationship through the effect of realized profits. As can be seen, the coefficient on current profits is negative and significant; that is, given fixed capacity choices by the two players, the higher the realized profit, the less likely they are to dissolve the relationship. For example, consider two otherwise identical pairings. In both pairings, both players choose the Pareto efficient capacity of 60; however, in the first pairing, the realized demand was 60, while in the second pairing, the realized demand was 1. Although subjects in both pairings made the same decision, the realized profit for each subject in the first pairing is 180, while it is -115 in the second pairing. Each subject in this second pairing is 20.5 percentage points more likely to

Table 6 The Decision to Dissolve

	Random Effects		Fixed Effects		R.E. Logit	
Own Choice	0.003^{***}	(0.001)	0.004^{**}	(0.001)	0.009	(0.006)
Match's Choice	-0.008^{***}	(0.002)	-0.008^{**}	(0.002)	-0.058^{***}	(0.017)
Current Round Profit	-0.001^{***}	(0.000)	-0.001^{**}	(0.000)	-0.008^{***}	(0.002)
$\mathbf{1}$[Pairing Duration = 2]	-0.078^{***}	(0.009)	-0.063^{***}	(0.010)	-0.684^{***}	(0.094)
$\mathbf{1}$[Pairing Duration ≥ 3] $\cdot$ Pairing Dur.	-0.011^{***}	(0.001)	-0.009^{***}	(0.001)	-0.283^{***}	(0.035)
Cycle Index	-0.012^{***}	(0.001)	-0.012^{***}	(0.001)	-0.110^{***}	(0.017)
Pairing Number	0.005^{***}	(0.002)	0.002	(0.001)	0.024^{**}	(0.012)
Constant	0.470^{***}	(0.055)	0.443^{***}	(0.063)	0.017	(0.465)
Observations	6162		6162		6162	
R^2 / Log. Like	0.206		0.184		-1935.26	

Note: Robust standard errors (clustering at session level) in brackets. * $p < 0.1$; **: $p < 0.05$; ***: $p < 0.01$. For random-effects models, we included indicator variables for each session. For all models, we included indicator variables for four different regions where capacities fall: (i) $k_i = k_{3-i}$, (ii) $k_i < k_{3-i}$, (iii) $D \leq k_{3-i} < k_i$ and (iv) $k_{3-i} < \min\{k_i, D\}$, but omit them from the table.

dissolve relationship than subjects in the first pairing. Thus, above and beyond the decisions made by players in a pairing, (bad) luck influences whether or not the relationship lasts.

Lastly, we also see that the frequency of dissolving relationships declines across cycles (the coefficient on *Cycle Index* is negative and significant), but that, within a cycle, subjects who have experienced more pairings are more likely to dissolve (the coefficient on *Pairing Number* is positive and significant). This could be reflective of a subject type – i.e., a person who only wants one-shot relationships. We also investigated whether demographics, risk preferences or our other cognitive measures affected the decision to dissolve a relationship. Our results, which we omit in the interest of brevity, suggest that none have a significant effect on the decision to dissolve.

4.5. Do Subjects Dissolve Optimally?

While Table 6 provides the determinants of the decision to dissolve a pairing and suggests that subjects respond to random factors, it does not allow us to conclude whether subjects dissolve too much or too little. We can get a sense of this by comparing future profits depending on whether or not the subject chose to dissolve in the current period. Because profits also depend on other factors, such as absolute misalignment in the previous period (i.e., $|k_{i,t} - k_{3-i,t}|$), we consider the following simple fixed-effects regression:

$$\begin{aligned} \text{Future Profit}_{i,t+1} = \beta_0 + \beta_1 \cdot \texttt{maintain}_{i,t} + \beta_2 \cdot \texttt{maintain}_{i,t} \cdot |k_{i,t} - k_{3-i,t}| \\ + \beta_3 \cdot \texttt{dissolve}_{i,t} \cdot |k_{i,t} - k_{3-i,t}| + \mu_i + \epsilon_{i,t} \end{aligned} \quad (3)$$

We consider two measures of future profits: (i) the realized profit in the next period and (ii) the average profit over all future periods in the current cycle. The former informs us of the immediate

Table 7 Future Payoffs and The Decision to Dissolve

(a) Fixed Effects Regression

	Next Period Profit		Ave. Future Profit	
Maintain	22.177***	(2.923)	17.040**	(3.608)
Maintain × $\lvert k_i - k_{3-i}\rvert$	−0.754**	(0.205)	−0.616***	(0.090)
Dissolve × $\lvert k_i - k_{3-i}\rvert$	−0.346	(0.263)	−0.288	(0.175)
Constant	41.272***	(4.293)	47.518***	(3.608)
Observations	5538		5538	
R^2	0.045		0.094	

Note: Robust standard errors (clustering at session level) in brackets. * $p < 0.1$; **: $p < 0.05$; ***: $p < 0.01$.

(b) Average Future Profits & Rematching Over Cycle (Cycles Lasting 7+ Periods)

Cycle Per.	Dissolve	Maintain
1	49.36	56.86
2	44.71	58.09
3	41.12	57.05
4	40.98	62.71
5	41.19	60.25
6	41.75	61.74
7	41.72	61.74

cost/benefit of dissolving, while the latter examines the long term consequences, which may be different because the next relationship(s) may take time to develop.

The results are in Table 7(a). Observe that β_1 is significantly positive and β_2 is significantly negative. We also see that, for both measures of future profits, β_3 is not significantly different from 0. This makes sense because misalignment in period t should not affect earnings in future periods when the relationship dissolves. The implication of these results is that, unless misalignment is sufficiently large, it is better to maintain relationships, both in the short-run and overall. For example, using the point estimates, next period profits will be higher from dissolving if misalignment is greater than $\hat{\beta}_1/(\hat{\beta}_3-\hat{\beta}_2) = 54.36$, while average future profits will be higher from dissolving if misalignment is greater than 51.95. Looking at the marginal effect of maintaining a relationship at various levels of misalignment, we can say, first, that it is significantly better to maintain a relationship at the 5% level or better for misalignment less than or equal to 30 (using either measure of future profits) and second, it is never significantly better to dissolve, even for extremely high misalignment. (For example, to test for the significance of the marginal effect of maintaining when misalignment is equal to X, we conduct the hypothesis test: $\beta_1 + X(\beta_2 - \beta_3) = 0$ from (3).) Our results show that we can reject this hypothesis at the 5% level for $X < 30$, meaning that it is statistically better to maintain for misalignment 30 or less. Yet despite this, subjects choose to dissolve relationships 27.7% of the time when misalignment was between 1 and 45.

Table 7(b) takes a different approach. Specifically, it looks at average future profits (until the end of the cycle) depending on whether subjects dissolved or not in the first seven periods of a cycle. This gives us insight into the quality of the rematching pool. As can be seen, average future profits are always higher for the group of subjects who choose to maintain their relationship than those who request to dissolve. More interestingly, we see that the average future payoffs from maintaining increase modestly across cycle periods, while there is a substantial drop for those who dissolve between cycle periods 1 and 2 and 2 and 3. Thus, the rematching pool seems to deteriorate over time. To see this differently, for each cycle period, we can calculate the chance that a randomly drawn future payoff realization from the population who chose to dissolve exceeds a randomly drawn future payoff realization from the population who chose to maintain. For the first three cycle periods the chances are 39.6%, 34.3% and 31.9%, respectively. Thus, the later in a cycle that one dissolves, the lower are the chances that dissolving will lead to higher earnings.

4.6. Summary of TBA Versus IBA

Our results show that, on average, indefinitely binding agreements (IBA) lead to significantly higher profits. However, we identify several nuances. Indeed, the flexibility to dissolve relationships can be beneficial if used judiciously – subjects who only occasionally dissolve their relationships earn as much or more than the highest earning subjects in the IBA treatment. Yet our results suggest that subjects tend to dissolve relationships too much and not always for the right reason; in particular, (bad) luck appears to play a influence the decision to dissolve. We also provide evidence of the decreasing quality of the rematching pool over time, which further explains why subjects who dissolve very often do worse on average.

5. Robustness: What Drives the Poor Performance in TBA?

In this section we propose four (not mutually exclusive) explanations for the worse average performance of TBA vs IBA. We report the results of treatments 3–5, which we designed to analyze these explanations, as well as some additional data analysis.

5.1. Uncertainty About the Rematching Pool

A feature of the TBA institution is that subjects are rematched from within the sub-population of subjects whose pairings also dissolved in the previous period. Table 7 above suggests dissolutions which take place later in a cycle impact profits negatively more significantly than those which happen earlier, which could be due to the deteriorating quality of the rematching pool over time. While we consider this a realistic feature of the institution and not a design flaw, it is interesting to see how behavior changes when the benefits from dissolving are fixed and certain. Therefore, we conducted the TBA-FP treatment in which, at the end of every period, subjects could either maintain their relationship or terminate it and receive a fixed payment of 50ECU per period until the end of the cycle. Looking at Table 2 we see that this fixed payment amount of 50ECU falls between the average profit obtained per cycle by subjects who, in the original TBA treatment had 2 or fewer pairing and the average profit per cycle by subjects who had 3 or more pairings.

It also falls close to halfway between the expected profit of 91.5ECU associated with the Pareto-efficient Nash equilibria and the expected profits of 3 from the worst equilibrium. We conjecture that performance should be higher in TBA-FP than in TBA because (i) like the IBA institution, subjects will only have one pairing, providing stronger incentives to build a profitable relationship; (ii) unlike IBA, if the relationship sours, subjects can escape and get a guaranteed payoff; and (iii) since subjects do not enter a rematching pool that is worse than the overall population, they do not suffer the negative consequences of playing in this pool.

Table 8 Summary Statistics for TBA-D and TBA-FP (Average Capacity, Misalignment and Profits)

	Average Capacity	Average Misalignment	Average Profit	Average Pairing Duration	% Pairings Never Dissolved
TBA	33.72	9.59	51.38	2.24	27.24‡
TBA-FP	43.79	7.87	65.75♯	9.88	75.00
TBA-D†	43.97	9.31	65.08	3.46	36.84
$p-$value of test: TBA vs. Other Treatment:					
TBA-FP	< 0.001	0.135	0.001	< 0.001	< 0.001
TBA-D	< 0.001	0.695	0.001	< 0.001	0.010

Note 1: $p-$values are generated from random-effects regressions (with standard errors corrected for clustering at the treatment-session level), where we include indicator variables as the relevant explanatory variables. In all cases, we also include indicator variables for the different sessions.
Note 2: ♯ This includes all periods, even when subjects are receiving the fixed payment after dissolving. We believe this accurately reflects the institution. However, the results are even stronger if these cases are omitted.
Note 3: † For the TBA-D treatments, for the Average Pairing Duration and % Pairings Never Dissolved, we omit those cases in which a subject attempted to dissolve but was not allowed.
Note 4: ‡ This number is slightly different from the number reported in Table 2 because, in that table, a subject would be counted as having only one pairing if they dissolved in the period in which the cycle ended. Thus they had no chance to be rematched. In this table, to be consistent with TBA-FP, we would count such a subject as having dissolved.

Table 8 provides summary statistics on various performance metrics. As can be seen, subjects in TBA-FP choose significantly higher capacities, are better aligned and earn significantly higher profits. We also see that subjects maintain their pairings substantially longer (almost 4 times as long) and that fully 75% of the pairings never dissolved, compared with only 27.24% of initial pairings that never dissolved in TBA. Thus, players appear to be more patient with their match, which allows them to build a profitable relationship.

Remark 3. In this treatment, because 50 was the fixed termination payoff, one might expect this number to become salient in the decision to dissolve – even though subjects should not be concerned with profit realizations due to low demand. This is indeed the case as we observe that subjects were significantly more likely to dissolve if their realized profit in that period was less than 50. This further substantiates our results from TBA that one potential drawback of flexibility is that subjects react to bad luck.

Table 9 **Average Future Profits & Rematching Over The Cycle (Cycles Lasting 7 or More Periods) in TBA-D**

Cycle Per.	Dissolve (Allowed)	Previously Denied	Maintain
1	58.80	67.52	69.29
2	58.53	48.82	69.78
3	61.97	45.36	66.12
4	64.55	55.23	72.68
5	60.42	48.45	70.78
6	55.12	43.05	72.26
7	70.73	49.94	70.24

5.2. Reducing the Flexibility to Dissolve

We have established that subjects tend to dissolve too much and this is one of the reasons why the overall performance under TBA is worse than under IBA. In this section, we provide further evidence by considering a new treatment where we reduce subjects' ability to dissolve relationships. In the TBA-D treatment, subjects were told that there was a 10% chance that each pairing would be unbreakable and therefore would last for an indefinite duration. Subjects would only learn they were actually indefinitely bound to their match following a request to dissolve which was then denied by the software. Note that, in this event, the subject's match **was not** notified that a dissolution request was made but denied.

Table 8, shows that subjects in the TBA-D treatment choose significantly higher capacities and earn significantly higher payoffs than in the TBA treatment. We also see that the average pairing duration (conditional on being allowed to dissolve) is 1.22 periods longer than in TBA and that almost an extra 10% of the pairings never dissolve. Both of these are significantly different from TBA. Table 9 replicates Table 7(b) in reporting the average future earnings depending on which state a player is in the current cycle period. In the TBA-D treatment, there are three such states: (i) the subject requested a dissolution which was granted, (ii) the subject had a dissolution request denied either in the current or an earlier period or (iii) the subject willingly chose to maintain the relationship. Consistent with Table 7(b), subjects whose pairing dissolved earn less than subjects who willingly maintain their relationship. Even more interesting, we see that subjects whose attempt to dissolve was denied earn the least of all three groups. This suggests that, rather than encouraging long run behavior, finding out that one's dissolution request was denied leads to a souring of the relationship. Therefore, we propose that limiting flexibility by denying some dissolution requests increases profits because it forces subjects to work harder to make their relationship work since a failed relationship either leads to a rematching pool which is worse, or to being stuck in a relationship which becomes dysfunctional.

Table 10 The Prevalence of Teaching (i.e., $k_i > k_{3-i}$ for T or More Consecutive Periods)

	Number of Periods, T, of Teaching			
Treatment	2	3	4	5
IBA	0.264	0.176	0.106	0.069
TBA	0.182	0.094	0.043	0.029
$p-$value	0.011	0.005	0.014	0.040

Note: The table reports the fraction of times a subject chose a higher capacity than her match for at least T consecutive periods. The $p-$value is derived from the results of a paired $t-$test, using the session average as the unit of observation.

Remark 4. In an online appendix, we provide further analysis of our TBA-FP and TBA-D treatments. In terms of cycle dynamics, we show that behavior in the TBA-FP treatment shares similarities with both the IBA and TBA treatments, which makes sense because like IBA, subjects have only one partner per cycle but like TBA, subjects can dissolve that relationship. The TBA-D treatment is more closely aligned with the TBA treatment. In both treatments, we show that the decision to dissolve is less influenced by one's own capacity, their match's capacity and profit realizations (though with an interesting nuance in TBA-FP). Moreover, risk preferences do not strongly influence the decision to dissolve.

5.3. Sophisticated Teaching

Sophisticated teaching, which we define as the act of trying to influence one's match to take "better" actions, is like an investment. It involves paying a short-term cost – in the form of taking a sub-optimal action from a myopic perspective – in order to try to shift the future behavior of one's match such that payoffs will be higher in the future. Past research (Terracol and Vaksmann 2009, Hyndman et al. 2012) comparing fixed and random matching shows that players are less willing to invest in teaching under random matching (i.e., when new pairings are formed in each period). This is because they are less likely to reap the benefit of such an investment. The same logic suggests that the relative incentives for teaching are stronger in IBA since the partners in a pairing know they are bound together for an indefinite amount of time.

We define a *teaching episode* as a series of at least T consecutive periods wherein a subject chose a higher capacity than their match, and the pairing was maintained for all T periods. For example, suppose that there were 5 periods and $T = 3$. Suppose that a player chooses a higher capacity than her match in periods 2, 3 and 4 (but lower in periods 1 and 5). Periods 2, 3 and 4 constitute a teaching episode and the frequency of teaching is 60%. Table 10 contains results on the frequency of teaching given this definition as we vary T. For all values of T we considered, such behavior is approximately twice as likely to occur in the IBA treatment compared to the TBA treatment.

Table 11 Do Subjects Respond to Teaching? (Fixed Effectsl Dep. Var.: $\Delta k_{i,t}$)

	IBA		TBA	
Number of Period Match Teaches	1.537^{**}	(0.461)	2.047^{**}	(0.444)
Constant	-1.800^{**}	(0.386)	-2.010^{***}	(0.292)
Observations	2982		3408	
R^2	0.030		0.042	

Note: Robust standard errors (clustering at session level) in brackets. * $p < 0.1$; **: $p < 0.05$; ***: $p < 0.01$.

Table 12 Average Profits Pre-Teaching, While Teaching and Post-Teaching

	Average Profits			$p-$values		
Treatment	Pre-Teaching	Teaching	Post-Teaching	Pre v. T	T v. Post	Pre v. Post
TBA	43.82	52.37	57.53	0.003	0.444	0.120
IBA	48.41	40.34	64.38	0.043	0.012	0.047

Note: The $p-$value is derived from the results of a paired $t-$test, using the session average as the unit of observation.

Table 11 looks at the responsiveness to teaching, showing the results of a regression in which the dependent variable is the change in capacity between two periods and the explanatory variable is the lagged number of periods in which their match has chosen a higher capacity. In both the IBA and TBA treatments, the coefficient on one's match's teaching is positive and significant at the 5% level (and the coefficients are not statistically different across treatments). This suggests that, in both IBA and TBA, subjects respond to the teaching efforts by their matches by eventually raising their capacity levels.

Finally, we ask whether teaching is profitable. Consider (Subject, Cycle) pairs in which a *teaching episode* lasting three or more periods occurs (during which the pairing is maintained). There are potentially three stages for the cycle: (i) a pre-teaching phase, (ii) the teaching episode and (iii) a post-teaching phase (during which the pairing may even dissolve). Given the short-term investment/long-term reward nature of teaching, if teaching is successful, we would expect $\bar{\Pi}^{\text{post-teach}} > \bar{\Pi}^{\text{pre-teach}} > \bar{\Pi}^{\text{teaching}}$.

Table 12 shows that average profits are highest in the post-teaching phase of the cycle. For the IBA treatment, they are significantly higher than in both other phases. For the TBA treatment, we have the directionally correct comparative static but the result is not significant. Thus is appears that teaching pays; that is, it increases the profits of the teacher.

In conclusion, our data suggest that subjects engage in teaching behavior and that they do so more in IBA than in TBA. Further, in both treatments, teaching is effective and profitable.

Remark 5. We also note that teaching is more prevalent in the TBA-FP and TBA-D treatments than in the TBA treatment, which is not surprising given the longer average pairing length.

Moreover, we largely replicate the results from Tables 11 and 12 for these treatments.

5.4. Endogenous Institution Choice

Finally, we turn to the issue of whether endogenous selection into an institution can help overcome the performance penalty of the TBA institution relative to IBA. In our institution choice (IC) treatment, at the beginning of each cycle, all subjects chose whether they wanted to be bound with the same person until the random termination of the cycle (i.e., IBA), or whether they wanted the option to dissolve a relationship at the end of every period (i.e., TBA).

Table 13 Determinants of Institution Choice

(a) TBA Choice

Cycle	Freq.
1	0.789
2	0.724
3	0.618
4	0.579
5	0.671
6	0.632
7	0.553
8	0.566

(b) Random-Effects Regression (Dep. Var.: Choose TBA)

	(1)		(2)	
Willing. to Take Risk (Incent.)	-0.017^{**}	(0.009)	-0.014^{**}	(0.006)
Cognitive Reflection Test	-0.061^{***}	(0.013)	-0.052^{***}	(0.014)
Need for Cognition	-0.001	(0.004)	0.000	(0.001)
Grade Point Average	-0.178^{**}	(0.087)	-0.118^{**}	(0.048)
Male	-0.031	(0.078)	-0.007	(0.060)
Cycle Index	-0.027^{***}	(0.006)	-0.003	(0.010)
(Lag IBA) × (Lag Ave. Profit)			-0.004^{***}	(0.001)
(Lag TBA) × (Lag Ave. Profit)			0.003^{***}	(0.001)
Constant	1.626^{***}	(0.285)	1.226^{***}	(0.155)
Observations	608		532	
R^2	0.094		0.342	

Note 1: Robust standard errors (clustering at session level) in brackets. * $p < 0.1$; **: $p < 0.05$; ***: $p < 0.01$.
Note 2: The "Risk Willingness" is measured such that higher numbers indicate a greater willingness to take risk.

In Table 13(a) we show the frequency with which subjects chose the TBA institution for each cycle. In the first cycle, almost 80% of subjects chose the TBA institution. This is remarkable because our previous analysis showed that the IBA institution led to higher profits than the TBA institution. However, as we can also see, there is a consistent move towards IBA in later cycles. Table 13(b) shows that there are, in fact, differences between the type of subjects who choose the TBA and IBA institutions. Interestingly, subjects who are more willing to take risks are significantly less likely to choose TBA. Thus, the inflexibility of the IBA institution is perceived as risky. This finding is consistent with Hyndman et al. (2014) who showed, in a finite horizon game, that payoffs were more variable under fixed than random matching. (However, since neither matching protocol from Hyndman et al. (2014) are identical to either the TBA or IBA institutions, some caution is warranted in comparing the results.) Two of our three measures of cognitive ability (GPA and Cognitive Reflection Test) also indicate that higher scoring subjects are less likely to choose TBA.

Finally, the coefficient on the cycle index is significantly negative. The second column, which includes average profits from the previous cycle interacted with previous institution shows that the trend towards the IBA institution is driven by past experience with the institution. Specifically, subjects are more likely to switch institutions the lower were their profits in the previous cycle. All of the demographic variables maintain their sign and significance in column (2).

We now compare the (exogenously imposed) IBA and TBA treatments with our IC treatment, in which subjects choose between IBA and TBA. Table 14 reports summary statistics depending on the institution and whether the institution was chosen or exogenously imposed. Subjects who choose into their institution appear to choose a significantly higher capacity than their counterparts when the same institution was exogenously imposed. There does not appear to be any significant difference in the degree of misalignment. Combining these two results, we see that profits are significantly higher in our institution choice treatment. This suggests that there may be a signaling effect to endogenous institution choice: subjects who choose IBA know that they will be matched with other subjects who also chose the IBA institution. This may signal a mutual commitment to building successful long term relationships. Similarly, subjects who choose TBA know that they will be paired with subjects who also chose the TBA institution. Thus it is clear that they can replace a poor performing partner, or see themselves be replaced, which may push them to choose higher capacities. Fan et al. (Forthcoming) find evidence for endogenous selection as a signaling device in a related context. Specifically, subjects who select into a group with a membership fee subsequently exert higher effort in a minimum effort game.

Table 14 Comparing Institution Choice and Exogenous Institution Treatments

	In TBA Institution			In IBA Institution		
	Chosen (IC)	Imposed (TBA)	$p-$value	Chosen (IC)	Imposed (IBA)	$p-$value
Capacity	41.28	33.72	0.000	40.65	37.05	0.000
Misalignment	8.66	9.59	0.838	8.36	9.39	0.455
Profits	61.68	51.38	0.001	63.54	56.48	0.017

Note: $p-$values measure statistical differences between the chosen and imposed institution through the significance of the coefficient of an indicator variable for the IC treatment in random effects regressions with standard errors clustered at the treatment-session level.

Although allowing for endogenous selection leads to higher (and approximately equal) average payoffs for subjects in both institutions, some of the disadvantages of the TBA institution are still present in the IC treatment. First, there is a strong negative relationship between average payoffs and the number of pairings a subject had. Thus, while there is positive signaling value to endogenous institution choice, it does not change the fact that subjects who choose flexible relationships appear to dissolve too much, which is to their detriment relative to subjects who dissolve less frequently. Moreover, subjects who chose into the TBA institution are significantly more likely to dissolve (by 8 percentage points) than subjects who were exogenously placed into it. The combination of these two effects explains why subjects perform on average worse in the IC treatment compared those who chose into the IBA institution, yet better than when the TBA institution was exogenously imposed.

6. Concluding Remarks

In this paper, we compare the behavior of subjects operating under two different institutional arrangements. In our indefinitely binding agreements (IBA) treatment, subjects play a stage game in fixed pairs of two over cycles lasting an indefinite number of periods, while, in our temporarily binding agreements (TBA) treatment, they can request to dissolve the existing relationship at the end of each period and be rematched with a different subject in the next period. The stage game is a two-person newsvendor game, which is a coordination game with multiple Pareto rankable equilibria where payoffs are increasing in the subjects' actions (*capacities*) as well as in an exogenous random component (*demand*).

On average, we find that subjects in the IBA institution earned about 10% more than their TBA counterparts, and this is largely attributable to them choosing higher average capacities under IBA. However, even though the *average* performance is worse in TBA, average payoffs are actually significantly higher for the near-30% of subjects who remained matched with the same person over the entire cycle in the TBA treatment. Moreover, these long-lasting relationships benefit from positive reinforcement, i.e., capacities generally increase over the cycle and come closer to the efficient level. In contrast, the average tendency in IBA was towards negative reinforcement, with capacities declining over the course of the cycle.

The reason that the TBA institution performs, on average, worse than the IBA institution is partly driven by subjects over-using their flexibility to dissolve. While subjects often dissolve relationships due to misalignment (i.e., large differences in their chosen capacity levels), they also respond to random factors, that is a low realization of demand. Additionally, subjects do not seem to recognize that the rematching pool is worse than the overall subject pool. Indeed, except for extreme levels of misalignment, our results suggest that subjects would be better off maintaining their relationship, rather than dissolving it. Two of our follow-up treatments – TBA-FP and TBA-D – which modified the TBA institution by replacing the rematching pool with a fixed, exogenous termination payoff (TBA-FP) or by denying some dissolution requests (TBA-D) led to fewer dissolution requests, longer-lasting relationships, higher capacity decisions and, therefore, higher profits – even higher than in IBA. Thus, an important result of our paper is that unfettered flexibility to dissolve relationships is detrimental. Instead, a more limited flexibility, which still encourages long-run thinking may be advantageous.

We also showed that far-sighted teaching (the mechanism by which a subject encourages her match to raise his capacity by consistently playing high capacity values herself) is less frequent in TBA than in IBA. Subjects respond to the teaching efforts of their match approximately equally in both treatments and these efforts turn out to be equally beneficial, so the fact that there is less of it in TBA leads to worse subject performance.

Finally, we show that endogenous selection into an institution is profit-enhancing and eliminates the performance difference between TBA and IBA. Interestingly, we find that most subjects opt for the flexibility of TBA with a trend towards the IBA institution in later cycles that appears to be driven by their experience in the previous cycle. We also show that there are clear differences in the type of subjects who prefer each institution. The IBA institution is viewed as risky with more risk-averse subjects opting for TBA and, interestingly, subjects with higher self-reported GPAs and who scored higher on the Cognitive Reflection Test, were more likely to choose IBA. Thus, a plausible interpretation of these results is that endogenous selection into an institution is a way for subjects to signal their attitude towards commitment in a relationship, which reduces strategic uncertainty and allows for more profitable interactions.

The overall lesson of our paper to managers is that both the TBA and IBA institutions offer advantages and disadvantages. Those groups in TBA who were able to build and maintain successful long-term relationships performed, by far, the best of all subjects. Furthermore, TBA does appear to facilitate sorting amongst players, which helps to ensure that high quality players are eventually matched. This is even more likely to be the case when subjects can choose the institution (TBA or IBA) under which they operate. On the other hand, it appears to be the case that the IBA institution incentivizes long-run thinking, which is advantageous overall. Moreover, it prevents groups from dissolving because of bad luck. In practice, there may be factors specific to the environment that suggest whether flexibility is desirable or whether it is best to commit to a long-lasting relationship. An important corollary is that endogenous selection into an institution – which, arguably, best-fits the real world – signals something about one's type. The effect of this signal reduces the performance gap between TBA and IBA but does not completely eliminate all of the negative aspects of the TBA institution. When forming relationships, managers should think carefully about what the proposed scope of the relationship says about them and what type of partner(s) it attracts.

There are several interesting areas for future research. First, one feature which is missing from our experiments is that, when first matched, subjects have no information about the player they are matched with. In reality, partnerships are generally not initiated at random but are created by the conscious decisions of the parties involved. It would be interesting to see whether an explicit group formation stage, similar to Riedl et al. (2016), would increase the overall profitability of the TBA institution. Second, because we kept the expected cycle length equal across all treatments, we are not able to study the impact of the perceived expected length of the relationship on the subjects' behavior, which would be interesting to analyze. Third, it may also be interesting to consider a hybrid version of our TBA and TBA-FP institutions in which the subjects are given three options: (i) maintain their relationship, (ii) dissolve and enter the rematching pool or (iii) dissolve and take an exogenous termination payoff. This setting may yield interesting insights into how subjects perceive the quality of the rematching pool. Another interesting avenue would be to consider a treatment in which subjects have to pay a fee to dissolve a relationship, as is the case in many real-life partnerships (e.g., lawyer fees, severance packages, etc.).

Acknowledgments

We would like to thank conference participants at the Economic Science Association (2013 North American Meeting). We also thank Matthew Embrey and Guillaume Fréchette for valuable comments, and Malcolm Kass for his assistance in running the experiments. Kyle Hyndman gratefully acknowledges the National Science Foundation (SES-1025044) and the Institute for Research in Experimental Economics (IFREE, Small Grants Program) for financial support. Both Kyle Hyndman and Dorothée Honhon thank the University of Texas at Dallas for financial support.

References

Bamford, James, David Ernst, David G. Fubini. 2004. Launching a world-class joint venture. *Harvard Business Review* February.

Beer, Ruth, Hyun-Soo Ahn, Stephen Leider. 2017. The informational and incentives effects of supplier awards. Working Paper.

Bernard, Mark, Jack Fanning, Sevgi Yuksel. 2014. Finding cooperators: Sorting through repeated interaction. Working Paper.

Bernstein, Fernando, Gregory A. DeCroix. 2004. Decentralized pricing and capacity decisions in a multitier system with modular assembly. *Management Science* **50**(9) 1293–1308.

Bolton, Gary, Claudia Loebbecke, Axel Ockenfels. 2008. Does competition promote trust and trustworthiness in online trading? an experimental study. *Journal of Management Information Systems* **25**(2) 145–169.

Bostian, AJ A., Charles A. Holt, Angela M. Smith. 2008. Newsvendor "pull-to-center" effect: Adaptive learning in a laboratory experiment. *Manufacturing & Service Operations Management* **10**(4) 590–608.

Brandts, J., D. Cooper. 2006. A change would do you good: An experimental study of how to overcome coordination failure in organizations. *American Economic Review* **96**(3) 669–693.

Cachon, Gerard P., M. Lariviere. 2001. Contracting to assure supply: How to share demand forecasts in a supply chain. *Management Science* **47**(5) 629–646.

Cacioppo, John T., Richard E. Petty. 1982. The need for cognition. *Journal of Personality and Social Psychology* **42**(1) 116–131.

Camerer, Colin F., Teck-Hua Ho, Juin-Kuan Chong. 2002. Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. *Journal of Economic Theory* **104**(1) 137–188.

Center for Digital Strategies. 2006. *Strategic Partnering: Managing Joint Ventures and Alliances (Thought Leadership Roundtable on Digital Strategies)*. Tuck School of Business, Dartmouth University.

Davis, Andrew, Kyle Hyndman. 2018. An experimental investigation of managing quality through monetary and relational incentives. *Management Science* **64**(5) 2345–2365. Http://dx.doi.org/10.1287/mnsc.2016.2716.

D'Evelyn, Sean. 2013. Dump, date or marry: Endogenous group formation with varied contract length. Working Paper.

Fan, James, Anthony Kwasnica, Douglas Thomas. Forthcoming. Paying for teamwork: Supplier coordination with endogenously selected groups. *Production and Operations Management* .

Frederick, Shane. 2005. Cognitive reflection and decision making. *Journal of Economic Perspectives* **19**(4) 25–42.

Grundberg, Sven, John D. Stoll. 2012. Nokia sales face more hurdles. The Wall Street Journal, `http://www.wsj.com/articles/SB10001424052702304898704577480340752531300`.

Han, S.L., D.T. Wilson, S.P. Dant. 1993. Buyer-supplier relationships today. *Industrial Marketing Management* **22**(4) 331–338.

Hendricks, K. B., V. R. Singhal. 2005. Association between supply chain glitches and operating performance. *Management Science* **51**(5) 695–711.

Honhon, Dorothée, Kyle Hyndman. 2016. Flexibility and reputation in repeated prisoner's dilemma games. Working Paper.

Hyndman, Kyle, Matthew Embrey. Forthcoming. The econometrics of experiments. Stephen Leider Elena Katok, Karen Donohue, eds., *Handbook of Behavioral Operations Management*. Wiley.

Hyndman, Kyle, Santiago Kraiselburd, Noel Watson. 2013. Aligning capacity decisions in supply chains when demand forecasts are private information: Theory and experiment. *Manufacturing & Service Operations Management* **15**(1) 102–117.

Hyndman, Kyle, Santiago Kraiselburd, Noel Watson. 2014. Coordination in games with complementarities: An experiment on fixed vs random matching. *Production and Operations Management* **23**(2) 221–238.

Hyndman, Kyle, Erkut Özbay, Andrew Schotter, Wolf Ehrblatt. 2012. Convergence: An experimental study of teaching and learning in repeated games. *Journal of the European Economic Association* **10**(3) 573–604.

Hyndman, Kyle, Antoine Terracol, Jonathan Vaksmann. 2009. Learning and sophistication in coordination games. *Experimental Economics* **12**(4) 450–472.

Kalwani, M.U., N. Narayandas. 1995. Long-term manufacturer supplier relationships: do they pay off for supplier firms? *Journal of Marketing* **59**(1) 1–16.

Knez, Mark, Colin F. Camerer. 1994. Creating expectational assets in the laboratory: Coordination in 'weakest-link' games. *Strategic Management Journal* **15** 101–119.

Kovach, Steve. 2014. Microsoft closes its $7.2 billion purchase of nokia. Business Insider, `http://www.businessinsider.com/microsoft-closes-nokia-acquisition-2014-4`.

KPMG. 2009. *Joint Ventures: A tool for growth during an economic downturn*. KPMG International.

Kwon, H. Dharma, Steven A. Lippman, Kevin F. McCardle, Christopher S. Tang. 2010. Project management contracts with delayed payments. *Manufacturing & Service Operations Management* **12**(4) 692–707.

Larsen, M. M., T. Pedersen, D. Stepniov. 2010. *LEGO Group: An Outsourcing Journey*. Case 9B10M094, Ivey Management Services.

Liker, Jeffrey K., Thomas Y. Choi. 2004. Building deep supplier relationships. *Harvard Business Review* December, 104–113.

Lunsford, J. L. 2007. Boeing scrambles to repair problems with new plane. *Wall Street Journal* (Dec. 7).

Neuville, J. 1997. La stratégie de la confiance: Le partenariat industrial observé depuis le fournisseur. *Sociologie du Travail* **20**(3) 297–319.

Özer, Ö., Yanchong Zheng, Kay-Yut Chen. 2011. Trust in forecast information sharing. *Management Science* **57**(6) 1111–1137.

Page, Talbot, Louis Putterman, Bulent Unel. 2005. Voluntary association in public goods experiments: Reciprocity, mimicry and efficiency. *Economic Journal* **115** 1032–1053.

Riedl, Arno, Ingrid M.T. Rohde, Martin Strobel. 2016. Efficient coordination in weakest-link games. *Review of Economic Studies* **83**(2) 737–767.

Roth, Alvin E., J. Keith Murnighan. 1978. Equilibrium behavior and repeated play of the prisoner's dilemma. *Journal of Mathematical Psychology* **17**(2) 189–198.

Schweitzer, Maurice E., Gerard P. Cachon. 2000. Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science* **46**(3) 404–420.

Shokoohyar, Sina, Anyan Qi, Elena Katok. 2016. Project management under risk sharing. Working Paper.

Tang, Christopher S., Joshua D. Zimmerman, James I. Nelson. 2009. Managing new product development and supply chain risks: The boeing 787 case. *Supply Chain Forum: An International Journal* **10**(2) 74–86.

Terracol, Antoine, Jonathan Vaksmann. 2009. Dumbing down rational players: Learning and teaching in an experimental game. *Journal of Economic Behavior & Organization* **70** 54–71.

Tomlin, B. 2003. Capacity investments in supply chains: Sharing the gain rather than sharing the pain. *Manufacturing & Service Operations Management* **5**(4) 317–333.

Uzzi, B. 1996. The sources and consequence of embeddedness for the economic performance of organizations: the network effect. *American Sociological Review* **61**(4) 675–698.

van Huyck, John, Raymond Battalio, Richard Beil. 1990. Tacit coordination games, strategic uncertainty and coordination failure. *American Economic Review* **80**(1) 234–248.

Wang, Y., Y. Gerchak. 2003. Capacity games in assembly systems with uncertain demand. *Manufacturing & Service Operations Management* **5**(3) 252–267.

Wilson, Alistair J., Hong Wu. 2017. At-will relationships: How an option to walk away affects cooperation and efficiency. *Games and Economic Behavior* **102** 487–507.

Xu, Xin, Yao Zhao. 2013. Incentives and coordination in project-driven supply chains. Working paper.