# A simple experimental test of the Coase conjecture: fairness in dynamic bargaining

Jack Fanning and Andrew Kloosterman

December 18, 2017

**Abstract**

In each round of an infinite horizon bargaining game, a proposer proposes a division of chips, until a responder accepts. The Coase conjecture predicts that incomplete information about responders' preferences for fairness will lead to almost immediate agreement on an equal payoff split when discounting between rounds is small. We experimentally test this prediction when chips are equally valuable to both bargainers and when they are worth three times as much to proposers, and compare outcomes to an ultimatum game. Behavior offers strong support for the theory. In particular, when chips are more valuable to proposers, initial offers, initial minimum acceptable offers, and responder payoffs are significantly higher in the infinite horizon than in the ultimatum game, while proposer payoffs are significantly lower.

The Coase conjecture (Coase (1972)) is central to modern bargaining theory. It predicts that a seller who faces a buyer with private information about her value will sell her good almost immediately at close to the buyer's lowest possible value when both agents are patient.[1,2] The seller's problem is that she must compete against her future self in an infinite horizon. For any price she offers today, high value buyers will have the greatest incentive to accept. If today's offer is rejected, therefore, the seller should likely infer that she faces a low value buyer, creating an incentive to cut her price tomorrow. But in which case high value buyers should only purchase if today's price is already low (this is known as demand withholding).

A rigorous theoretical proof of Coase's idea was ultimately provided by Gul, Sonnenschein, and Wilson (1986). It serves as a foundational result for many recent and more complex bargaining models (e.g. two sided reputational models in the style of Abreu & Gul (2000)). The core insight is ubiquitous: with patient bargainers, one-sided asymmetric information leads to almost immediate agreement favorable to the informed party (as though negotiation was exclusively with her "toughest" type).

Nonetheless, the theory is non-obvious. Another intuition is that asymmetric information will lead to inefficient delay, with an uninformed seller trying to hold out for a high price. Testing whether it holds in practice is therefore important. Laboratory experiments, providing a controlled environment, would seem ideally suited to that task. Discouragingly

[1] So long as she obtains a positive profit doing so, what is known as the "gap" case in the literature.

[2] In fact, Coase's original idea was that a durable goods monopolist facing downward sloping demand (an equivalent problem) would be forced to sell at (close to) marginal cost.

for the theory, results from the existing experimental literature have run almost entirely counter to even its qualitative predictions. In particular, initial prices are typically increasing in the discount factor and game's horizon (see our literature review at the end of this section for details). Indeed, Reynolds (2000) concludes by saying: "It appears doubtful that any experimental design would generate results consistent with the Coase conjecture."

One potential confound of such results is that there is an additional source of asymmetric information. An extremely well-established finding in the bargaining literature, exemplified by the simple ultimatum game, is that some people care about fairness. Subjects routinely reject offers which give them a much smaller payoff than their opponent (for a recent survey see Guth and Kocher (2014)). Preferences for fairness are heterogeneous and private, and therefore represent a second source of asymmetric information. Previous Coase conjecture experiments considered the canonical model of a seller facing a privately informed buyer, but their hypotheses were based on money maximizing subjects, and these need not hold even approximately in the presence of private information about fairness concerns. In particular, a seller who cares about fairness can resist rapidly dropping her price if she obtains sufficient disutility from selling to a high value buyer at an "unfair" low price.[3] Money maximizing sellers may then profitably imitate such behavior.

In this paper we provide a simple experimental test of the Coase conjecture by focusing exclusively on the asymmetric information about naturally occurring fairness preferences which has potentially confounded previous experiments. We consider an infinite horizon bargaining game, where in each round a proposer can propose any division of 100 chips between herself and a responder. If the responder accepts this offer the game ends, otherwise the value of chips to both parties are discounted by $\delta$ and the game continues into the next round. Each round, we use the "strategy method" to elicit responders acceptance decision by asking them for their minimum acceptable offer.[4]

If there is any positive probability that a responder is a fair type, that is, one who is unwilling to accept any unequal division, then for $\delta$ close to 1, the Coase conjecture predicts almost immediate agreement on an almost equal monetary payoff split (we choose $\delta = 0.95$). The translation of the Coasean reasoning is straightforward: if a proposer's offer of less than an equal monetary split is rejected, she should increase her belief that her opponent cares about fairness and so increase her offer in the next round. Anticipating this, even purely money motivated responders should reject offers that are not close to

[3]See Fanning (2014).

[4]We elicit responders' "strategy" in each round, not for the full infinite horizon game. We did this both in order to save time and to have a clear measure of responder demand witholding. Oosterbeek et al (2004) shows that this method tends to increase offers in ultimatum games, but, any such effect should not effect our comparisons of results across treatments that all use this same method.

equal monetary payoffs.

The difficulty of statistically testing this prediction is that it is somewhat imprecise (how close to an equal payoff split is close enough?). We therefore compare the infinite horizon game outcomes to those of ultimatum games (which are equivalent to the $\delta = 0$ case of our infinite horizon game).

The Coasean theory predicts that initial offers and responder payoffs should be weakly lower in the ultimatum game, proposer payoffs should be weakly higher, while initial minimum acceptable offers should be strictly lower (due to demand withholding). The weakly qualifier relates to the fact that if fair types are very likely then an equal payoff split offer can be money maximizing in the ultimatum game too. Indeed, the existing literature has documented that in ultimatum games where chips are equally valuable to proposers and responders (i.e. the standard ultimatum game), proposer offers are already close to an equal monetary payoff split.

Theory also predicts, however, that equal payoff split offers are less likely in ultimatum games when chips are worth more to the proposer than the responder (this increases the incentive to make an unequal offer). Moreover, that prediction is supported by previous experimental evidence from Kagel, Kim & Moser (1996). In addition to the standard equal chip value setting, therefore, we consider treatments where chips are worth three times as much to proposers as responders. We view this second comparison as offering the cleanest test of the theory where we expect the predictions noted above to hold strictly.

Our infinite horizon results are remarkable close to the predictions of the Coase conjecture. Regardless of chip values, average initial offers, minimum acceptable offers, proposer payoffs and responder payoffs were less than 9% away from an equal monetary division. Remarkably, more than 92% of final chip allocations were within 10% of an equal division.

When chips are equally valuable to both parties, the weak inequalities of our predictions are indeed weak: initial offers and payoffs in the infinite horizon are statistically indistinguishable from the ultimatum game. Nonetheless, initial minimum acceptable offers are significantly higher in the infinite horizon (in line with theory's strict prediction). When chips are worth more to proposers than responders, then all the treatment differences are strict: initial offers, initial minimum acceptable offers, responder payoffs and efficiency are significantly higher in the infinite horizon, while proposer payoffs are significantly lower. Moreover, the differences are qualitatively large, in particular, responder payoffs were 52% higher in the infinite horizon and proposer payoffs were 28% lower.

We believe these results represent strong evidence in favor of the Coase conjecture. In this simple bargaining environment, responders correctly understood that rejecting an unfair offer would most likely lead a proposer to raise her offer in the next round. Anticipating this, proposers made fair offers from the start of the game.

In addition to providing support for the theory generally, it also specifically suggests that fairness concerns may have large effects on bargaining outcomes more broadly. Even in settings where most people care little for fairness, the Coase conjecture implies that the mere possibility that they care can lead to an almost completely fair outcome (although of course determining what is fair may be more difficult in more complex settings). This may, for instance, help explain the findings of Cullen & Pakzad-Hurson (2017) in Task Rabbit data. When multiple workers privately contracted with an employer to do the same work in the same location, low paid workers frequently renegotiated their wages, and conditional on doing so obtained exactly the wage of the highest paid worker (the highest paid workers and those in different locations did not renegotiate). Co-location appears to have enabled workers to discover their peers' wages. Fairness concerns and the Coase conjecture can then explain why renegotiation fully equalized wages.

The remainder of this section discusses related literature. Section 2 then highlights our theoretical predictions, Section 3 provides details on the experimental design, Section 4 presents the results and Section 5 is a conclusion.

## Related literature

Most Coase conjecture experiments have found evidence that contradicts the theory's comparative statics.[5] Guth, Ockenfels & Ritzberger (1995) matched a seller to 10 buyers with uniformly distributed values, for either two or three trading rounds and three different discount factor combinations. Subjects were either inexperienced or had training on the theoretical model. Behavior violated theoretical comparative statics on the discount factor and prices were typically above the static monopoly level (i.e. above the single round commitment price, far above the Coasean prediction). Rapoport, Erev & Zwick (1995) paired a seller to a single buyer with uniformly distributed values in a long horizon setting and three different discount factors. Contrary to theory, initial prices were increasing in the discount factor and above the static monopoly level. Reynolds (2000) compared treatments where a seller faced either one buyer or five buyers with uniformly distributed values, and either one, two or six trading rounds (and a constant discount factor). Again, contrary to theory, initial prices increased in the trading horizon, although they were below the static monopoly benchmark. Cason and Reynolds (2005) paired a seller with a buyer holding either a high or low value, for either one or two trading rounds with four possible continuation probabilities, and imposed a restricted grid of possible offers. Initial offers didn't vary much with the continuation probability and not generally in the direction suggested by theory. The authors argue that quantal

[5]In addition to the one-sided asymmetric information experiments highlighted here, some experiments consider two sided asymmetric information (e.g. Embrey, Fréchette, and Lehrer (2015) investigate the reputational bargaining model of Abreu and Gul (2000)).

response equilibrium captures some important features of the results.

On the other hand, other experiments related to the Coase conjecture offer at least partial support for the theory. Cason and Sharma (2001) considered a certain demand setting where one seller faced two buyers, one of whom always had a high value and the other a low value, and an uncertain demand treatment where with small probability both buyers had the same value. The bargaining horizon was indefinite, with treatments also varying the probability that the game would continue into another round. Theory predicts high initial prices for certain demand regardless of continuation probability, and lower prices for uncertain demand, which are declining in the continuation probability (for Coasean reasons). However, the authors hypothesized that subjects would view both treatments as comparable and uncertain. Initial prices in both demand treatments were indeed closer to theory's predictions for uncertain demand. Those prices declined in the continuation probability for certain demand, in line with theory's prediction for uncertain demand, but they do not decline in the uncertain demand treatment. Guth, Kroger and Normann (2004) considered a two round bargaining game where a seller faces a buyer with uniformly distributed values, and both parties have private discount factors. They found support for theory's prediction that initial prices are *increasing* in the seller's discount factor.

Slembeck (1999) investigated a repeated ultimatum game with a fixed partner (and equally valuable chips for both players), that might be expected to share a similar Coasean prediction with our infinite horizon game. Subjects only played the repeated game once. Average offers across all twenty rounds were close to an equal split, although this was also true in to a control treatment where subjects were randomly rematched after each round. Rejection rates were significantly higher when partners were fixed, however, in line with our finding of higher minimum acceptable offers in the infinite horizon.

Related to our infinite horizon game with unequal chip values, Roth and Malouf (1979) considered a dynamic unstructured bargaining game where two subjects had to agree on a division of the probability of winning a prize, worth three times as much to one of the players. Both subjects could send freeform chat messages and propose any division at any time within 12 minutes. Agreements clustered around two distinct fairness norms, equal probability (50% probability for each subject) and equal expected payoffs (75% probability for the low prize subject). We did not expect an equal chip division in our setting to have a similar normative appeal to an equal probability of getting some prize (as opposed to nothing)[6], however, a different bargaining protocol may also explain our different results.

[6] Equal expected monetary payoffs may have appeared very unfair to risk averse high prize subjects, given their large chance of getting nothing.

## 2 Theory and hypotheses

We propose a very simple model, in which agents care about fairness to different extents. To simplify our analysis we consider a model with a continuous action space in which responders see proposers' offers before choosing their Minimum Acceptable (MA) offer. Nonetheless, it is possible to show that the conclusions still hold in a fine, discrete action space game with simultaneous choices (our actual game), in which there is a vanishingly small probability that agents tremble over their action choices.[7] Our solution concept is Perfect Bayesian equilibrium, which requires sequential rationality at all information sets and beliefs determined by Bayes' rule wherever possible. Off equilibrium path beliefs are unimportant for the analysis.

Each agent is a fair type with probability $\lambda$ and is otherwise normal. A fair type's utility function when she receives \$$x_i$ and her bargaining partner receives \$$x_j$, is $u_i^F(x_i, x_j) = x_i - x_j \mathbb{1}_{[x_j > x_i]}$. This ensures that a fair type prefers disagreement to any agreement that gives her opponent a higher monetary payoff. A normal (money maximizing) type's utility function is simply $u_i^N(x_i, x_j) = x_i$. It does not affect the analysis if proposers are assumed to always have normal preferences. An agreement is a division of chips for the proposer and the responder, $(c_P, c_R)$ where $c_P = 100 - c_R \in [0, 100]$. If agents reach an agreement $(c_P, c_R)$ in round $t \in \mathbb{N}$ (time is discrete), then monetary payoffs are $x_i = \delta^{t-1} e_i c_i$, where $e_i$ is agent $i$'s exchange rate of chips to money in round 1 and $\delta \in [0, 1)$. This is theoretically equivalent to having a fixed chip to money exchange rate and agents' who discount future payoffs exponentially using discount factor $\delta$. No agreement is captured by $t = \infty$.

In both the infinite horizon and ultimatum game, at the start of a round a proposer can make an offer, $c_{R,t} \in [0, 100]$ which the responder observes before accepting or rejecting. In the infinite horizon game, if an offer is accepted in round $t$, then the game ends, otherwise it proceeds to round $t + 1$. In the ultimatum game, the game ends after round 1 regardless.

Let $\bar{c}_R = \frac{100 e_P}{e_R + e_P}$ be the offer of chips which implies equal payoffs for both players, and let $\bar{m} = \bar{c}_R e_R = (100 - \bar{c}_R) e_P$ be the implied monetary payoff (so $\bar{c}_R = 50$ if $e_P = e_R$ and $\bar{c}_R = 75$ if $e_P = 3e_R$). The next proposition establishes the Coase conjecture for the infinite horizon game. It says that there is a (generically) unique equilibrium. Moreover, for any positive probability of fair types, $\lambda \in (0, 1]$, if players are sufficiently patient, $\delta \approx 1$, then there is almost immediate agrement on an equal payoff split. This is an approximately efficient outcome.

**Proposition 1.** *In the infinite horizon game there is a generically unique equilibrium*

[7] Without such trembles our use of the "strategy method" in each round prevents sequential rationality from having any bite.

*path of play. For any $\varepsilon > 0$ and any $\lambda \in (0, 1]$ there is some $\bar{\delta}_{\varepsilon,\lambda} < 1$ such that if $\delta \geq \bar{\delta}_{\varepsilon,\lambda}$, in any equilibrium $c_{R,1} \geq \bar{c}_R - \varepsilon$, while for both players (and both types), payoffs are within $\varepsilon$ of $\bar{m}$. If $\lambda = 0$ then $c_{R,1} = 0$, responder payoffs are $0$ and proposer payoffs are $100e_P$ (for both types).*

The proposition's proof is standard and is relegated to the Online Appendix. It involves first showing that a proposer must offer $\bar{c}_R$ within a finite number of rounds (such an offer must be accepted). This holds because the proposer can always guarantee a positive continuation payoff by offering $\bar{c}_R$, and so if she makes less generous offers, she must expect normal types to sometimes accept, but in which case the updated probability of facing a fair responder must increase and reach one. This induced finite horizon allows us characterize a (generically) unique equilibrium by backward induction. The equilibrium features offers which slowly increase up to $\bar{c}_R$ (making normal types indifferent to waiting). The final step is to show that there is an upper bound, $T$, on the number of rounds until the proposer offers $\bar{c}_R$, that is uniform for all $\delta \in (0, 1)$. Given the option of waiting to accept, responder payoffs are at least $\delta^{T-1}\bar{m}$ (which is within $\epsilon$ of $\bar{m}$ for large $\delta$).

The cutoff $\delta$ needed for the Coasean prediction to hold depends on the precise details of the model. Nonetheless, the general Coasean prediction of approximately immediate agreement, and approximately equal payoffs, for sufficiently high $\delta$ is seemingly robust to more general forms of fairness preferences (e.g. see Lopomo & Ok (2001)). In particular, this is the case when there is a continuum of types who differ continuously in their concern for fairness with only a positive density refusing any offer less than $\bar{c}_R$. The reasoning is identical: a proposer can guarantee a positive continuation payoff by offering $\bar{c}_R$ and so she must make this offer in a (uniformly) bounded number of rounds.

**Hypothesis 1.** *In the infinite horizon game, initial offers, initial MA offers and final agreed offers will be close to $\bar{c}_R$, payoffs of all agents will be close to $\bar{m}$, and any efficiency loss through bargaining will be small.*

Exactly how close to the limit predictions the data need to be to satisfy this hypothesis is a matter of judgement. To provide a cleaner baseline with which compare our results, we also consider the ultimatum game. In the ultimatum game, there is no guarantee of approximately equal initial offers, equal payoffs or an efficient outcome. Normal responders will accept any positive offer (a MA offer of 0), and so if $\lambda > \bar{\lambda} = \frac{e_P}{e_R + e_P}$, the proposer will offer $c_{R,1} = \bar{c}_R$, which all responders accept (recall that $\lambda > \bar{\lambda}$ in the infinite horizon we also have $c_{R,1} = \bar{c}_R$). For smaller $\lambda$, however, she will offer $c_{R,1} = 0$, which only normal types accept. To see this, notice that $\lambda > \bar{\lambda}$ if and only if $\bar{m} > (1 - \lambda)e_P 100$.

This implies that initial offers, responder payoffs and efficiency[8] should be weakly higher in

[8] One conclusion from the proof of Proposition 1 in the appendix is that when $\lambda > \bar{\lambda}$, there is agreement in round 1 in the infinite horizon game too (not just almost immediate agreement).

the infinite horizon while proposer payoffs should be weakly lower, although the inequality need not be strict if fair types are very likely ($\lambda > \bar{\lambda}$). Initial MA offers should always be strictly higher, however, due to demand withholding. Fair types always set MA offer equal to $\bar{c}_R$. If normal responders chose an MA offer of 0 in the infinite horizon (as well as the ultimatum game) then proposers would offer 0 in round 1 and $\bar{c}_R$ in round 2 for large $\delta$, making the responders strategy suboptimal.

As previously noted, the possibility that offers are close to an equal monetary split even in the ultimatum game is not just a theoretical concern, this is typical in the literature when each agent faces the same chip to money exchange rate. Notice, however, that the cutoff $\bar{\lambda}$ is increasing in the ratio $\frac{e_P}{e_R}$. A higher ratio requires the proposer to sacrifice more chips to buy off fair types, while she can still demand all the chips when facing a normal type (she gets relatively more when being greedy).[9] This is in line with the finding of Kagel, Kim and Moser (1995) that offers are far from an equal monetary split and there is significant disagreement when $e_P = 3e_R$. The Coasean prediction of Hypothesis 1 is not affected by these exchange rates, however, and so we expect to see greater treatment differences between the infinite horizon and ultimatum bargaining when $\frac{e_P}{e_R}$ is larger.

**Hypothesis 2.** *Initial offers, responder payoffs and efficiency will be weakly higher, proposer payoffs will be weakly lower and initial MA offers will be strictly higher in the infinite horizon game than in the ultimatum game. Strict differences in these variables are more likely when $\frac{e_P}{e_R}$ is larger.*

Finally, we reiterate that our model of preferences shouldn't be taken literally. The literature is clear that in the ultimatum game (regardless of exchange rates) proposers do not offer $c_{R,1} = 0$, nor do (any) responders accept such offers. Our qualitative predictions, however, are robust to more complex models.

# 3 Experimental Design

We ran 12 sessions with a total of 236 subjects from the undergraduate student population at the University of Virginia. They were recruited through the VeconLab and Darden BRAD lab pools of students who had signed up to participate in experiments. No subject participated in more than 1 session. The two treatment variables were the horizon of the game, infinite or one-shot, and the exchange rates for chips, equal or 3 times as valuable to proposers. Using the standard $2 \times 2$, design, there were 3 sessions of each of the 4

[9]This seems to be a robust prediction. Suppose the normal type was in fact modeled as in Fehr and Schmidt (1999) as a type with $u_i^N(x_i, x_j) = x_i - \alpha \max\{x_j - x_i, 0\}$ then she will accept any offer greater than $\underline{c}_R = \frac{100 e_P \alpha}{e_R(1+\alpha)+e_P\alpha}$ in the ultimatum game. A proposer should therefore offer $\underline{c}_R$ if $\lambda \leq \underline{\lambda} = \frac{\bar{c}_R - \underline{c}_R}{100 - \underline{c}_R}$ (and $\bar{c}_R$ otherwise) where $\underline{\lambda}$ is again increasing in $\frac{e_P}{e_R}$.

treatments described in the following table.[10]

Treatments

| Name | Horizon | Exchange Rates | Subjects |
|---|---|---|---|
| 1:1 DYN | $\infty$ | $e_P = e_R = 1$ | 60 |
| 3:1 DYN | $\infty$ | $e_P = 3$ and $e_R = 1$ | 60 |
| 1:1 UG | 1 | $e_P = e_R = 1$ | 60 |
| 3:1 UG | 1 | $e_P = 3$ and $e_R = 1$ | 56 |

In each match, subjects were matched in pairs and engaged in the infinite horizon or ultimatum bargaining game (depending on which treatment was run in their session).[11] Each round was implemented with simultaneous moves in which the proposer made an offer and the responder an MA offer, and the offer was accepted if and only if it was larger than the MA offer.[12] In the infinite horizon treatments, discounting was implemented by multiplying each side's chips by $0.95^{t-1}$ when agreement was reached in round $t$.

Subjects received feedback about their own outcomes at the end of each round. In particular, a responder saw the proposer's offer, while a proposer only saw if their offer was accepted or rejected (but not the MA offer of the responder).

Each session consisted of 10 matches where the matching procedure was the turnpike design.[13] That is, the subjects were given a fixed role, proposer or responder, and then matched with every participant of the other role exactly once and in a way that if say proposer A matches to responder B who later matches to proposer C who then matches to responder D, proposer A will match with responder D before proposer C does.

All terms were neutrally framed. The experiment was programmed and run in z-tree (Fischbacher 2007). Subjects were paid for 1 randomly selected match and average earnings were \$25.96.

# 4 Results

Our results are presented in three subsections. The first tests our two hypotheses, the second examines behavior in infinite horizon bargaining in greater detail, and third describes evidence for subject learning. Except for that third subsection we evaluate the

[10]We actually used exchange rates of either $e_P = e_R = \$0.25$ or $e_P = 3e_R = \$0.75$, but to simplify the exposition of the results we rescale these to $e_P = e_R = 1$ or $e_P = 3e_R = 3$.

[11]Potentially these games could have gone on forever. Fortunately, none did.

[12]Offers and MA offers could be made to the nearest one-hundredth of a chip.

[13] In one session of Treatment 3:1 UG, recruitment error led to only 16 participants. In this session, only 8 matches were possible.

last five matches to allow subjects to first learn the game environment.[14]

We use non-parametric tests to calculate significance. Unless explicitly stated otherwise these are Wilcoxon signed ranks tests for single sample tests (comparing data to theoretical benchmarks), or Wilcoxon-Mann-Whitney tests for two sample tests (comparing data from two treatments). We test hypotheses match by match with each subject contributing a single data point, and report the number of matches (out of 5) for which differences are significant at the 5% level.[15] It would not be correct to interpret a significant difference in any one match as clear evidence of a (non-random) treatment effect because multiple hypotheses are being tested, however, we typically find that either all 5 matches deliver significant differences or none do so. The turnpike matching design ensures that subjects should independently maximize utility for each match, without considering the effect of their behavior on future matches. However, to the extent that subjects learn about population behavior from previous matches independence might still be violated. As a robustness check, therefore, we also report Wilcoxon-Mann-Whitney tests for session-average data (6 observations).[16,17]

## 4.1 Hypothesis tests

Figures 1-6 display means, medians, and the interquartile range for round 1 offers, round 1 (MA) minimum acceptable offers, accepted offers, proposer payoffs, responder payoffs, and efficiency by match and treatment respectively. We illustrate variability with the interquartile range as much of the data is heavily skewed.

We first compare outcomes in the dynamic treatments to the theoretical benchmark of an immediately agreed equal monetary split (Hypothesis 1) and then compare these to ultimatum game outcomes (Hypothesis 2).

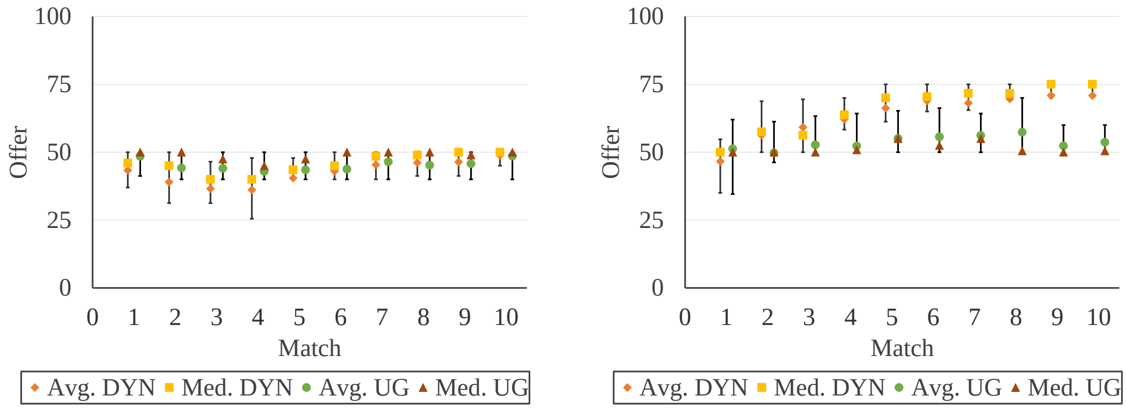
Figures 1 and 2 display round 1 offers and MA offers. Not only are means and medians close to $\bar{c_R}$ (an equal monetary split) in both infinite horizon treatments, but the interquartile range exhibit very little variation by the last few matches. The average offer across matches 6-10 was 45.96 in 1:1 DYN and 69.69 in 3:1 DYN, while the average MA

[14] Guth, Ockenfels and Ritzberger (1995) speculate that the lack of opportunity to learn about their complicated game environment explains why theory performed so poorly in their experiment. In a different infinite horizon setting, the repeated prisoner's dilemma, a survey by Dal Bó and Fréchette 2017 find that theory performs much better once subjects have had a chance to learn.

[15] Notice that for payoffs, efficiency and final accepted offers, independence would be badly violated by averaging across rounds 6-10. For example, Proposer A would affect the averaged payoffs of Responder B and C whom she is paired with in matches 6 and 7 respectively.

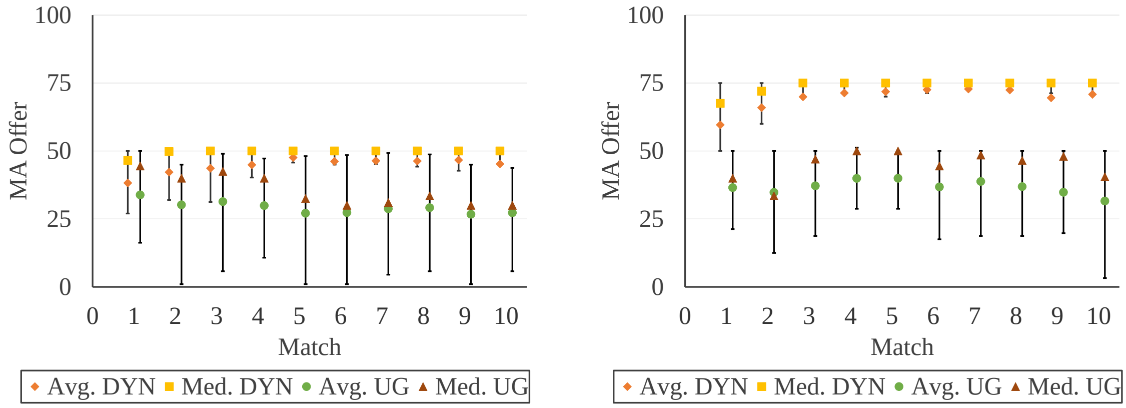
[16] In fact, for matches 9 and 10 when comparing 3:1 exchange rates we only have 5 observations due to the recruitment issue detailed in footnote 13. In this case we use a 10% significance threshold, the highest one feasible (when all sessions of one treatment are above all those of the other).

[17] Using session average data for Wilcoxon Signed Ranks tests can never give a significant difference with only 3 data points.




(a) 1:1 treatments (b) 3:1 treatments

**Figure 1.** Offers by Match

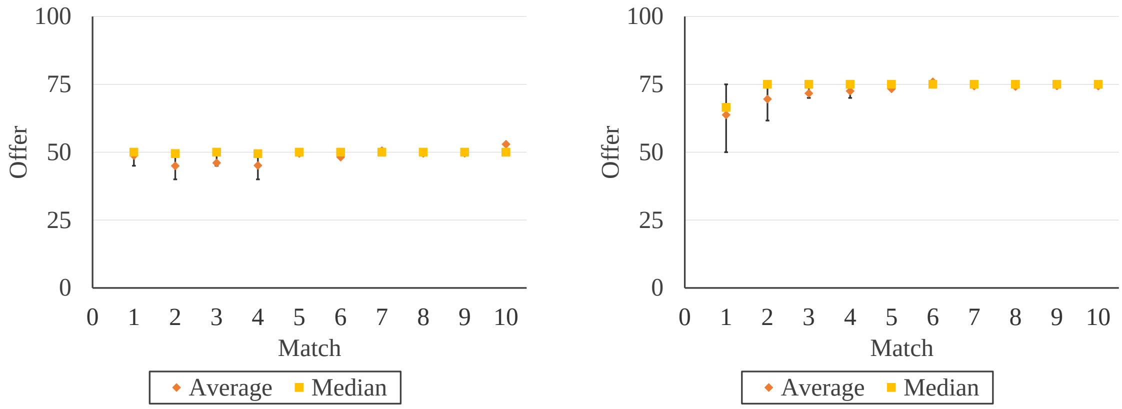



(a) 1:1 treatments (b) 3:1 treatments

**Figure 2.** MA Offers by Match

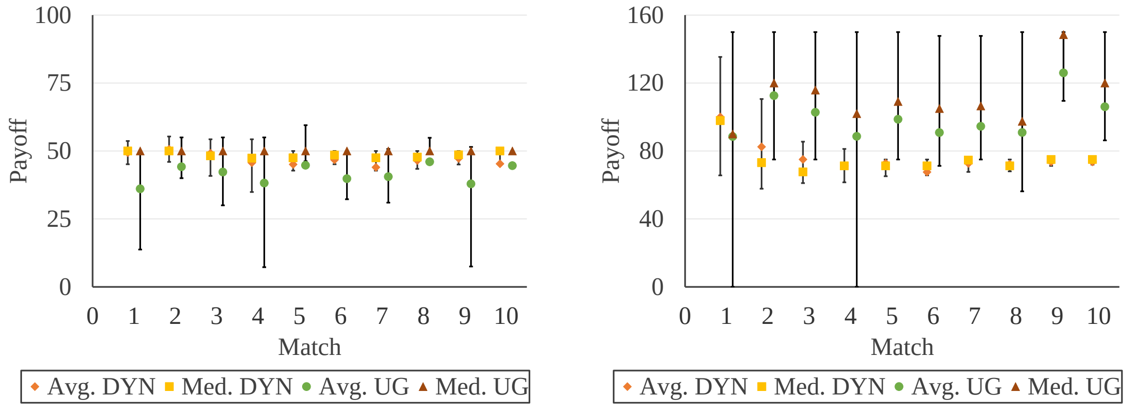
offer was respectively 46.17 and 71.64. Each of these averages is less than 9% from $\bar{c}_R$. Statistically, offers are significantly less than $\bar{c}_R$ for every one of the last 5 matches for each treatment while the MA offers are significantly less than $\bar{c}_R$ in 3 of the last 5 matches for each treatment. When offers and MA offers differ from $\bar{c}_R$, they almost always are less than it. Close to, but possibly below, $\bar{c}_R$ is precisely what the Coase Conjecture predicts and so this is consistent with the first statement of Hypothesis 1. The fact that initial offers were on average slightly lower than MA offers suggests that, at least occasionally, there was delay in reaching agreement.

Figure 3 illustrates perhaps our most striking result, that final accepted offers are almost exactly $\bar{c}_R$. They are not significantly different from it in any of the last 5 matches of either treatment. The average across the last 5 matches is 50.16 for 1:1 DYN and 74.67 for 3:1 DYN, less than 0.5% from $\bar{c}_R$. Moreover, there is very little variation around this average, highlighted by the degenerate interquartile ranges. Indeed, in 1:1 DYN, 67% of accepted offers are exactly 50, and 92% are within 5 of 50. In 3:1 DYN, 76% of accepted




(a) 1:1 treatment (b) 3:1 treatment

**Figure 3.** Accepted Offers by Match

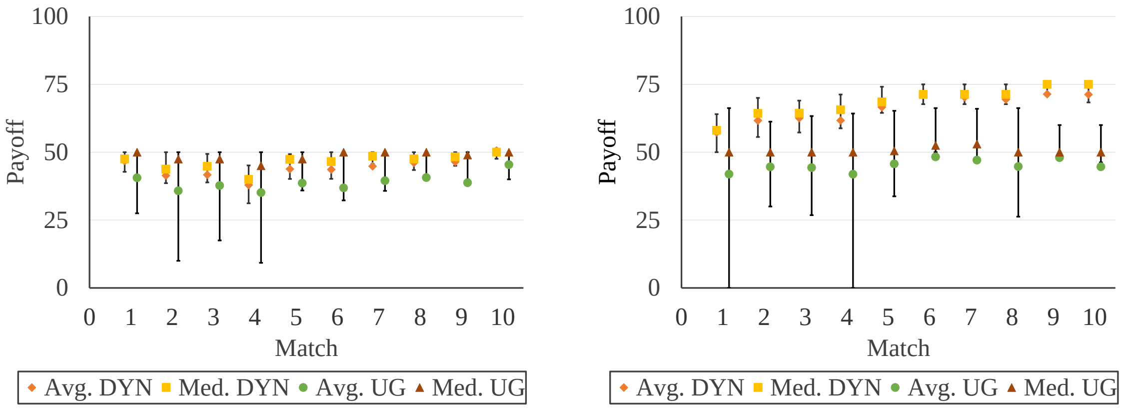



(a) 1:1 treatments (b) 3:1 treatments

**Figure 4.** Proposer Payoffs by Match

offers are exactly 75 and 97% are within 5 of 75. This is again in line with Hypothesis 1.

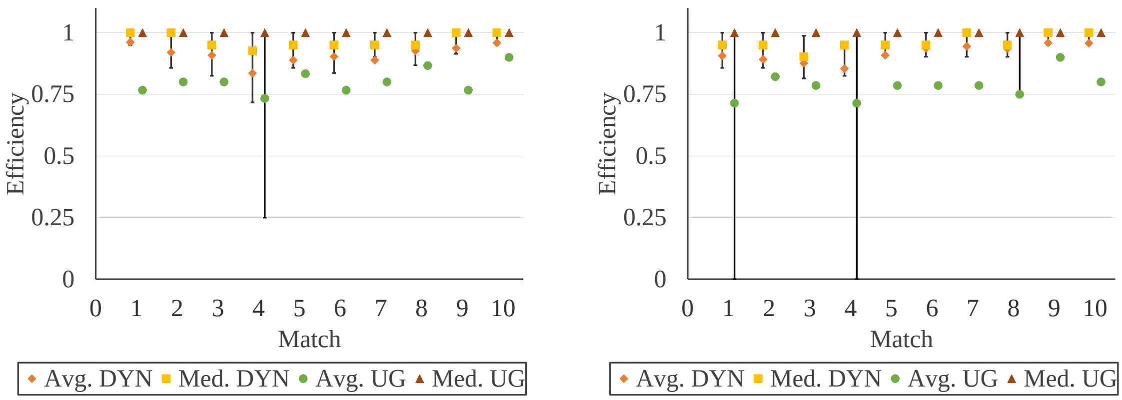
Figures 4 and 5 display payoffs for proposers and responders respectively. Hypothesis 1 predicted payoffs close to $\bar{m}$ for both players. The average proposer payoff in the last 5 matches is 45.96 and 71.95 for 1:1 DYN and 3:1 DYN respectively, while the average responder payoff is 46.32 and 70.71. All these averages are less than 9% from $\bar{m}$, and moreover, the interquartile ranges indicate little variation. Statistically, payoffs are significantly less than $\bar{m}$ for all of the last 5 matches for proposers in Treatment 1:1 DYN and for responders in both treatments, while proposer payoffs in 3:1 DYN are significantly different in only 2 matches. The difference between proposer and responder payoffs can also be directly compared using a matched-pairs Wilcoxon signed ranks test. This difference is significant for only 1 of the last 5 matches for each treatment. While the small differences in magnitude of payoffs from $\bar{m}$ is again in line with the Coasean prediction, it should be noted that theory would predict proposer payoffs slightly larger than $\bar{m}$ if not exactly equal (because proposers should always be able to offer $\bar{c}_R$ and have




(a) 1:1 treatments

(b) 3:1 treatments

**Figure 5.** Responder Payoffs by Match




(a) 1:1 treatments

(b) 3:1 treatments

**Figure 6.** Efficiency by Match

it immediately accepted).

Figure 6 reports efficiency, which is recorded as $\delta^{t-1}$ when there is agreement in round $t$ (and 0 if there is no agreement in the ultimatum game). Across the last 5 matches, average efficiency is 92% and 95% for 1:1 DYN and 3:1 DYN respectively. This is again in line with Hypothesis 1's prediction of little inefficiency. Indeed, a majority of pairs reached agreement in round 1 and at least 70% did so by round 2 for both treatments. Statistically, efficiency is nonetheless significantly less than 100% for all of the last 5 matches in both treatments.[18]

In summary, we view the small differences in magnitude between observed outcomes and an immediate agreement on an equal monetary split as strong evidence in support of Hypothesis 1 and by extension the Coase Conjecture.

[18]Statistical significance here is not surprising as efficiency is bounded above by 100% so all deviations are negative.

We next compare infinite horizon game outcomes to those in the ultimatum game in order to assess Hypothesis 2. We first compare the treatments in which proposers and responders have the same exchange rate (1:1 DYN vs 1:1 UG). As seen in Figure 1, round 1 offers are essentially identical for both game horizons and there are no statistical differences for any of the last five matches (which is robust to session-average comparisons). Of course, theory allows for no strict treatment difference in this case, because if fair types are sufficiently likely, an equal split offer may be optimal even in the ultimatum game (this is why Hypothesis 2 allows for weak differences). In fact, we can verify that this was the case. Given the empirical distribution of MA offers in the last 5 matches, the offer which maximized proposer payoffs in the ultimatum game was exactly $\bar{c}_R = 50$.

Even with no difference in offers, Hypothesis 2 predicted that round 1 MA offers should be strictly higher in the infinite horizon than in the ultimatum game (due to demand withholding). This is indeed what we find, as illustrated in Figure 2. The difference is significant for all of the last 5 matches (which is robust to session-average comparisons). Moreover, the magnitude of the difference is large, across the last 5 matches, the average MA offer is 66% larger in the infinite horizon than in the ultimatum game (46.17 vs 27.83). This suggests that most subjects who didn't care much about fairness, nonetheless understood the logic of the Coase Conjecture (i.e. they could profitably hold out for close to an equal split in the infinite horizon).[19]

Given that initial offers were close to an equal split for both game horizons it is unsurprising that payoffs are also close to each other for both proposers and responders, as seen in Figures 4 and 5. As a result, there is just one case of a significant treatment difference for payoffs, match 8 for proposers where payoffs are significantly lower in the infinite horizon game, and this difference is not robust to session-average comparisons.

Figure 6 suggests that mean efficiency is slightly higher in the infinite horizon than in the ultimatum game. Indeed, averaging over the last 5 matches efficiency is 82% in the ultimatum game compared to 92% in the infinite horizon. Nonetheless, the only significant difference is in match 8. This is perhaps not surprising given that the medians and interquartile ranges are tightly clustered around 100% for both treatments. However, it may also reflect the fact that the ranksum test is not entirely appropriate here, because efficiency in the ultimatum game is binary rather than continuous, whereas in the infinite horizon it can be any multiple of 0.95. On the other hand, our robustness check using session averages finds significant differences in 3 of the 5 matches.

[19] One fact that isn't entirely in line with theory is that 57% of round 1 MA offers were (at least) $\bar{c}_R$ in 1:1 DYN compared to only 20% in 1:1 UG. Theory predicts that only (perfectly) fair types should have such high acceptance thresholds in either treatment. Other subjects should accept slightly less than $\bar{c}_R$ in the infinite horizon when anticipating $\bar{c}_R$ in the following round (e.g. a money maximizing type should accept anything above $\delta\bar{c}_R = 47.5$). This perhaps suggests that subjects used rules of thumb rather than a precise tradeoff between higher future offers and the cost of delay.

According to Hypothesis 2, the best test for the Coase Conjecture (with strict treatment differences) is the comparison between the ultimatum game and the infinite horizon when chips are worth three times as much to proposers (3:1 DYN vs 3:1 UG). This is because there is less incentive for proposers to make equal split offers in the ultimatum game for these exchange rates. Not only do we observe significant differences where theory predicts, but these differences are large in magnitude

Averaging across the last 5 matches, round 1 offers are 26% higher in the 3:1 DYN than in 3:1 UG (69.69 vs 55.29), while MA offers are as much as 98% higher (71.64 vs 36.10). The treatment differences are significant for every one of the last 5 matches for both variables (and robust to session-average comparisons). Unlike for 1:1 exchange rates, the empirical distribution of MA offers in the ultimatum game implies that the expected payoff maximizing offer is not $\bar{c}_R = 75$ but rather 50. This incentive to make offers below $\bar{c}_R$ in the ultimatum game is exactly what is needed to deliver strict treatment differences (for offers, payoffs and efficiency).

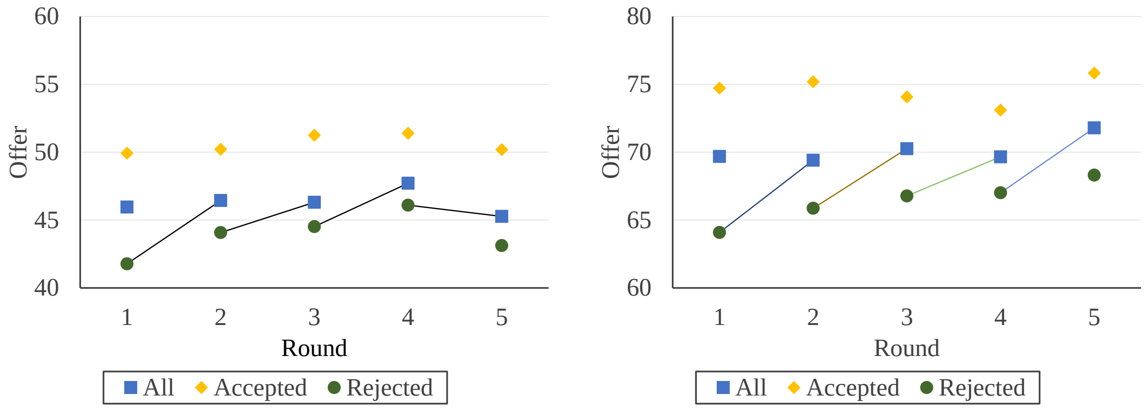
As a consequence, proposer payoffs are 28% lower (71.95 vs 99.81), while responder payoffs are 52% higher payoffs (70.71 vs 46.57), averaging across the last 5 matches. For both roles, these payoff differences are significant for every one of the last 5 matches (the comparisons are all robust to session-averages for responders, but only robust for 3 out of 5 matches for proposers).

Finally, average efficiency in the infinite horizon, at 95%, is again higher than in the ultimatum game, at 80%, averaging across the last 5 matches. This difference is not significant for any match, however, recall that the ranksum test is perhaps not entirely appropriate here. For session averages, the difference is significant for 4 of the 5 matches.

In summary, subject behavior is very much consistent with Hypothesis 2. While large strict treatment differences for some variables are only observed when the exchange rate ratio is 3:1 rather than 1:1, this is exactly what theory led us to expect. The finding that infinite horizon outcomes are both very close to an immediate agreement on an equal monetary split (Hypothesis 1) and closer to that outcome than the ultimatum game (Hypothesis 2), appears to offer strong support for the Coase Conjecture.

## 4.2 Infinite Horizon Bargaining

In this subsection, we explore subject behavior in the infinite horizon game in greater detail. As previously mentioned, a majority of matches reached agreement in round 1 for both infinite horizon treatments, however, here we focus on those that did not. The (unique) equilibrium of the infinite horizon game which provides the basis for the Coase conjecture predicts that if proposers do not offer an equal monetary split $\bar{c}_R$ immediately,




(a) Treatment 1:1 DYN

(b) Treatment 3:1 DYN

**Figure 7.** Average Offers by Round

then after a less generous offer is rejected it should be increased (although not above $\bar{c}_R$). In the simple two type model developed in the theory section this rate of increase of offers is $\frac{1-\delta}{\delta} \approx 5\%$ in order to make money maximizing responders indifferent to waiting. With a richer set of "somewhat fair" types in the model the rate of increase may be smaller, but it should not be larger or else all responders would find it worthwhile to wait for $\bar{c}_R$.

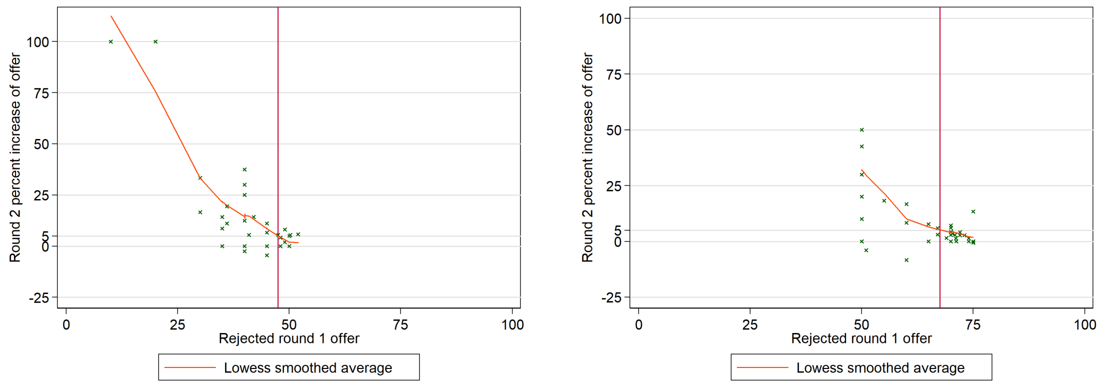
Figure 7 plots the average offers in each round for all offers, accepted offers, and rejected offers for the first 5 rounds.[20] Lines are placed between all offers in round $t$ and rejected offers in round $t-1$ to highlight the change in offer after an offer is rejected.

The figure illustrates several things. First, we confirm the previous result that accepted offers are almost exactly $\bar{c_R}$. This seems to hold regardless of how long the bargaining game lasts. Second, we see that proposers do indeed increase their offers after rejection. This is true for all but the change from round 4 to 5 in Treatment 1:1 DYN (and that case accounts for only $n = 7$ observations).

The average increase of round 2 offers given a rejected round 1 offer is as much as 11% in 1:1 DYN (41.78 to 46.44) and 8% in 3:1 DYN (64.09 to 69.41), although the rate is slightly lower between later rounds. This suggests that it was profitable on average for responders to reject the round 1 offers which they in fact rejected. However, this clearly does not mean that it was profitable to reject all offers, in particular those already close to $\bar{c}_R$. To provide a clearer assessment of which offers were worth rejecting, Figure 8 plots rejected round 1 offers against the percentage increase of offer in round 2 as well as a smoothed lowess line of best fit.[21] The lowess line is monotonically decreasing in a proposer's initial offer for both treatments. In 1:1 DYN the lowess line falls below 5% at 47.5, which is exactly $\delta\bar{c}_R$ (if proposers never offer above $\bar{c}_R$ then money maximizing

[20] 92% of matches last less than or equal to 5 rounds.

[21] The 3:1 DYN figure excludes one round 1 offer of 60 which fell 100% to 0.




(a) Treatment 1:1 DYN

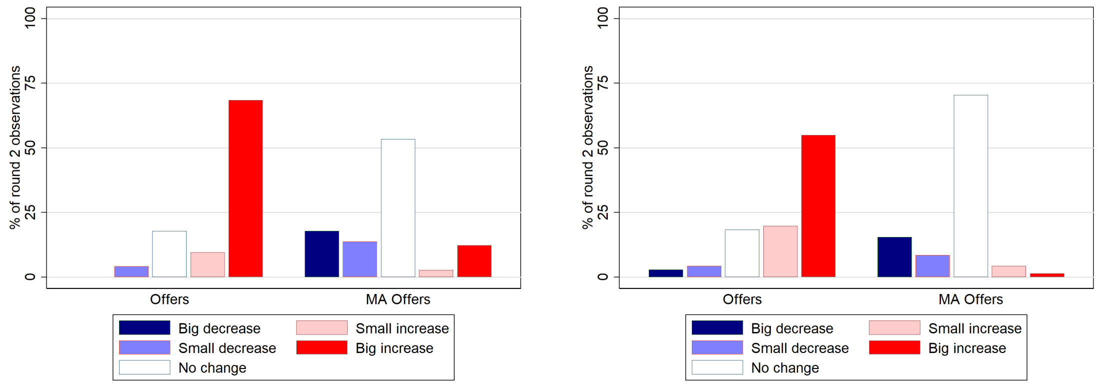
(b) Treatment 3:1 DYN

**Figure 8.** Percentage increase of round 2 offer by rejected round 1 Offer

responders should always accept offers above $\delta \bar{c}_R$). In 3:1 DYN the lowess line falls below 5% at 67.5. Below these cutoffs, proposers' were not able to credibly commit to not substantially increasing offers, making it optimal for even money maximizing responders to reject. Very few round 1 offers of at least $\bar{c}_R$ were rejected (15% in 1:1 DYN and 7% in 3:1 DYN), but the average increase among these was 1.5% in 1:1 DYN and 2.5% in 3:1 DYN (a majority of such offers in each treatment remained constant).

Theory's predictions for MA offers are less clear. In our simple two type model, a money maximizing responder mixes over her acceptance decision. This mixing could potentially be done at the population level, however, so that each responder's MA offers are constant. In fact, this was typically the case. Figure 9 shows the fraction of round 2 offers and MA offers which represent a big (strictly more than 5%) decrease, a small (between 5% and 0%) decrease, no change, a small increase and a big increase compared to round 1. More than 68% (in 1:1 DYN) and 54% (in 3:1 DYN) of rejected round 1 offers increased by more than 5%, while 53% (in 1:1 DYN) and 70% (in 3:1 DYN) of MA offers did not change at all. It seems that while proposers' increased their offer following a rejection (eventually up to $\bar{c}_R$), responders were more content to wait for that to happen. This is broadly consistent with the predicted equilibrium dynamics. To statistically confirm that impression, we compare the change in round 2 offers averaged at the proposer level (4.11 in 1:1 DYN and 4.77 in 3:1 DYN) with the negative change (change$\times(-1)$) in round 2 MA offers averaged at the responder level (0.78 in 1:1 DYN and 2.41 in 3:1 DYN). The null hypothesis of no difference between these changes is rejected at 1% level in a rank sum test for both exchange rates.

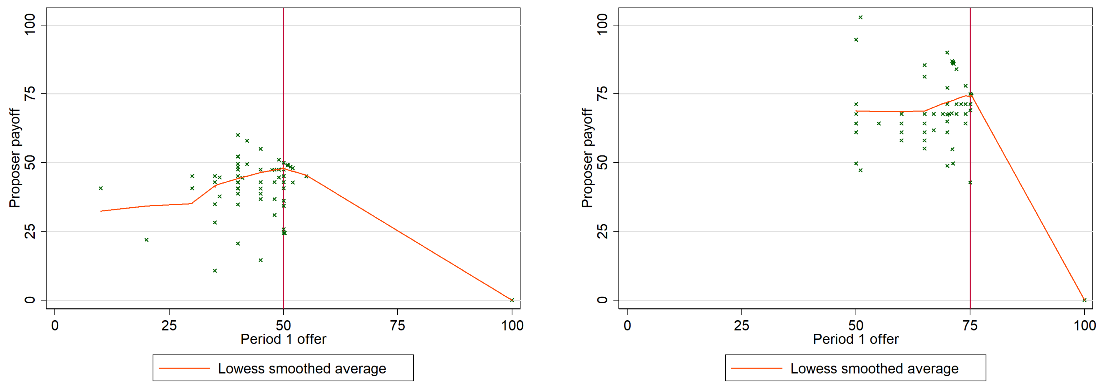
Figure 8 shows that given proposer behavior, responders should have optimally rejected (round 1) offers which were not close to an equal monetary split. Figure 2 shows that on average responders did in fact do this. Given responder behavior, what was a proposers' optimal strategy? We can attempt an answer to this question by treating a proposer's




(a) Treatment 1:1 DYN

(b) Treatment 3:1 DYN

**Figure 9.** Change in Offers and MA Offers in round 2




(a) Treatment 1:1 DYN

(b) Treatment 3:1 DYN

**Figure 10.** Initial Offers and Proposer Payoffs

round 1 offer as a proxy for her (infinite horizon) strategy. Figure 10 plots a proposer's payoff in a match against her round 1 offer as well as a smoothed lowess line. In both treatments the lowess line is initially monotonically increasing and then monotonically decreasing, reaching a maximum at exactly $\bar{c}_R$.[22] Figure 1 established that proposers' average round 1 offers were indeed very close to $\bar{c}_R$. Such approximate optimality from both proposers and responders suggests that behavior was in fact not very far from (a Coasean) equilibrium.

## 4.3 Learning

Finally, we investigate how behavior changed over the course of the experiment. Figures 1-6 suggest that learning was more important for 3:1 DYN than for other treatments.

[22]Notice that if responders always accept offers above $\bar{c}_R$, then this line mechanically decreases for offers above $\bar{c}_R$.

Much of this learning occurs in the first few matches.

Comparing the first and last match in 3:1 DYN, proposers' round 1 offers increased by 52% (46.62 to 70.85), while responders' MA offers increased by 19% (59.62 to 70.83). A matched-pairs Wilcoxon signed ranks test finds that both changes are significantly different from zero at the 1% level. Proposer payoffs decreased 26% (99.98 to 73.60) from the first to the last match, while responder payoffs increased by 24% (57.28 to 71.25), with both changes significant at the 1% level. Average efficiency increased from 91% to 96%, which is significant at the 10% level (for the efficiency change associated with proposers).[23]

By contrast in the other treatments there was little evidence for learning. In 1:1 DYN, the 13% increase of round 1 offers between the first and last match (43.33 to 48.88) is significant at the 10% level, but all other changes are insignificant. In 1:1 UG, the 24% increase in proposer payoffs (36.07 to 44.60) is significant at the 10% level, but all other changes are insignificant. Finally in 3:1 UG the 11% increase of offers (51.25 to 56.80) is significant at the 10% level, but all other changes are insignificant.

Learning in 3:1 DYN, seems to occur in two stages. First, responder behavior converges to the Coasean prediction, and only then does proposer behavior adapt. Round 1 MA Offers (and final accepted offers) are already close to $\bar{c}_R = 75$ by round 3, with an essentially degenerate interquartile range. Round 1 offers, meanwhile take longer to converge (the interquartile range overlaps with $\bar{c}_R = 75$ for the first time only in match 5).

It is perhaps not surprising that learning was most important in one of the infinite horizon treatments, a more complicated bargaining environment. Within the infinite horizon treatments, the fact that significant learning occurs only for the 3:1 exchange rate, is perhaps because of proposers' stronger temptation to make unequal payoff offers in this case (as in 3:1 UG). Only with experience did they come to understand that such low ball offers were not worthwhile in the infinite horizon (due to the Coase conjecture), and merely led to delay.

# 5 Conclusion

It is often difficult to obtain enough experimental control to convincingly test important theoretical results in lab. Subjects' heterogenous, unobserved preferences for fairness represent an important potential confound for game theoretic predictions. In this paper, rather than ignoring such naturally occurring preferences, we embraced them, and sought

[23] For payoffs and efficiency, the assumption of independence may be quite suspect here, because Proposer A and B's change of payoff may both be affected by the behavior of Responder C (with whom they are paired with in matches 1 and 10 respectively).

to directly use this source of private information to test the Coase conjecture. For an infinite horizon bargaining game in which proposers make all offers, the theory predicts that patient players will agree almost immediately on close to an equal monetary split. This prediction seems to perform very well. Initial offers, initial minimum acceptable offers, proposer payoffs, responder payoffs and efficiency are all less than 9% away from the immediate equal split benchmark (theory's limit prediction as players become infinitely patient).

One problem with our theoretical prediction is that it lacked precision (how close is close?). We, therefore, also compared the infinite horizon game outcomes to those in an ultimatum game. Our finding that it is impossible to statistically distinguish outcomes (apart from minimum acceptable offers) for the two different game horizons when chips are equally valuable to both players, is consistent with the theory (because immediately offering an equal split may always be profitable when fair types are very prevalent). According to theory, however, strict treatment differences are more likely when chips are worth more to proposers than responders, which is exactly what we find. Not only are initial offers, initial minimum acceptable offers and proposer payoffs significantly lower in the infinite horizon, and responder payoffs significantly higher, but the magnitude of the treatment differences is large. We view our findings as strong evidence that subjects comprehend the basic message of the Coase conjecture, that with one sided asymmetric information the uniformed party must "give in" to the informed party almost immediately (as if she was bargaining with the informed party's toughest type).

It might be argued that while the Coase conjecture performs well in our narrowly defined setting, this is at the expense of sacrificing its reach. As a result, we may not be able to make a clear prediction about what will happen in Coase's original setting of a seller facing a buyer who is privately informed about her value (because the real world is messy and has additional sources of private information, including those about fairness preferences). Two responses are in order. First, bargaining theorists have long acknowledged that clear predictions are rare when there are multiple sources of private information. Some progress has been made, however, for instance in two-sided reputational models (Abreu & Gul (2000), and those models are still built on Coasean foundations (with one-sided private information). In this light, our experiment is important because it shows that those foundations appear secure. Second, our result highlights how fairness preferences in particular, may play an outsized role in real world bargaining situations. Even in settings where people are unlikely care much about fairness (e.g. CEOs of multinational corporations), the mere possibility that they do combined with the Coase conjecture may be enough to profoundly affect outcomes.

# References


[1] Abreu, Dilip and Gul, Faruk. Bargaining and Reputation. *Econometrica* 2000, 68(1), pp. 85-117.

[2] Cason, Timothy and Sharma, Tridib. Durable Goods Monopoly, Coasean Dynamics, and Uncertainty: Theory and Experiments. *Journal of Political Economy* 2001, 109(6) pp. 1311-1354.

[3] Cason, Timothy and Reynolds, Stanley. Bounded Rationality in Laboratory Bargaining with Asymmetric Information. *Economic Theory* 2005, 25, pp. 553-574.

[4] Coase, Ronald. Durability and Monopoly. *The Journal of Law and Economics* 1972, 15(1), pp. 143-149.

[5] Cullen, Zoe and Pakzad-Hurson Bobak. Equilibrium Effects of Pay Transparency. *Working Paper* 2017.

[6] Dal Bó, Pedro and Fréchette, Guillaume. On the Determinants of Cooperation in Infinitely Repeated Games: A Survey. *Journal of Economic Literature* 2017, forthcoming.

[7] Embrey, Matthew, Fréchette, Guillaume, and Lehrer, Stephen. Bargaining and Reputation: Experimental Evidence on Bargaining in the Presence of Irrational Types *The Review of Economic Studies* 2015, 82(2), pp. 608-631.

[8] Fanning, Jack. Bargaining, Fairness, and Commitment. *Working Paper* 2014.

[9] Fehr, Ernst and Schmidt, Klaus. A Theory of Fairness, Competition and Cooperation. *Quarterly Journal of Economics* 1999, 114, pp. 817-868.

[10] Fischbacher, Urs. z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics* 2007, 10(2), pp. 171-178.

[11] Gul, Faruk, Sonnenschein, Hugo, and Wilson, Robert. Foundations of Dynamic Monopoly and the Coase Conjecture. *Journal of Economic Theory* 1986, 39(1), pp. 155-190.

[12] Guth, Werner, and Kocher, Martin. More than Thirty Years of Ultimatum Bargaining Experiments: Motives, Variations, and a Survey of the Recent Literature. *Journal of Economic Behavior and Organization* 2014, 108, pp. 396-409.

[13] Guth, Werner, Kroger, Sabine, and Norman, Hans-Theo. Durable-Goods Monopoly with Privately Known Impatience: A Theoretical and Experimental Study *Economic Inquiry* 2004, 42(3), pp. 413-424.

[14] Guth, Werner, Ockenfels, Peter, and Ritzberger, Klaus. On Durable Goods Monopolies: an Experimental Study of Intrapersonal Price Competition and Price Discrimination over Time. *Journal of Economic Psychology* 1995, 16(2), pp. 247-274.

[15] Kagel, John, Kim, Chung, and Moser, Donald. Fairness in Ultimatum Games with Asymmetric Information and Asymmetric Payoffs. *Games and Economic Behavior* 1996, 13(1), pp. 100-110.

[16] Lopomo, Giuseppe and Ok, Efe. Bargaining, Interdependence, and the Rationality of Fair Division. *RAND Journal of Economics* 2001, 32(2), pp. 263-283.

[17] Oosterbeek, Hessel, Sloof, Randolph, and Van de Kuilen, Gijs. Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis. *Experimental Economics* 2004, 7, pp. 171-188.

[18] Reynolds, Stanley. Durable-Goods Monopoly: Laboratory Market and Bargaining Experiments. RAND Journal of Economics 2000, 31(2), pp. 375-394.

[19] Rapoport, Amnon, Erev, Ido and Zwick, Rami. An Experimental Study of Buyer-Seller Negotiation with One-sided Incomplete Information and Time Discounting. *Management Science* 1995, 41(3), pp. 377-394.

[20] Roth, Alvin and Malouf, Michael. Game-Theoretic Models and the Role of Information in Bargaining. *Psychological Review* 1979, 86, pp. 574-594.

[21] Slembeck, Tilman. Reputations and Fairness in Bargaining - Experimental Evidence from a Repeated Ultimatum Game With Fixed Opponents. *Discussion Paper* 1999.

# 6 Online Appendix (not for publication)

**Proof of Proposition 1**

I first define a series of equilibrium offers $q(n) = \delta^n \bar{c}_R$, which will apply when there are $n$ rounds of bargaining remaining. Notice that this makes a normal responder indifferent between accepting immediately or waiting for an additional round.

The equilibrium probability that the responder is fair with $n$ rounds of bargaining remaining is $\eta(n)$, and the proposer's value function divided by her exchange rate $e_P$ is $W(n)$. We shall refer to this as simple her value function (rescaling by $e_P$ in this way does not affect the analysis). A simple application of Bayes' rule implies that if beliefs are given by $\eta(n)$ today and $\eta(n-1)$ tomorrow and only normal types accept, then a fraction $\frac{\eta(n-1)-\eta(n)}{\eta(n-1)}$ of responders do in fact accept today. The variables $\eta(n)$ and $W(n)$ are defined recursively from $\eta(n-1)$ and $W(n-1)$ starting with $\eta(0) = 1$ and $W(0) = 100 - \bar{c}_R$ using the equality:[24]

$$\begin{aligned} &\frac{\eta(n-1)-\eta(n)}{\eta(n-1)}(100 - q(n)) + \frac{\eta(n)}{\eta(n-1)}\delta W(n-1) \\ =&\frac{\eta(n-1)-\eta(n)}{\eta(n-1)}(100 - q(n-1)) + \frac{\eta(n)}{\eta(n-1)}W(n-1) \end{aligned} \tag{1}$$

with $W(n)$ defined as the value at which equality is obtained. This means that $\eta(n)$ represents the belief at which the proposer is just indifferent between offering $q(n)$ today followed by $q(n-1)$ tomorrow, or immediately offering $q(n-1)$. Although the lower offer may be desirable other things equal, it involves some delay. Meanwhile, $W(n)$ simply represents good accounting for the proposer's value function. Using the equality $q(n-1) - q(n) = (1-\delta)q(n-1)$ and rearranging the above equation gives:

$$\frac{\eta(n-1)-\eta(n)}{\eta(n-1)} = \frac{W(n-1)}{q(n-1)+W(n-1)} \tag{2}$$

Plugging this in to evaluate $W(n)$ gives:

$$W(n) = \frac{100W(n-1)}{q(n-1)+W(n-1)} \tag{3}$$

[24]The second line of this equality is equal to $W(0)$ if $n = 1$ and $\frac{\eta(n-2)-\eta(n)}{\eta(n-2)}(100 - q(n-1)) + \frac{\eta(n)}{\eta(n-2)}\delta W(n-2)$ if $n = 2$.

We are now ready to characterize the equilibrium.

**Lemma 1.** *There is a generically unique equilibrium path. If $\lambda \in (\eta(N), \eta(N-1))$ then bargaining must finish by round $N$. On the equilibrium path offers are:*[25]

$$c_{R,t} = q(N-t) \textit{ for } t \in \{1, ..., N\}$$

*Beliefs that responder is a fair type are:*

$$\mu_t = \eta(N-t) \textit{ for } \textit{ for } t \in \{2, ..., N\}$$

Notice that equation 2 implies $\eta(1) = \frac{\bar{c}_R}{100} = \frac{e_P}{e_R+e_P}$, so that if $\lambda > \bar{\lambda} = \frac{e_P}{e_R+e_P}$ then we must have $c_{R,1} = \bar{c}_R$, as claimed in the main text.

Given the characterization of the lemma the proof of the Proposition is almost immediate. Notice that $W(n-1) \geq 100-\bar{c}_R$ and $q(n-1) \leq \bar{c}_R$ imply that the probability of acceptance in each round is bounded away from zero:

$$\frac{\eta(n-1)-\eta(n)}{\eta(n-1)} = \frac{W(n-1)}{q(n-1)+W(n-1)} \geq \frac{100-\bar{c}_R}{100} > 0$$

Hence, given any $\lambda > 0$, Bayesian updating then implies for $t \geq 1$ that $\mu_{t+1} \geq \mu_t \frac{100}{\bar{c}_R} \geq \lambda \left(\frac{100}{\bar{c}_R}\right)^t \geq 1$. Clearly, therefore, bargaining must end before round $T = \left\lceil \frac{ln\left(\frac{1}{\lambda}\right)}{ln\left(\frac{100}{\bar{c}_R}\right)} \right\rceil + 1$ for any $\delta$, or else $\mu_{T+1} > 1$. Given this, we must have that $c_{R,1} \geq q(T-1) = \delta^{T-1}\bar{c}_R$, and responder payoffs for both types are certainly greater than $\delta^{T-1}\bar{c}_R e_R$ (given their option of waiting to accept the offer $\bar{c}_R$), while proposer payoffs are certainly greater than $\bar{m}$ (given the option to offer $\bar{c}_R$ immediately). Choosing $\bar{\delta}_\lambda < 1$ appropriately, therefore, gives the result (notice in particular that if the responder expected payoff is greater than $\bar{m} - \epsilon$ then the proposer's payoff must be smaller than $\bar{m} + \frac{e_P\epsilon}{e_R}$).

### Proof of Lemma 1

First, we claim that bargaining must end if the proposer ever offers at least $\bar{c}_R$. Let $\overline{q}$ be the supremum of equilibrium offers and initially suppose that $\overline{q} > \bar{c}_R$. In which case, there exists some $\varepsilon \in (0, \overline{q} - \bar{c}_R)$ such that $\overline{q} - \varepsilon > \delta\overline{q}$ so that all responders should accept, but in which case the proposer should never make an offer above $\overline{q} - \varepsilon$, a contradiction. Hence, $\overline{q} \leq \bar{c}_R$, and if the proposer every offers $\bar{c}_R$ in equilibrium, it must certainly be accepted (as $\bar{c}_R > \delta\bar{c}_R$).

A history $h_t$ is sequence of past offers. Let $\mu_{t+j}(h_t)$ and $c_{R,t+j}(h_t)$ describe beliefs and

[25]For $\lambda = \eta(N)$ there are two possible price and belief paths one corresponding to the equilibrium where $\lambda \in (\eta(N), \eta(N-1))$, the other corresponding to the equilibrium where $\lambda \in (\eta(N+1), \eta(N))$.

offers in round $t+j$ consistent with some continuation equilibrium after $h_t$ (such that offers have not yet hit $\bar{c}_R$, ending bargaining). The proposition then follows from the statements below, which are proved by induction on $n$.

1. If $\mu_t(h_t) > \eta(n)$ then bargaining last at most $n$ more rounds including the current one, $c_{R,t+j}(h_t) \geq q(n-1-j)$ for $j \in \{0,...,n-1\}$ and $\mu_{t+j}(h_t) \geq \eta(n-1-j)$ for $j \in \{1,...,n-1\}$.

2. If $\mu_t(h_t) \in (\eta(n), \eta(n-1))$ then the proposer's continuation value is given by:

$$V_{n-1}(\mu_t(h_t)) = \frac{\eta(n-1) - \mu_t}{\eta(n-1)}(100 - q(n-1)) + \frac{\mu_t}{\eta(n-1)}W(n-1)$$

3. If $\mu_t(h_t) \leq \eta(n)$ and $c_{R,t} \in (q(n), q(n-1))$ then $\mu_{t+1}(h_t, c_{R,t}) = \eta(n-1)$.

4. If $\mu_t(h_t) \leq \eta(n)$ and $c_{R,t} < q(n)$, then $\mu_{t+1}(h_t) \leq \eta(n)$.

5. If $\mu_t(h_t) = \eta(n)$ then the proposer's continuation value is $W(n)$.

6. If $\mu_t(h_t) < \eta(n)$ then bargaining lasts at least $n+1$ more rounds including the current one with $c_{R,t+j}(h_t) \leq q(n-j)$ and $\mu_{t+j}(h_t) \leq \eta(n-j)$ for $j \in \{0,...,n\}$.

Consider the above claims for $n = 0$, and let $\eta(-1) = 1$ and $W(-1) = q(-1) = 100 - \bar{c}_R$. 1) and 2) and 3) are then true immediately because their antecedents cannot be activated. 4) is true because $\mu_{t+1}(h_t, c_{R,t}) \leq 1 = \eta(0)$. For 5), recall that any offer of at least $\bar{c}_R$ must be accepted. If $\mu_t(h_t) = 1$ then the best possible continuation value the proposer can obtain is from proposing $\bar{c}_R$ immediately, because the fair responder will not accept less. This gives a continuation value of $W(0) = 100 - \bar{c}_R$. For 6) given that the game has not ended it is immediate that if bargaining lasts at least 1 round, and given that both types will immediately accept an offer of $\bar{c}_R = q(0)$ we must have $c_{R,t}(h_t) \leq q(0)$.

Given that the claims are true for arbitrary $(n-1)$ we proceed to show that they must also be true for $n$, with $n > 0$. The true statements for (n-1) referred to by **1),...,6)** while the statements to be proven for $n$ are instead referred to by **1'),...,6')**.

**1')** we need only consider $\mu_t(h_t) \in (\eta(n), \eta(n-1)]$ because for $\mu_t(h_t) > \eta(n-1)$ this is true by claim **1**. We first make a subclaim that there can be at most a finite number of rounds $j$ such that $\mu_{t+j}(h_t) \leq \eta(n-1)$. This is ultimately because the proposer always has the option of offering (fractionally more than) $\bar{c}_R$, which will be accepted and thus ensures a continuation value of *at least* $100 - \bar{c}_R$.

Suppose there is an equilibrium offer path which lasts an infinite number of rounds without agreement and gives the proposer $100 - \bar{c}_R$ but never has $\mu_{t+j}(h_t) > \eta(n-1)$ for any $j \in \mathbb{N}$. Let $a_{t+s}(h_t)$ be the equilibrium implied probability of acceptance in round

$t + s$ (conditional on $h_t$). Given that the proposer can obtain $100 - \bar{c}_R > 0$ for any $\mu_t(h_t) in[\eta(n-1), \eta(n)]$ we must have:

$$\lim_{m \to \infty} \sum_{j=0}^{m} \delta^j a_{t+j}(h_t) 100 \geq 100 - \bar{c}_R$$

However, for $N$ such that $\delta^N < \frac{100-\bar{c}_R}{200}$ we must have:

$$\sum_{j=0}^{N} a_{t+j}(h_t) > \frac{100 - \bar{c}_R}{200}$$

And so in $N$ rounds at least fraction $\frac{100-\bar{c}_R}{200}$ of the responders accept (all normal types). The same argument can be repeated $k$ times implying that:

$$\mu_{t+kN}(h_t) \geq \mu_t \left( \frac{200}{100 + \bar{c}_R} \right)^k$$

The right hand side of this equation is greater than 1 for large $k$, giving a contradiction, and so eventually we must have $\mu_{t+j}(h_t) > \eta(n-1)$.

Let $t'$ be the supremum of times such that $\mu_{t'}(h_t) \leq \eta(n-1)$. For $n = 1$ bargaining must end in round $t'$ with $c_{R,t'} = \bar{c}_R$. For $n > 1$, we must have $\mu_{t'+1}(h_t) > \eta(n-1)$. This immediately implies that $c_{R,t'+1}(h_t) \geq q(n-2)$ by claim **1)**. If $c_{R,t'} < q(n-1)$ the normal type responder would then strictly prefer to wait an extra round to obtain the lower price (because $c_{R,t'} < \delta q(n-2)$), which would imply $\mu_{t'}(h_t) = \mu_{t'+1}(h_t)$ a contradiction. This means that $c_{R,t'}(h_t) \geq q(n-1)$.

If $\mu_{t'}(h_t) < \eta(n-1)$ we must have $c_{R,t'+j}(h_t) \leq q(n-j)$ for $j \leq n$ by claim **6)** and so in conjunction with the argument of the previous paragraph $c_{t'+j} = q(n-j)$. This in turn implies the proposer's continuation payoff in round $t'$ is given by $V_{n-1}(\mu_{t'}(h_t))$ where this is defined in equation 3. If $\mu_{t'}(h_t) = \eta(n-1)$ on the other hand, then the proposer's continuation value is $W(n-1) = V_{n-1}(\mu_{t'}(h_t))$ by claim **5)**.

Now suppose $t' \neq t$ and consider time $t'-1$ where $\mu_{t'-1}(h_t) \in (\eta(n), \eta(n-1)]$. For the proposer's offer to be accepted with positive probability it must be that $c_{R,t'-1}(h_t) \geq q(n)$ or else the responder would wait for the offer of $c_{R,t'}(h_t) = q(n-1)$. But in which case this strategy cannot obtain the proposer the continuation value of $V_{n-1}(\mu_{t-1})$ which we know she could obtain by offering (fractionally more) than $q(n-1)$. To see this notice that her continuation value in the supposed equilibrium can be written as follows:

$$\begin{aligned}
&\frac{\mu_{t'}(h_t)-\mu_{t'-1}}{\mu_{t'}(h_t)}(100-c_{R,t'-1}(h_t))+\frac{\mu_{t'-1}(h_t)}{\mu_{t'}(h_t)}\delta V_{n-1}(\mu_{t'}(h_t))\\
=&\frac{\mu_{t'}(h_t)-\mu_{t'-1}(h_t)}{\mu_{t'}(h_t)}(100-c_{R,t'-1}(h_t))+\frac{\mu_{t'-1}(h_t)}{\mu_{t'}(h_t)}\frac{\eta(n-1)-\mu_{t'}(h_t)}{\eta(n-1)}\delta(100-q(n-1))+\frac{\mu_{t'-1}(h_t)}{\eta(n-1)}\delta W(n-1)\\
\leq&\frac{\eta(n-1)-\mu_{t'-1}(h_t)}{\eta(n-1)}(100-q(n))+\frac{\mu_{t'-1}(h_t)}{\eta(n-1)}\delta W(n-1)\\
<&V_{n-1}(\mu_{t'-1}(h_t))
\end{aligned}$$

Where the first inequality follows because of the $c_{R,t'-1}(h_t) \geq q(n)$, and the second inequality follows from the fact that $\mu_{t'-1}(h_t) > \eta(n)$, and from remembering that $\eta(n)$ is defined as the lowest belief at which the third line would be equal to the fourth (see equation 1). And so finally we have a contradiction, proving that $t' = t$. We have established that if $\mu_t(h_t) \in (\eta(n), \eta(n-1)]$ then $c_{R,t}(h_t) \geq q(n-1)$ and the proposer's continuation payoff must be at least $V_{n-1}(\mu_{t'}(h_t))$. Furthermore, $\mu_{t+1}(h_t) \geq \eta(n-2)$ for $n > 1$ or else the proposer would not obtain $V_{n-1}(\mu_{t'}(h_t))$. The claims for $c_{R,t+j}(h_t) \geq q(n-j)$ and $\mu_{t+j}(h_t) \geq \eta(n-1-j)$ for $j \geq 1$ then follow from claim **1)**.

**2')** This is simply accounting given claim **1')** and **6)**.

**3')** By assumption $\mu_t(h_t) \geq \eta(n)$ and $c_{R,t} \in (q(n), q(n-1))$. Suppose then that equilibrium acceptance of this offer implies $\mu_{t+1}(h_t, c_{R,t}) < \eta(n-1)$, then by claim **6)** $c_{R,t+j}(h_t, c_{R,t}) \leq q(n-j)$ for $j > 0$, but in which case all high types would optimally accept $c_{R,t}$ and so $\mu_{t+1}(h_t) = 1$. But that in turn implies $c_{R,t+1}(h_t) = \bar{c}_R$. This presents a contradiction for $n > 1$ and immediately implies the claim **3')** for $n = 1$. In either case, however, it must be true that $\mu_{t+1}(h_t) \geq \eta(n-1)$.

If on the other hand $\mu_{t+1}(h_t, c_{R,t}) > \eta(n-1)$ for $n > 1$ then by claim **1)** we have $c_{R,t+1}(h_t, c_{R,t}) \geq q(n-2)$, but in which case it is not optimal to accept $c_{R,t}$ in round $t$, and so $\mu_t(h_t, c_{R,t}) = \mu_{t+1}(h_t) < \eta(n-1)$, a contradiction. And so, $\mu_{t+1}(h_t, c_{R,t}) = \eta(n-1)$.

**4')** By assumption $\mu_t(h_t) \leq \eta(n)$ and $c_{R,t} < q(n)$. Suppose that $\mu_{t+1}(h_t, c_{R,t}) > \eta(n)$. By claim **1')** this implies that $c_{R,t+1}(h_t, c_{R,t}) \geq q(n-1)$, and so the normal type of responder would not optimally accept, ensuring $\mu_t(h_t) = \mu_{t+1}(h_t, c_{R,t}) \leq \eta(n)$, a contradiction.

**5')** If $\mu_t(h_t) = \eta(n)$, then the proposer can obtain at least the value $V_{n-1}(\eta(n)) = W(n)$ as given by equation 3, by offering (fractionally more than) $c_{R,t} = q(n)$ or $c_{R,t} = q(n-1)$ (by claims **3**, **3'** and **5**). Offering any particular $c_{R,t} \in [0, q(n)) \cup (q(n), q(n-1))$ cannot obtain this value. In particular, notice that offering $c_{R,t} < q(n)$ implies $\mu_{t+1}(h_t, c_{R,t}) = \mu_t(h_t)$ by claim **4')**, and so it simply delays proceedings, and means the proposer cannot obtain a payoff as great as $W(n)$. This means the proposer's equilibrium continuation value must be exactly $W(n)$.

**6')** Suppose $\mu_t(h_t) < \eta(n)$, then by Claims **3') and 5** the proposer can guarantee a value $V_n(\mu_t(h_t))$ by offering (fractionally more than) $q(n)$. Notice that $V_n(\mu_t(h_t)) > V_{n-j}(\mu_t(h_t))$ for $j > 0$ (for $j = 1$ this follows given how $\eta(n)$ is defined), which implies that offering strictly more than $q(n)$ cannot obtain such a utility. This then must imply that $c_{R,t} \leq q(n)$.

If the time $t$ offer is not accepted with positive probability then $\mu_{t+1}(h_t) = \mu_t(h_t) < \eta(n-1)$ and so claim **6')** is true immediately due to claim **6)**. If on the other hand this time $t$ offer is accepted with positive probability but $c_{R,t+j} > q(n-j)$ for some $j > 0$, then it is not optimal for the normal responder to accept the time $t$ offer, (she prefers instead to wait until time $t+j$), contradicting a positive probability of acceptance. Finally if $\mu_{t+j}(h_t) > \eta(n-j)$ for some $j > 0$ then this would imply $c_{t+j}(h_t) > q(n-j)$ by claim **1)**, but we have just claimed that this leads to a contradiction.

This proves statements **1')** through **6')**. These being true for all $n$ implies that any equilibrium must have the properties described in the Lemma. To prove that such an equilibrium does exist, all that remains is to specify is off path strategies and beliefs for the proposer, the responder's strategy being fully outlined above.

Suppose that $\mu_t \in (\eta(n), \eta(n-1)]$. If the proposer makes an offer, which is strictly less than $q(n)$, then as specified above no responder accepts. Given the responder continuation strategies laid out above then the proposer is content to follow the price path $(q(n-1), q(n-2), ..., q(0))$ from the next round onward making the responders rejection optimal.

If the proposer instead offers $c_{R,t} \in (q(n-k), q(n-k-1)]$ with $k \geq 0$ then as specified above, a responder will accept so that $\mu_{t+1} = \eta(n-k-1)$. Given responder continuation strategies the proposer is then indifferent between charging the offer path $((q(n-k-1), q(n-k-2), ..., q(0))$ or the path $((q(n-k-2), q(n-k-3), ..., q(0))$. By mixing and following the first price path with probability $\frac{c_{R,t}-\delta q(n-k-2)}{\delta(q(n-k-1)-q(n-k-2))}$ and the second with the complimentary probability, we can ensures that responders' behavior is optimal. If an offer above $\bar{c}_R$ is rejected, we can specify that $\mu_t(h_t) = 1$ forever afterwards and the proposer optimally offers $\bar{c}_R$, which is always accepted by both types.

Strategies then represent optimal choices for both the responder and proposer by construction at every possible history given the other's strategy, and beliefs are determined by Bayes' rule wherever possible, hence this is an equilibrium.