Abstract

How will agents behave when bargaining in the face of an upcoming deadline? If irrational types exist, committed to their bargaining positions, rational agents will imitate this tough behavior to gain reputational benefits, even though this may result in the deadline being missed. Notably, if agents are patient and irrational types are committed to fixed demands then agreement must necessarily follow a U shape: deals are made either at time zero, or arbitrarily close to the deadline, with a positive fraction of rational agents failing to agree. But what demands will rational agents make? The model gives clear answers when the the probability of commitment is small. Maintaining the fixed demand assumption for irrational types, rational agents should ask for half the available surplus regardless of their relative discount rates, guaranteeing this payoff as the behavioral perturbation vanishes. If irrational types can commit to time varying demands, however, then rational agents should imitate a generalized Rubinstein (alternating offers) demand to secure the associated time zero payoff in the limit. The link between reputational and alternating offer models is driven by the properties of symmetry and a respect for backward induction, common to both. In addition to insights into the mechanics of reputational bargaining, the findings also raise questions about the correct assumptions to make about types and the motivation of their commitment.

1 Introduction

This paper is about bargaining in the face of a deadline when some agents are crazy, committed to their demands, a setup adapted from the infinite horizon reputational bargaining model of Abreu and Gul [2000] (henceforth AG). The motivation is twofold: firstly the model can help explain ‘deadline effects’ in bargaining, such as 11th hour deals and substantial disagreement; secondly, when the prior probability of commitment is small, it gives clear answers about agents’ optimal strategies, yielding insights into the nature of reputational bargaining in more general contexts.

Developing the first motivation, bargaining with deadlines often seems to resemble brinkmanship, with parties holding out on incompatible demands until the deadline is imminent, and even then sometimes failing to agree entirely. An example I shall use to illustrate various points in the paper is the 2011 debt ceiling negotiations. In May of that year the US Treasury warned that
it would run out of money on August 2 unless Congress raised its borrowing capacity. Throughout the summer a tense standoff ensued between Republicans, demanding any rise in the debt ceiling be accompanied by extensive government spending cuts, and Democrats, demanding tax rises in any deficit reduction deal. Only on July 31 was an agreement announced, and only then passed into law on August 2 itself.\footnote{\url{http://en.wikipedia.org/wiki/United_States_debt−ceiling_crisis_of_2011}} In addition to raising the debt ceiling, the final deal included spending cuts, but no tax rises and was widely regarded in the press as a major Republican victory. Such success, despite the Democrats controlling both the Senate and White House, was attributed to the Republicans’ ‘irrational’ unwillingness to compromise.

The exact timing of the deadline is assumed to be uncertain in the model, it is continuously distributed on some finite interval. The presence of agents who are genuinely unwilling to compromise, means rational agents will choose to imitate ‘tough’ behavior in order to gain reputational benefits. Although the risk that the deadline may accidentally pass then creates incentives for a rational agent to reach immediate agreement, the hope that a possibly rational opponent will back down first creates incentives to hold out, leading to delay and potentially disaster.

An uncertain deadline is a natural modeling choice for many environments. In particular, despite the Treasury’s claims some Wall Street analysts claimed the country could remain solvent until August 10, or even August 15. The reputational advantage of acting tough suggests Republicans’ behavior may not have actually been crazy at all, but merely an attempt to gain a strategic advantage, just as Democrats instance on tax rises in any deal ultimately turned out to be a bluff. Finally, the incentive to hold out when the true deadline is uncertain suggests that the final deal to avoid default was not inevitable, and future standoffs may well end differently.

Under this modelling setup, I show how deadline effects must necessarily occur so long as agents are relatively patient and behavioral types are simple, in the sense of not changing their demands over time. Bargaining must then either result in an agreement at time zero, or at the 11th hour, or result in disagreement. Moreover, rational agents are among those who must sometimes miss the deadline.

Despite this, delay and disagreement typically will not occur when the prior probability of commitment types is sufficiently small. Indeed the model yields tight predictions about the optimal behavior of rational agents in this case, which do not depend on the relative likelihood of particular behavioral types, so long as the type space is sufficiently rich. These predictions represent the second motivation for the model outlined above, and represent the paper’s most important theoretical contribution.

The exercise of making the probability of commitment (vanishingly) small can be viewed as a implementing the minimal perturbation of a complete information bargaining game. The findings can then provide a robustness check on complete information results, or a selection criterion if multiple equilibria exist. Alternatively, taking the reputational model more seriously, the limit results highlight important forces at work that should be relevant even for larger probabilities of commitment. Doing this in a non-stationary environment (the bargaining problem appears very different 6 months before a deadline, than when only hours or minutes remain) leads to a richer understanding of reputational forces that should be relevant far beyond the current context, and are obscured in a stationary infinite horizon model, such as AG’s.
I distinguish between two kinds of commitment behavior. The simple types, described above, are committed to obtaining a fixed share of the bargaining surplus, while sophisticated commitment types may vary their demands over time. Only simple types are considered in AG. Abreu and Pearce [2007] (henceforth AP) allow for sophisticated types when agents bargain over an enforceable long term contract governing play in an infinitely repeated game. They find, however, that such types are irrelevant for limit payoffs, so long as there is a simple type committed to the stationary strategy of always demanding her Nash Bargaining With Threats payoff (Nash [1953]) and playing her Nash threat action so long as it is refused. In this paper by contrast, sophisticated types do matter.

If all behavioral types are simple then, regardless of her discount rate, each agent should imitate a type demanding half the available surplus, guaranteeing this payoff as the probability of commitment becomes small. If some behavioral types are sophisticated, however, in particular if there is a type committed to what I call the generalized Rubinstein demand (Rubinstein [1982]), then rational agents can secure (almost) the associated time zero payoff by imitating it.

The generalized Rubinstein demand is defined as the solution to a complete information (no behavioral types) alternating offers game for the deadline environment, when the time between offers becomes arbitrarily small. This demand reflects both agents’ discount rates and the risk of missing the deadline. It varies over time (potentially non-monotonically). Agent i’s demand is a convex combination of her infinite horizon Rubinstein demand $r_i + r_j$, and $\frac{1}{2}$; it converges to the latter as the agent becomes convinced the deadline is imminent.

The intuition for the simple types result is that rational agents should consider how they will behave at 10 seconds to midnight. When the prior probability of commitment is small, equilibrium requires that bargaining must continue until almost the last possible minute with some small positive probability. At that point both agents have approximately the same (high) impatience for a deal as the deadline is imminent, and so can expect an equal share. The optimality of demanding half the surplus at the end of bargaining then rolls back to the beginning. A more precise explanation is wrapped up in the war of attrition logic, which describes the equilibrium, and is explored more thoroughly in the text.

The intuition for why the Rubinstein demand is always successful when allowing sophisticated types is more subtle. As noted, the Rubinstein demand converges to $\frac{1}{2}$ as the deadline becomes increasingly imminent. Working back from that point, however, discount rates can play a role if demands can vary, and indeed the backward induction logic of the Rubinstein demand turns out to exactly balance the rates at which two agents build reputation (for being committed to their demands) in equilibrium. The result is driven by two features that are common to both reputational bargaining and the alternating offers model. First, they are both symmetric, providing equal innate bargaining power to agents, and second, they both have a strong respect for backward induction. Again, a more precise explanation must elaborate on the nature of a reputational equilibrium in more detail.

Regarding the symmetry of both solutions, it must be noted that the alternating offers protocol has been criticized for its sensitivity to assumptions such as the relative frequency of agents’ offers; the ‘natural’ equal offer frequency model, however, treats agents symmetrically as the absolute time between offers becomes small. AG show that this symmetry is fundamental to the reputational model, with outcomes becoming independent of the relative frequency of agents’
offers as the absolute time between offers shrinks to zero.

This finding that Rubinstein demand is the model’s canonical type, relates to the broader attempt to provide payoff bounds through reputational effects in games, when a wide variety of possible types is entertained. This literature goes back at least to Fudenberg and Levine [1989], who show that a patient long-run player can guarantee her Stackelberg payoff against a short run opponent, so long as there is a positive probability of such a commitment type. To my knowledge, however, this is the first attempt to find such optimal types in a non-stationary environment.

Although AG do find that with simple types and a stationary infinite horizon, rational agents can obtain their stationary Rubinstein payoff, \( \frac{r_i}{r_{j+1}} \), as the probability of commitment becomes small, the result offers a slight contrast to AP and Kambe [1999], who both highlight the relevance of the Nash solution. Kambe’s model differs slightly from AG: agents can make any demand before facing a fixed probability of becoming committed to it, resulting in a unique pure strategy equilibrium with immediate agreement.

In stationary bargaining environments there is a very close link between the Nash and Rubinstein solutions. Under standard axioms on time preferences, discount factors and utility functions are not uniquely determined. This allows Binmore et al. [1986] to show that under a suitably chosen utility transformation, the two solutions exactly coincide as time between offers becomes arbitrarily small. Even without such a transformation the Rubinstein demand may be regarded as the Nash solution with bargaining weights inversely proportional to discount rates.\(^2\) For non-stationary environments by contrast, there is no general link between the Nash and Rubinstein solutions, as shown in Coles and Muthoo [2003].

The finding that the Rubinstein type is canonical in the current non-stationary setting, might then seem to suggest that the link between the reputational model and the Nash solution is largely coincidental. There is one important feature of the Nash solution, which remains relevant, however, which is that of symmetry. Like Nash, the alternating offer and reputational models want to treat equal agents equally, but do so while respecting backward induction and treating discount rates as one important reason for discrimination.

Another important takeaway from the paper is a question raised by the contrasting findings with simple and sophisticated type spaces: are sophisticated types plausible? If not then the ‘advice’ for rational agents bargaining in the face of a deadline changes dramatically. This is important if we take the reputational model seriously and view the limit results as illustrating forces which will be relevant for larger prior probabilities of commitment.

Intuitively, an agent who changes her demand over time does not seem committed to it, indeed this almost seems true by definition. She must sometimes refuse to consider an offer which only moments before she would happily have taken and sometimes refuse to consider an offer which she will happily accept at a later date. Furthermore, the model requires behavioral types to be transparent, identifiable from each other, for reputational effects to work. This effectively means that future demand profiles must be announced at time zero, making non-stationary demands

\(^2\)It may be countered that there appears to be no such link between the models when outside options are present. Even when agents are equally patient and outside options are \( d^i < \frac{1}{2} \) the Rubinstein demand splits the surplus equally, while agent \( i \)’s Nash solution is \( \frac{1}{2}(1 + d^i - d^j) \). Even here, however, if discount rates are instead interpreted as a constant risk of negotiation breakdown (in which case agents actually obtain their outside option) then \( i \)’s Rubinstein demand is also \( \frac{1}{2}(1 + d^i - d^j) \).
even less plausible; after announcing their demands two known behavioral agents must willfully delay an agreement, which they are committed to accepting at a later date.

Wolitzky [2011] shows this transparency is required by adapting AP’s model to allow types to distinguish themselves (from other types) only as bargaining unfolds; for instance, he allows a ‘soft’ behavioral type that outwardly resembles a tough Nash Bargaining With Threats type, but concedes to any opponent’s demand probabilistically over time. This may lead rational agents to never imitate behavioral types, with almost any efficient, feasible, and individually rational payoff attainable in equilibrium. Part of the appeal of simple types may then be precisely their transparency: announcing that you will never accept anything less than half of the surplus, whether credible or not, is easy to understand.

More generally, the issue with sophisticated types seems to concern their underlying motivation. The behavior of simple types can be relatively easily be derived from payoff types. For instance, agents may have differing views of a ‘fair’ surplus division and very strong (!) preferences in favor of that division. It would, however, seem very difficult to extend that motivation to non-stationary demands, as views of fairness should be time invariant. This is not to suggest that simple commitment types are obviously correct, the preference justification given above may seem farfetched too. The point is rather to highlight the tension in the reputational setup, where crazy behavior is exogenously imposed instead of being endogenously derived, and to suggest that more thought be given to motivating commitment.

This question about the reasonableness of some commitment types has been raised before. Notably, Mailath and Samuelson [2001] investigate the incentives for firms to develop a reputation for high quality in the presence of inept behavioral types. Their justification for ignoring Stackelberg leader types is that such types would need to have extremely strange preferences compared to a normal firm, whereas inept types may have very similar preferences, but a higher cost of effort.

The rest of the paper is structured as follows: in the remainder of this section I review additional related literature; section 2 lays out the model; section 3 establishes the existence of a unique equilibrium; section 4 explores some of the basic differences from the AG model, and in particular investigates deadline effects, section 4.3; section 5 presents the main theoretical results when the probability of commitment is small; and section 6 concludes. Unless stated otherwise, all proofs are in the Appendix.

1.1 Further related literature

The reputational bargaining literature grew out of the broader literature of reputational effects in games started by Milgrom and Roberts [1982], Kreps and Wilson [1982] and Kreps et al. [1982]. Building on the previously mentioned Fudenberg and Levine [1989], important contributions in understanding reputational payoff bounds in games with two long run agents are provided in Schmidt [1993], Cripps et al. [1996], Celetani et al. [1996], Aoyagi [1996] and Evans and Thomas [1997]. Myerson [1991] was the first to formally introduce behavioral types into the bargaining problem.

Building on the current paper Fanning [2013a] and Fanning [2013b] consider reputational bargaining in two additional non-stationary environments. Both again consider simple and sophis-
ticated behavioral types, finding that a generalized Rubinstein type is canonical in the larger type space.

The first paper considers bargaining over multiple issues with multiple deadlines. It provides insights into agenda setting: rational agents would like the issue which they care relatively more about debated second. The second paper considers bargaining in the face of uncertainty about the future bargaining environment, which is later revealed. With only simple types there are incentives for rational agents to make aggressive demands that are optimal only in states of the world in which the bargaining environment (relatively) favors them. This can cause demands to diverge, with no agreement being reached until the resolution of uncertainty, even as the probability of commitment becomes arbitrarily small.

Also building on the AG model, Abreu et al. [2012] consider reputational types which are simple in the demands they make, but may differ in the time at which they make their initial demand. They find that when there is uncertainty about agents discount factors, non-Coasean outcomes can occur in which rational patient agents delay initial demands in order to separate from non-patient agents even as the probability of commitment becomes small.

Wolitzky [2012] considers non-stationary commitment bargaining in a model similar to Kambe [1999]. He finds that the maxmin bargaining posture in a stationary infinite horizon environment is a fixed surplus share plus compensation for any delay in reaching agreement.

Motivated by the findings of the current paper regarding the justification of commitment types, Fanning [2013c] shows how if a seller has known ‘reasonable’ stationary preferences for fairness (paying a continuous increasing utility cost for unequal material outcomes) and the buyer’s valuation of a good is uncertain, the unique equilibrium may be non-Coasean with high initial prices even as the time between offers becomes small. When the seller may or may not be fair, equilibrium can resemble a war of attrition with normal sellers and high valuation buyers imitating the (potentially non-stationary) commitment behavior of fair sellers and low value buyers. Nonetheless, important differences from the crazy type model remain, such as the existence of multiple equilibria.

Inderst [2005] also considers a crossover model between reputational bargaining and bargaining in the presence of unknown valuations, assuming a possibly committed seller and a buyer with unknown valuation. He motivates simple, fixed demand, commitment types for the seller by the possibility that she may be an employee of a distant principal without the authority or incentives to lower prices.

Deadlines have of course been incorporated into numerous bargaining models, with those deadlines modeled in many different ways. I focus here on papers that are concerned with deadline effects, that is, delay prior to the deadline and possible disagreement.

In an alternating offer model with a hard deadline in which agents must wait to receive offers before making a counteroffer, an initial proposer will strategically delay her first offer in order to present her opponent with an ultimatum at the deadline. Ma and Manove [1993] extend this setup to allow offers to arrive with a stochastic delay and show the possibility of equilibrium disagreement and rejection of offers which arrive early.

Fershtman and Seidmann [1993] posits that after rejecting an offer agents becomes committed to not accepting lower ones, perhaps because of the loss of face this would require. This means
offers prior to a hard deadline allow an opponent to increase her commitment power. With a random proposer identity in each period, an agent will offer her opponent none of surplus and reject compromise offers prior to the deadline, in order to secure all the surplus if she proposes at the deadline, and her commitment demand otherwise.

Ponsati [1995] directly assumes a concession game structure in which agents can agree to one of two alternatives prior to a hard deadline. There is two sided incomplete information about valuations, which may be negative. The unique equilibrium involves probabilistic concession by both agents at the deadline and no concession on some interval strictly prior to the deadline.

Simsek and Yildiz [2008] allow for agents who are optimistic (no common prior) about their future bargaining power (for instance regarding proposal rights). A jointly inconsistent expectation of having bargaining power at the deadline and hence extracting a large payoff, causes agreement to be delayed until close to that deadline.

Fuchs and Skrzypacz [2012] consider a hard deadline when a buyer has an unknown valuation. Even as offers become frequent the seller obtains positive surplus by retaining the option to make a ‘final’ offer at the deadline, where a positive atom of trade occurs.

Spier [1992] investigates a model of pretrial negotiation, with one-sided incomplete information about the liability of the defendant and fixed costs of delay in each period. Settlement behavior follows a U shape: agreement occurs either immediately, or just prior to the hard deadline, with some cases proceeding to court. This effect persists even as agents make offers frequently because discounting cannot screen between defendant types who (absent common fixed costs) would prefer to settle later. This agreement pattern is highly similar to the deadline effects observed in this paper, although the mechanics of the models are very different, in particular my model requires two sided incomplete information.

Roth et al. [1988] empirically demonstrates a deadline effect in a relatively unstructured bargaining experiment, agreement rates spike prior to the deadline and many subjects fail to agree. I compare the results of that paper to the current model in section 4.3. Ockenfels and Roth [2006] finds somewhat comparable deadline effects in ebay auctions.

2 The model

In this section, I first outline the structure of the model, before discussing the assumptions, and drawing some some initial conclusions about the nature of any equilibrium.

2.1 The fundamentals

Agents 1 and 2 bargain over the division of surplus whose value is 1 (dividing a dollar), and discount the future at rate \( r' \geq 0 \). There is some deadline \( t \) by which time bargaining must be completed or else the surplus is lost. \( t \) is distributed according to the distribution \( G \) on \([0, T]\) with continuous density \( g \) and \( g(T) > 0 \). If parties attempt to make an agreement at some time \( t < T \), with probability \( G(t) \) the deadline has already passed. The assumption \( r' + g(t) > 0 \) ensures agents would strictly prefer a given surplus division to be agreed immediately rather
than waiting.

The structure of bargaining is as follows: at time zero agent 1 announces a complete bargaining position $\alpha^1 \in [0, 1]^{0,T}$. $\alpha^i(t)$ details the surplus share which she will demand from agent 2 at time $t$ if a deal has not already been reaching and the deadline has not already passed. Following this agent 2 can either immediately agree to the offer $(1 - \alpha^1(0))$, or make a counter demand $\alpha^2 \in [0, 1]^{0,T}$. Following these announcements agent i can concede to j’s current offer $(1 - \alpha^j(t))$ at any $t \leq T$. If both agents concede at exactly the same time, one of the two proposed divisions is implemented with probability $\frac{1}{2}$.

Reputational dynamics come from a prior probability that agents are irrational $\zeta > 0$. A committed type of agent i is identified by a bargaining position $\alpha^i \in (0, 1)^{0,T}$: a type $\alpha^i$ always demands $\alpha^i(t)$ at time $t$ and accepts any offer greater or equal to that, rejecting smaller offers. $C^i$ is the finite set of irrational types, who conditional on irrationality, have probability $\pi^i(\alpha^i)$. I assume that for all irrational types $\alpha^i(t)$ is continuously differentiable with $(1 - \alpha^i(t))(\frac{800}{1000} + r^i) > -\alpha^i(t)$. This last condition, that an irrational type’s offer, $(1 - \alpha^i(t))$, does not increase too quickly, ensures that if agent j is certain she faces an irrational agent i, she would always rather concede immediately than wait for a later, more advantageous demand of that irrational agent. To simplify the statement of some theorems, I assume that $\max(\alpha^1(0) \in C^1) + \min(\alpha^2(0) \in C^2) > 1$. Simple types’ demands satisfy $\alpha^i(t) = 0$, but this need not be true for sophisticated types.

Given the game description above, a rational agent’s strategy must describe her initial demand function choice and her decision about when (if ever) to concede to her opponent, given demands. The solution concept is Perfect Bayesian Equilibrium.

### 2.2 Discussion

I first address the assumption made above about the stochastic deadline. As discussed in the Introduction uncertainty about the exact time at which the deadline arrives is natural for many environments. In the example of the 2011 debt ceiling negotiations, this reflected the possibility that Treasury might have defaulted prior to August 2, or as late as August 15.

In addition to its intuitive plausibility the continuous nature of the deadline distribution ensures that outcomes do not depend on the identity of the agent who has the right to make a ‘final’ offer before a known hard deadline. Notice that ‘almost final’ offers made close to $T$ in the stochastic deadline model, concern a situation in which the deadline is almost certain to have already passed.

It is not important to the model whether agents find out if the deadline has passed while they are bargaining, or if they only find out after attempting to strike make a deal. Even if agents are initially uninformed, they will act on the assumption that the deadline has not passed as their actions only have payoff consequences in that case. A final interpretation of the model is that agents with linear utility bargain over the division of a perishable good, a fraction $G(t)$ of which is destroyed by time $t$.

A highly related model in which the deadline is hard at $T$, but deals take an unknown time to implement can also be defined. If the density of deals implemented at time $t + w$ following
an agreement at time $t$ is given by $g(T - w)$ then one recovers exactly the same probability $G(t)$ that a deal agreed at time $t$ will miss the deadline of $T$. The only difference between the models concerns the exact timing of implemented deals, which matters for discounting payoffs. The additional layer of complexity in evaluating payoffs makes describing this uncertain implementation model slightly less transparent, however, when $r' = r' = 0$ the models are formally identical. All the results developed in the paper can be extended almost without change to this uncertain implementation model. I lay out the slight adjustments one must make to those results in the Appendix, section 7.1. The motivation for this related model in the context of the debt ceiling example is that the complexities of passing a bill into law, which takes a positive but uncertain amount of time, meant the Treasury’s borrowing authority might not have been raised by August 2, even though the principles of an agreement were struck on July 31. In particular, even a single senator can filibuster a bill for at least two days, despite the objections of 99 colleagues.

The second issue which needs to be discussed is the continuous time nature of bargaining. With no behavioral types, in the subgame following demand choices one has a standard war of attrition, allowing multiple equilibria with ‘always concede’ and ‘never concede’ in particular being equilibrium strategies. Incorporating demand choice, any division of the pie may then be an equilibrium. Taking the limit of a discrete time bargaining game, with alternating offers for instance, will result in a uniquely defined outcome. Why then would we be interested in the continuous time game?

The answer is that the game defined above represents a series of short cuts and simplifications, which are reasonable to make in the context of reputational bargaining. As mentioned, AG consider a fully discrete time infinite horizon game, with simple reputational types, before showing that as the absolute time between offers becomes small, all equilibria converge in distribution to the unique equilibrium of a continuous time war of attrition game, regardless of the relative frequency of agents’ offers. I am jumping straight into the continuous time game, without reproving AG’s full convergence result for the slightly different bargaining environment.

The reasoning behind AG’s convergence proof is that if an agent was ever revealed as rational, but her opponent is possibly irrational, she must concede to her opponent (almost) immediately if the time between offers is sufficiently small. The logic is Coasean (Coase [1972]): the known rational agent, i, must become increasingly pessimistic about j’s rationality the longer neither has conceded, and so eventually must be convinced of her commitment and so concede. Sufficiently close to a final date $t'$ by which i must concede, however, even a rational j will not concede. But facing no concession by rational or committed j agents on $[t' - \varepsilon, t']$ i must optimally concede at $t' - \varepsilon$. This means that an agent’s strategy effectively reduces to ‘keep imitating a behavioral type’ or ‘concede’.

In the context of the continuous time bargaining game with behavioral types and deadlines, defined above, offers are effectively arbitrarily frequent already, resulting in an immediate, not approximate, concession game structure. This is shown in the lemma below which is based on the proof of AG, Proposition 5.

**Lemma 1.** If at some $t' \in [0, T)$ agent $j$ is possibly committed to her demand, and agent $i$ is certainly not, agent $i$ must concede immediately in any equilibrium.

Given the actual setup of the model, the only time at which $i$ could ever reveal herself as rational
is at time zero. One can, with some added complexity, also allow agents to reveal rationality without conceding at other times in the \((0, T^*)\) while maintaining the continuous time bargaining structure. The lemma above, however, shows that rational agents would never actually want to take advantage of such opportunities, given the assumptions made about commitment types. In addition, the lemma implies that an agent’s initial demand choice can be reduced to deciding which irrational type to imitate, as she must imitate some type.

Another issue for discussion is the assumption of initial demand announcements. Why does each agent have to announce a complete bargaining plan, instead of making these announcements sequentially over time? One might think that this is again simply a mathematical simplification for the continuous time environment, and that similar assumptions could be obtained if agents only announced their initial continuous time bargaining function on some interval \([0, \varepsilon)\), before at \(t = \varepsilon\) announcing a function to cover demands on \([\varepsilon, 2\varepsilon)\) and so on. The assumption is, however, more substantive. The issue is related to the transparency of types, discussed briefly in the Introduction.

For reputational forces to work, it must be clear which type a rational agent is pretending to imitate. Without this, a soft type who concedes frequently to any opponent’s demand, may initially look outwardly similar to a tough type. As Wolitzky [2011] shows this can destroy any incentive for rational agents to imitate commitment types. The initial announcement of future demands then, in addition to its mathematical simplicity for a continuous time environment, plays an important role in allowing agents to ‘announce a type’.

A further issue that must be addressed is the assumptions made about commitment types. I have assumed that commitment types belong to a particularly class: effectively continuously differentiable demand functions. The reason for this is simplicity: this class will be shown to be big enough to illustrate the paper’s main results. I extend the model in section 5.5 to incorporate types which make discontinuous, and potentially history contingent demands. The purpose of the exercise is to demonstrate the robustness of the result that agents can guarantee their generalized Rubinstein payoff when the prior probability of commitment is small, if such a type exists. In that extension agents announce a complete history contingent bargaining function, effectively their type, in order to ensure transparency, however, the possibility of discontinuities means that Lemma 1 need no longer hold.

The special interest in simple types, stationary demands, is their intuitive plausibility compared to sophisticated types, non-stationary demands. The interest in sophisticated types, by contrast, is to allow for a sufficiently broad type space to answer questions about canonical types, those which the model ‘wants’ agents to adopt. In particular, the generalized Rubinstein demand will be non-stationary in the deadline environment.

### 2.3 Strategies and utility

As stated above, Lemma 1 implies that strategies can be reduced to having rational agents first deciding which types to imitate and second, given incompatible demands, deciding at what time (if ever) to concede. Furthermore, given an initially incompatible demand announcement and continuous demand paths, if \(\alpha^i(t) + \alpha^i(t) = 1\) at any \(t > 0\) then committed agents will concede, and rational agent i is guaranteed at least \(\alpha^i(t) = (1 - \alpha^j(t))\) in any continuation game.
by conceding immediately. The assumption $r' + g(t) > 0$ then implies any delay must strictly decrease one agent’s expected payoff, and therefore agreement must always be immediate. This ensures that it is without loss of generality to assume that rational agents always concede before committed agents.

This means a strategy for agent 1, $\sigma^1$, is described by a probability distribution $\mu_1$ on the set $C^1$, describing the bargaining postures adopted by rational agents, and a collection of cumulative distributions $F_{\alpha^1,\alpha^2}^1(t)$ is the total probability of agent 1 conceding to agent 2 by time $t$ following demands $\alpha^1$ and $\alpha^2$. The strategy of never conceding is captured by a concession time of $T$.

A strategy for agent 2, $\sigma^2$, is similarly made up of $\mu_2^2$ on $C^2 \cup \{Q\}$ specifying rational 2’s counterdemand to $\alpha^1$ (where Q is immediate acceptance), and $F_{\alpha^1,\alpha^2}^2$ which describes 2’s choice of concession time given demands $\alpha^1, \alpha^2$. Given $\mu_1$ and $\mu_2^2$, the posterior probability that agents are irrational after making their demands is given by:

\[
\bar{z}^1(\alpha^1) = \frac{z^1\pi^1(\alpha^1)}{z^1\pi^1(\alpha^1) + (1 - z^1)\mu(\alpha^1)}
\]

\[
\bar{z}^2(\alpha^2) = \frac{z^2\pi^2(\alpha^2)}{z^2\pi^2(\alpha^2) + (1 - z^2)\mu(\alpha^2)}
\]

Conditional on a war of attrition being started, with demands $\alpha^1, \alpha^2$, agent i’s expected utility from conceding at time $t$, given an opponent’s strategy $\sigma^j$, is:

\[
U^i(t, \sigma^j|\alpha) = \int_0^t e^{-r^j} \alpha^j(s)(1 - G(s))dF_{\alpha^1,\alpha^2}^j(s)
\]  

\[
+ (1 - F_{\alpha^1,\alpha^2}^j(t))e^{-r^j}(1 - G(t))(1 - \alpha^j(t))
\]

\[
+ \left(F_{\alpha^1,\alpha^2}^j(t) - F_{\alpha^1,\alpha^2}^j(t-\epsilon)\right)e^{-r^j}(1 - G(t))\frac{1}{2}(\alpha^j(t) + 1 - \alpha^j(t))
\]

A rational agent’s utility at the start of the war of attrition is therefore:

\[
U^i(\sigma|\alpha) = \int_0^T U^i(t, \sigma^j|\alpha)dH_{\alpha^1,\alpha^2}^j(t)
\]

Where $H_{\alpha^1,\alpha^2}^j(t) = \min\left\{\frac{F_{\alpha^1,\alpha^2}(t)}{1-\bar{z}^j(\alpha^1)}, 1\right\}$ describes a rational agent 1’s behavior, with a similar expres-

---

\[3\] Where: $F_{\alpha^1,\alpha^2}^j(t) - F_{\alpha^1,\alpha^2}^j(t-\epsilon) = F_{\alpha^1,\alpha^2}^j(t) - \lim_{s \to t} F_{\alpha^1,\alpha^2}^j(s)$
sion for agent 2. Finally a rational agent’s expected utility in the full game is:

\[
U_1(\sigma) = \sum_{\alpha_1} \mu_1(\alpha_1) (1 - z^2) \mu_{a_1}^2(Q) + z^2 \sum_{\alpha_2(0) \leq 1 - \alpha_1(0)} \pi_2(\alpha_2) \\
+ \sum_{\alpha_2(0) > 1 - \alpha_1(0)} U_1(\sigma|\alpha)((1 - z^2) \mu_{a_1}(\alpha_2) + z^2 \pi_2(\alpha_2))
\]

\[
U_2(\sigma) = \sum_{\alpha_1} ((1 - z^1) \mu_1(\alpha_1) + z^1 \pi_1(\alpha_1))
\]

\[
\times \left(1 - \alpha_1(0)) \mu_{a_1}^2(Q) + \sum_{\alpha_2(0) > 1 - \alpha_1(0)} U_2(\sigma|\alpha) \mu_{a_1}^2(\alpha_2)\right)
\]

The equations in 4 are exact analogues of agents’ payoffs in AG. The only difference between the games is that the payoff \(U_i^e(\sigma|\alpha)\) is determined in a slightly different manner, reflecting different continuation games. In what follows I frequently drop the subscripts of \(F_{\alpha_1, \alpha_2}\) and other variables, when the context should be clear, and also frequently refer to the equilibrium posterior probabilities of irrationality after demand choices, defined in equation 1, as simply \(z_i^e\).

3 Equilibrium

I first consider the game with only one irrational type for each agent. The following lemma shows that any equilibrium of the game must satisfy three simple properties.

**Lemma 2.** If there is a single behavioral type for each agent, \(C^i = \{\alpha^i\}^4\) for \(i = 1, 2\), then any equilibrium must satisfy the following properties:

(i) There is some time \(T^* < T\) such that both agents are indifferent between concession at any \(t \in (0, T^*)\)

(ii) Both agents reach a posterior probability of irrationality of 1 at exactly time \(T^*\)

(iii) At most one agent concedes with positive probability at time zero

The proof of this lemma is again indebted to AG, and their Proposition 1. The basic reasoning behind it is as follows. Both agents must reach probability 1 of irrationality at the same time because a rational agent would concede immediately when certainly facing a committed opponent. Only one agent can concede at time zero because if both conceded, then waiting until an instant after time zero before conceding would strictly increase agent’s profits. More generally, concession behavior must be continuous because mass concession (concession with strictly positive probability) by agent \(i\) at any time \(t > 0\) will induce agent \(j\) to wait to receive a profit bump on the interval \([t - \varepsilon, t]\), but in which case \(i\) would prefer to concede at \(t - \varepsilon\).

---

4Notice that if \(C^i = \{\alpha^i\}\) the assumption \(\max\{\alpha^1(0) \in C^1\} + \min\{\alpha^2(0) \in C^2\} > 1\) means that these types must be incompatible.
rather than wait until \( t \) to concede, a contradiction. But for agents to concede continuously on an interval \((0, T^*)\) they must obtain constant utility on that interval.

Given the above necessary characteristics of any equilibrium, it is easy to see that an equilibrium exists and is unique. For agent \( i \) to be indifferent to concession on the interval \((0, T^*)\), her utility \( U^i(t, \sigma^i|\alpha) \) from time \( t \) concession, defined in equation 2, must be differentiable with respect to \( t \) with a derivative of 0. This immediately implies the following equilibrium concession rate for agent \( j \):

\[
\frac{f^j(t)}{1 - F^j(t)} = \frac{(1 - \alpha^j(t))(r^j + \frac{g(t)}{1 - G(t)}) + \alpha'^j(t)}{\alpha^j(t) + \alpha^i(t) - 1} \tag{5}
\]

This uniquely defines \( F^j \) up to a boundary condition. To close the model let \( T' \) be the minimum time such that \( \alpha^j(t) + \alpha^i(t) = 1 \), or \( T \), if that minimum is not well defined. And let \( \hat{F}^j(t) \) be defined as follows for \( t \leq T' \):

\[
\hat{F}^j(t) = 1 - \exp \left( - \int_0^t \frac{(1 - \alpha^j(s))(r^j + \frac{g(s)}{1 - G(s)}) + \alpha'^j(s)}{\alpha^j(s) + \alpha^i(s) - 1} ds \right) \tag{6}
\]

In the special case of simple types \( \alpha'^j(t) = \alpha'^i(t) = 0 \) for all \( t \), this simplifies to:

\[
\hat{F}^j(t) = 1 - \left( e^{-r^j(1 - G(t))} \right)^{K^j} \tag{7}
\]

Where:

\[
K^j = \frac{1 - \alpha^j}{\alpha^j + \alpha^i - 1} \tag{8}
\]

\( \hat{F}^j(t) \) records the probability of concession by agent \( j \) at time \( t \) on the assumption that she does not concede with positive probability at time zero. For \( t < T^* \), one must therefore have \( F^j(t) \) given by:

\[
F^j(t) = 1 - c^j(1 - \hat{F}^j(t)) \tag{9}
\]

Where \( (1 - c^j) \) is the probability that agent \( j \) concedes at time zero. Given \( j \)'s posterior probability of commitment after demand choices, \( \bar{z}^j \), her equilibrium reputation for commitment at \( t \in [0, T^*] \), is given by:

\[
\bar{z}^j(t) = \frac{\bar{z}^j}{1 - F^j(t)} = \frac{\bar{z}^j}{c^j(1 - \hat{F}^j(t))}
\]

This reputation grows gradually with time according to agent \( j \)'s rate of concession to \( i \) in the war of attrition. Equilibrium requires both agents reach probability 1 of irrationality at the same time, \( \bar{z}^j(T^*) = \bar{z}^i(T^*) = 1 \). These two equations uniquely pin down the equilibrium’s two unknowns \( T^* \) and \( \frac{\zeta}{\bar{z}} \), where the latter is a single unknown, given that only one agent can concede with positive probability at time zero.

The intuitive way to solve for this equilibrium, is to work out which agent would ‘win the race’ to become irrational with certainty, assuming no initial mass concession by either agent. Let \( T^j < T' \) be the exhaustion time of agent \( j \), defined as the time by which she must necessarily
reach probability 1 of irrationality even without initial mass concession.\footnote{Given that \((1 - \alpha(t)) \left( r_i + \frac{r_j(t)}{1 - G(t_j)} \right) > -\alpha'(t)\) appropriate bounds on concession rates show that \(\hat{F}(T') = 1\). This combined with the fact that \(\hat{F}(t)\) is continuous and strictly increasing show that \(T_j < T'\) is well defined.}

\[
\frac{\bar{z}_j}{1 - \hat{F}_j(T_j)} = 1 \tag{10}
\]

Equilibrium requires that \(T^* = \text{min}(T_i, T_j)\). To ensure that both agents reach probability 1 of irrationality at the same time, one must then adjust the time zero concession of the agent for whom \(T_j > T^*\). That is:

\[
c_j = \frac{\bar{z}_j}{1 - \hat{F}_j(T^*)} \tag{11}
\]

I summarize these findings in the following proposition.

**Proposition 1.** If there is a single behavioral type for each agent, \(C_i = \{\alpha_i\}\) for \(i = 1, 2\), then there is a unique equilibrium defined by equations 6, 9, 10, 11.

The equilibrium requirement that both agents must be indifferent between conceding at any time on the interval \((0, T^*)\) immediately implies rational agent j’s utility at the start of the war of attrition is:

\[
U_j(\alpha) = (1 - c^i)\alpha_j(0) + c^i(1 - \alpha_i(0)) \tag{12}
\]

This shows agent j can obtain a payoff higher than her opponent’s initial offer only when there is positive probability of i conceding at time zero. It is clearly important therefore, other things equal, to have a low exhaustion time.

Agent j’s exhaustion time is reduced, the higher her posterior probability of irrationality after demand choices, \(\bar{z}_j\), and also the higher her rate of concession. At any given time agent j’s concession rate is given by equation 5. This is increasing in the hazard rate of the deadline distribution \(\frac{g(t_j)}{1 - G(t_j)}\), agent i’s rate of impatience \(r_i\), the generosity of j’s offer \((1 - \alpha_j)\) and the rate at which j’s demand increases \(\alpha'_j(t)\), but is decreasing in the size of agent i’s demand. Given that this rates can change greatly over time, however, the relationship between bargaining positions and the likelihood of an agent receiving initial mass concession is in general complicated.

With a unique equilibrium for any subgame with given demands and posterior probabilities of irrationality, one can turn to the demand choice game, allowing rational agents to imitate many different commitment types. This choice must involve agents mixing between different demands so that the resultant conditional probabilities of irrationality, and ensuing war of attrition games guarantee equal payoffs from each imitated demand. The structure of this demand choice game is identical to that presented in AG, with payoffs from choices given by the equations in 4. This allows us to immediately state the following proposition with the proof provided by AG’s Proposition 2.

**Proposition 2.** In any bargaining game an equilibrium exists. Furthermore, all equilibria yield the same distribution over outcomes.
4 Differences from the infinite horizon model

This section explores some of the differences from the infinite horizon reputational model of AG, and in particular investigates the possibility of deadline effects. It can be safely skipped by those interested only in the model’s limiting results as probability of commitment becomes small, which are presented in Section 5.

4.1 Non-monotonic concession

In AG’s model, making a more aggressive demand strictly decreases the likelihood that an opponent will concede at time zero, when holding the opponent’s demand and the initial probability of irrationality fixed. This makes intuitive sense because a larger demand by agent i reduces the ratio of i’s concession rate compared j’s, at every point in time, causing j’s reputation to grow relatively more quickly; this is shown in equation 13. Nonetheless, in the deadline world the comparative static need not hold. The reason is that larger demands slow down the concession rates of both agents. And this in turn may lead to more of the war of attrition being fought in an environment where effective discount rates \((r^i + \frac{g(t)}{1-G(t)})\) are relatively more favorable to the agent who increased her demand.

\[
\frac{f^i(t)}{1-F^i(t)} \frac{f^j(t)}{1-F^j(t)} = \frac{(1 - \alpha^i(t))(r^j + \frac{g(t)}{1-G(t)}) + \alpha'^i(t)}{(1 - \alpha^j(t))(r^i + \frac{g(t)}{1-G(t)}) + \alpha'^j(t)}
\]

To illustrate this possibility through an example consider the world with only simple commitment types where the deadline is distributed uniformly on the interval \([1, 2]\) while \(r^1 = \frac{1}{2}, r^2 = 1,\) and \(z^i = e^{-3} \approx 5\%\).\(^6\) The concession rates required to keep rational agents indifferent on the \([0, 2]\) interval are then given by:

\[
\frac{f^i(t)}{1-F^i(t)} = \frac{(1 - \alpha^i(t))(r^j + \frac{1}{2}) + \alpha'^i(t)}{\alpha^j + \alpha^i - 1}
\]

Initially consider \(\alpha^1 = 0.7\) and \(\alpha^2 = 0.4\). These demands imply that concession rates are equal (to 3) on the interval \((0, 1)\) and so exhaustion times are \(T^1 = T^2 = 1,\) meaning there is no initial mass concession by either agent.

By contrast, consider what happens if \(\alpha^2 = 0.5,\) while retaining \(z^i = e^{-3}.\) As a consequence of this more aggressive demand agent 1 concedes faster than 2 for \(t < 1,\) although both agents concede slower, and so neither reaches her exhaustion time by \(t = 1.\) For \(t \geq 1,\) however, agent 2 concedes faster than 1 (indeed increasingly so), with the result that 2’s reputation overtakes 1’s. Exhaustion times are given by \(T^1 \approx 1.44 > T^2 \approx 1.40,\) meaning agent 1 must make initial mass concession 15% of the time to meet equilibrium requirements; this strictly increases agent 2’s expected payoff.

\(^6\)This is a slight change to the model presented above where \(g\) was continuous, but it is an unimportant difference in this case; there clearly exist continuous densities arbitrarily close to this step density.
4.2 Extending bargaining opportunities

Intuitively one might expect that a more patient agent would prefer that any deadline occur later rather than earlier into the future, so that time zero concession is primarily determined by discount rates rather than the possibility of deadline expiration. With arbitrary sophisticated types is hard to know how to make this comparison, however, for simple types the comparative static does hold.

Consider delaying the arrival of the deadline by $\varepsilon$ time units in the subgame following initial demand announcements. One way to do this is to extend the model’s time horizon backwards so that instead of starting at time zero, it starts at time $-\varepsilon$. Bargaining then occurs on the interval $[-\varepsilon, T]$, although the deadline distribution on $[0, T]$ is unaffected, in particular $G(0) = 0$. For given simple demands and probabilities of irrationality this results in an equilibrium in exhaustion times that depend on $\varepsilon$ as follows. Remembering that $T_j$ is defined by the equation $\hat{F}_j(T_j) = 1 - \bar{z}_j$ one gets:

$$\frac{dT_j}{d\varepsilon} = -\frac{\partial \hat{F}_j(T_j)}{\partial \varepsilon} = -\frac{r_i^j}{r_i^j + \frac{g(T_j)}{1-G(T_j)}}$$

Assuming that $T_j = T^*$ one can then assess the effect on initial mass concession given that $\ln(c^i) = \ln(\bar{z}^i) - \ln(1 - \hat{F}_i(T^*))$:

$$\frac{d\ln(c^i)}{d\varepsilon} = -\frac{\partial \ln(1 - \hat{F}_i(T^*))}{\partial \varepsilon} - \frac{\partial \ln(1 - \hat{F}_i(T^*))}{\partial T^*} \frac{dT^*}{d\varepsilon} = \frac{(r^i - r^j)g(T^*)}{(r^i(1-G(T^*) + g(T^*))(\alpha^i + \alpha^j - 1)}$$

If $r^i > r^j$ this is negative and agent i must increases her initial mass concession as $\varepsilon$ increases, whereas if the reverse inequality holds she reduces initial concession. This implies that the patient agent unambiguously benefits from this shift in the deadline distribution, while the impatient agent suffers, for all demands.

The logic behind this result is that shifting the deadline backwards ensures that more of the war of attrition is fought on grounds favorable to the patient agent, when discount rates but not deadline expiration risk determine concession rates. The implication is that one should expect patient agents to seek technologies that will allow bargaining possibilities to be kept open as long as possible, while impatient agents will look for measures that shut those down more quickly.

4.3 Deadline effects

This aim of this section is show how if discounting is relatively small, and types are simple, then deadline effects must necessarily occur. That is, agreements follow a U shape: agents will either agree immediately (at time zero) sign an 11th hour deal or fail to agree entirely. Furthermore, some of the agents who miss the deadline will be rational, rather than behavioral.

In addition to low discount rates, I assume that the deadline distribution is tightly compressed within a narrow time frame. Both assumptions are reasonable in many bargaining situations. For instance during the debt ceiling negotiations, the parties impatience for a deal between May and August was most likely negligible compared to their valuation of the the terms of any
deal, and although any agreement reached before the end of July carried little risk of failure, without an agreement before August 15, default was almost inevitable. Although an intuitive understanding of deadline effects would seem to presume a tight deadline distribution, focused around time $T$, even with a more diffuse distribution, agreements will occur arbitrarily close to the unknown deadline, with rational agents failing to agree.

Agreement at time zero will frequently not be observable, an outside observer may believe the parties never in fact had conflicting interests at all. This means the agreement pattern described above may make it appear that whenever parties have conflicting positions, bargaining in the face of a deadline always results in either a last minute deal, or total failure. If conflict ever emerges, it will go right down to the wire.

I first consider the model with a single simple behavioral type for each agent, and then extend the results allowing for multiple simple types. I define $M(t)$ to be the total probability that agents have agreed to a deal (weakly) before time $t$, and therefore in particular $1 - M(T)$ is the total probability that agents fail to reach agreement.

**Lemma 3.** Consider a sequence of bargaining games $E_n = \{G_n, T, r^i_n, C^i, \pi^i, z^i| i=1,2\}$ with a single simple type for each agent, $C^i = \{r^i\}$ for $i = 1,2$, such that $G_n \overset{\delta_T}{\rightarrow} (r^1 \rightarrow 0$, and assume, without loss of generality, that $\frac{\ln(z^1)}{1-\pi^1} \leq \frac{\ln(z^2)}{1-\pi^2}$. Fix any $\varepsilon > 0$ then:

\[\lim M(0) = 1 - z^1(\frac{\varepsilon^1}{K^1}) < 1 - z^2\]

\[\lim (M(T - \varepsilon) - M(0)) = 0\]

\[\lim (M(T) - M(T - \varepsilon)) = \frac{K^1 + K^2}{K^1 + K^2 + 1} z^1 \left( (z^2)^{\frac{\varepsilon^1}{K^2}} - (z^2)^{\frac{\varepsilon^2}{K^2}} \right) > 0\]

\[\lim (1 - M(T)) = \frac{z^1}{K^1 + K^2 + 1} \left( (z^2)^{\frac{\varepsilon^1}{K^2}} + (z^2)^{\frac{\varepsilon^2}{K^2}} (K^1 + K^2) \right) > z^1\]

Where $K^i$ is defined in equation 8, and $G_n \overset{\delta_T}{\rightarrow}$ represents weak convergence of the deadline distribution to a degenerate hard deadline at time $T$. If instead $\frac{\ln(z^1)}{1-\pi^1} \geq \frac{\ln(z^2)}{1-\pi^2}$ then the above formulas must be relabelled, with agent 1 variables, switched with those of agent 2.

The proof of this result is very simple, if algebraic. It is based on the fact that $(1 - F^i(t))$, the total probability that agent $i$ concedes by time $t$ conditional on not conceding at time zero, is arbitrarily close to $(1 - G(t))^{K^i}$ for small $r^i$. Furthermore, $(1 - G(T - \varepsilon))$ is arbitrarily close to zero when the deadline distribution is concentrated around time $T$. And so, the total probability that an agent will concede by time $T - \varepsilon$, conditional on not conceding at time zero, must be approximately zero.

More intuitively, when $r^i$ is small and there is little risk of missing the deadline at any time before $T - \varepsilon$, very little concession is required to keep agents indifferent between conceding at one moment or the next. This means most agreements must take place either at time zero, or close to $T$, if they occur at all. Close to $T$ the risk of missing the deadline does still become large, however, prompting 11th hour deals.

The result above takes the commitment demands agents imitate as given. The next one shows a similar equilibrium structure when agents can choose among many different demands. The demand choice problem with multiple commitment types is complicated, with outcomes de-
pending on the relative distribution of irrational types in the population. This means that one difference between the results is that limiting probability that rational agents will reach an 11th hour deal, or disagree, cannot be given an exact analytical solution. The probability of both of these events must necessarily be bounded above zero, however. This is a fairly weak bound, given the generality of the distribution of commitment types permitted, however, these values can be made large in particular examples when the prior probability of aggressive commitment types is relatively large.

**Proposition 3.** Consider a sequence of bargaining games \( E_n = \{G_n, T, r'_n, C^i, \pi^i, z^i\}_{i=1,2} \) with simple types such that \( G_n \to \delta_T \) and \( r^i \to 0 \). Fix any \( \varepsilon > 0 \):

\[
\lim M(0) < 1 - \sum_{\alpha^1} \varepsilon^1 \pi(\alpha^1) \left( \sum_{\alpha^2 > 1-\varepsilon^1} \varepsilon^2 \pi(\alpha^2) \right)
\]

\[
\lim (M(T - \varepsilon) - M(0)) = 0
\]

\[
\lim (M(T) - M(T - \varepsilon)) > 0
\]

\[
\lim (1 - M(T)) > \sum_{\alpha^1} \varepsilon^1 \pi(\alpha^1) \left( \sum_{\alpha^2 > 1-\varepsilon^1} \varepsilon^2 \pi(\alpha^2) \right)
\]

Where \( G_n \to \delta_T \) represents weak convergence of the deadline distribution to a degenerate hard deadline at time \( T \).

The proof of Proposition 3 is based on Lemma 3. For each subgame in which rational agents make incompatible demands, \( \alpha^1 + \alpha^2 > 1 \), with probability bounded away from zero, Lemma 3 implies that these agents cannot reach agreement on the interval \((0, T - \varepsilon]\), although some make 11th hour deals and some fail to agree.

All that is required then, is to show that rational agents do imitate some incompatible demands with probability bounded away from zero in the limit. In particular, the proof shows that agents imitate the demands \( \max C^1 + \max C^2 > 1 \) with positive limit probability. To see this for agent 1, notice that if she did not then she would have a posterior probability of irrationality, \( \bar{z}^1 \), arbitrarily close to 1, and so any rational agent 2 would necessarily concede to her immediately in any war of attrition, giving 1 her best possible payoff (when deviating to imitate \( \max C^1 \) with probability 1). Similar reasoning applies to agent 2.

The above findings can be (loosely) compared to a bargaining experiment detailed in Roth et al. [1988] in which there was a substantial deadline effect. The effect occurred despite the experiment being initially designed to study Nash bargaining. The experimental game allowed subjects to reach agreement on the division of experimental lottery tickets. Tickets translated into a probability of winning a prize (rather than nothing) in a binary lottery, meaning that it should be possible to interpret tickets as agents’ utilities, no matter how valuable the prize. The value of the prize differed between the two experimental subjects, with prize pairs varying from \((\$10, \$15)\) to \((\$4, \$36)\) in different treatments. This setup creates a tension between agreements giving agents equal utility (equal tickets) and those giving agents equal expected monetary value (the low value agent must receive more tickets).

During bargaining unstructured messages could be sent between the subjects via computer. Proposals could also be sent, detailing a division of tickets. Bargaining ended after a deadline
of 9 minutes, or when identical proposals were sent by both subjects. Prior to the deadline there was no discounting of payoffs, \( r^i = 0 \), as these would be realized at the end of the experiment in all cases. The authors’ note that the limited processing power of the computers meant the last time at which an agent could agree to another’s offer was uncertain. Indeed the central server recorded some agreements up to 4 seconds after the official deadline (such agreements were honored and recorded at the deadline).

This setup in fact suggests a hard deadline with a stochastic delay in implementation, a model which is highly related model to the current one, and is outlined in the Appendix, section 7.1. With a hard deadline at time \( T \), if the density of deals implemented at time \( t + w \) following an agreement at time \( t \) is given by \( g(T - w) \), one obtains the probability \( 1 - G(t) \) that a deal agreed at time \( t \) it will be implemented before time \( T \). These two models are formally equivalent to each other when \( r^i = r^j = 0 \), and so the above results remain applicable. Under this interpretation, the assumption in Lemma 3 and Proposition 3 that \( G_n \leq \delta_T \) means that the probability of any strictly positive delay in implementing a deal becomes arbitrarily small.

The most notable findings of the experiment (as well as other experiments with a similar game structure reviewed in the paper) were agreements shifted towards an equal monetary split rather than equal utilities, a marked lateness in the time of agreements and a high frequency of disagreement. Despite having nine minutes, almost half of all agreements occurred in the last 30 seconds, with half of those occurring in the last 5 seconds. In fact half of all agreements in the last five seconds were made in the final second! Disagreement occurred 22% of the time.

The marked delay in agreement and the high frequency of disagreement certainly seems to be in line with a reputational bargaining model. One major difference is that there was that there was no obvious probabilistic concession at time zero. This might, however, be explained by the fact that there was a positive time involved in agents declaring their initial bargaining positions, and also the fact that prior to the last 30 seconds there was almost certainly no risk of disagreement created by computer delays, violating the assumption \( r^i + g(t) > 0 \). This suggests that any concession prior to the last 30 seconds may be regarded as initial.

Given the setup of the experiment, it would seem reasonable to expect higher value subjects to demand weakly less than half the tickets (\( \alpha^H \leq 0.5 \)) with low value subject demanding more (\( \alpha^L > 0.5 \)). In this case the model predicts high value subjects should concede at a faster rate than the low value subjects during the war of attrition (regardless of which agent concedes at time zero), and increasingly so as the deadline approaches. Nonetheless, the experimenters record that agreement times do not significantly correlate with the lower prize subject’s outcome.\(^7\) Nor do they report a difference between the terms of agreements prior to the last 30 seconds (potentially initial concession) and after that time. Although this seems to be evidence against the model, perhaps its most important implication is the confirmation of mixing among concession times, which is a key prediction of the war of attrition.

\(^7\)In one treatment (\$20,\$5) the low value prize subject’s outcome is significantly positively correlated, but this does not hold for the more extreme prizes (\$36,\$4) or less extreme prizes (\$14,\$6) and (\$15,\$10).
5 Limit analysis

This section presents the main theoretical results of the paper. These results concern the model’s predictions when the probability of irrationality is small and the type space is sufficiently rich.

I first restrict attention to simple commitment types who make stationary demands. I then consider the complete information (no commitment types) alternating offers bargaining problem in the face of a deadline, and characterize the solution as the time between offers becomes small. This provides necessary groundwork for considering sophisticated commitment types in the reputational model, which I consider next. After providing some simple comparisons of simple and sophisticated type limit results, I finally extend the model to allow for discontinuous demand reputational types, to test the robustness of the result for sophisticated types.

5.1 Simple types

The main result with simple types is that when the prior probability of commitment is small both agents can obtain half the surplus, regardless of their relative discount rates and regardless of the relative likelihood of particular types. This is proved in the following proposition.

Proposition 4. Let $E_n = \{G, T, r^i, C^i, \pi^i, z^i_n\}_{i=1,2}$ be a sequence of bargaining games with simple types in which $z^i_n \to 0$ such that $L > z^i_n > 1/L$ for some positive constant $L$, and let $V^i_n$ be agent $i$’s payoff. Suppose there exists an type $\alpha^i \in C^i$ such that $|\alpha^i - \frac{1}{2}| \leq \epsilon$, then:

$$\lim V^i_n \geq \frac{1}{2} - \epsilon$$

Proof. Given a sequence of bargaining games with $z^i_n \to 0$, and initial demand choices from a finite set, the probability of making a particular choice must converge to some limit (taking a subsequence if necessary).

Consider agent 2. If agent 1 makes a demand $\alpha^1 \leq \frac{1}{2} + \epsilon$ with positive probability in this limit, then agent 2 can obtain within $\epsilon$ of a half by accepting this demand. If agent 1 makes a demand $\alpha^1 > \frac{1}{2} + \epsilon$ with positive limit probability, then for any counterdemand made by agent 2 there is a bound on the posterior probabilities of irrationality for the two agents. That is, for some positive constant $L_2$ one has $L_2 > \bar{z}^1 \bar{z}^2$ sufficiently close to the limit. In particular, consider what happens if agent 2 always mimics a type such that $|\alpha^2 - \frac{1}{2}| \leq \epsilon$.

The simplified expression for $\tilde{F}(t)$ for stationary types is defined in equation 7, with $K^i$ defined in equation 8. These reveal that $\tilde{F}(T^*) = 1$. Given that $\bar{z}^i \to 0$ as $\bar{z}^i \to 0$ and $\bar{z}^i = 1 - \tilde{F}(T^*)$, this implies that $T^* \to T$.

In equilibrium mass concession is defined by equation 11, which means that:

$$\frac{c^1}{c^2} = \frac{\bar{z}^1}{\bar{z}^2} 1 - \tilde{F}(T^*) = \frac{\bar{z}^1}{\bar{z}^2} e^{\epsilon^2 K^1 T^*} (1 - G(T^*)) \frac{\alpha^1 - \alpha^2}{\alpha^2 - \alpha^1} \leq L_2 e^{\epsilon^2 K^1 T^*} (1 - G(T^*)) \frac{\alpha^1 - \alpha^2}{\alpha^2 - \alpha^1}$$

The second of the above equalities comes from plugging in for $\tilde{F}(t)$ using equation 7. The inequality then comes from the bound on $\bar{z}^i$ and the fact that $T^* < T < \infty$. Given $\alpha^1 > \alpha^2$, some simple comparisons of simple and sophisticated type limit results, I finally extend the model to allow for discontinuous demand reputational types, to test the robustness of the result for sophisticated types.
the right hand side of the above equation converges to 0 as $T^* \to T$, and therefore so must the left hand side, entailing not only that agent 2 wins the war of attrition but that she does so with probability approaching 1. This establishes a lower bound for agent 2’s limit payoff.

By demanding arbitrarily close to $\frac{1}{2}$ agent 1 can guarantee that either she faces a counterdemand arbitrarily close to $\frac{1}{2}$, or she wins the ensuing war of attrition with probability approaching 1, with this proved in identical fashion. □

That these limiting payoffs do not reflect discount rates at all is somewhat surprising. I previously showed in section 4.2 that with simple types, patient agents always preferred to extend the bargaining deadline further into the future. Proposition 4, however, shows that any deadline extension effect is entirely second order when the probability of commitment is small.

It is certainly not the case that an agent who makes a more aggressive demand ($\alpha^i > \alpha^j$) always concedes at a slower rate than her opponent irrespective of discount rates. Nonetheless, the aggressive agent does concede at a slower rate near the end of the deadline distribution, if the war of attrition is still going on. This is because agents are equally impatient for a deal at $t$ close to $T$: their ‘effective’ discount rates $\left(r' + \frac{G(t)}{1-G(t)}\right)$ are almost entirely determined by the hazard rate of the deadline distribution, which is exploding. At such a point if agent $j$ is even slightly more generous than agent $i$, $(1 - \alpha^j) > (1 - \alpha^i)$, she must be conceding faster (by an order of magnitude), helping her reach her exhaustion time first.

This argument is predicated on the war of attrition continuing until almost time $T$. But when the probability of irrationality is small this must be the case, because agents’ reputations are primarily built close to $T$. The reason for this is that for $r' < T$ concession rates are bounded because the hazard rate of the deadline distribution is bounded. This means $i$’s total probability of concession by $t'$, conditional on no concession at time zero, must be bounded away from 1. In turn, this implies reputations for commitment at $t'$, absent time zero concession, given by $\frac{z_j}{1-F_j(t')}$, are small whenever the initial probability of commitment, $\bar{z}_j$, is small. Exhaustion times, defined as the point at which reputations must reach probability 1, must therefore be arbitrarily close to $T$.

Intuitively, at the 11th hour agents have equal patience for a deal even when they have different discount rates and so can reasonably demand a deal on equal terms. The advantage of an agent demanding half the surplus at the end of bargaining, means there will be strong incentives for her to wait until the 11th hour unless her opponent is much more likely to be irrational. And so in equilibrium her opponent must indeed be much more likely to be irrational, which requires a high probability of initial mass concession.

5.2 The complete information alternating offers game

In point of comparison to the result with simple types, and to help set up the next one with sophisticated types, I examine a complete information alternating offers bargaining game for the stochastic deadline environment. That is, I temporarily consider a setting with no behavioral types, and a bargaining structure that resembles Rubinstein [1982]. Clearly this solution should depend to some degree on the relative discount rates of the two agents.

In the complete information alternating offers game each agent makes $N$ offers, with $\Delta$ time
ensuring that the horizon demand equation are approximately equal to 0, and so the solution is approximately equal to the infinite line distribution is centered around time $T$.

Some intuition for the solution can be found from the following exercises. If most of the deadline distribution is non-monotonic depending on the nature of the deadline distribution.

This game can be solved by backward induction. Let $t_k$ be the time of agent i’s $k^{th}$ offer. Agent 2’s final offer must demand the entire surplus, at which point her present discounted expected utility if the deadline has not yet passed is $v^2(t_k) = 1$. Other values are then defined recursively by the equations:

$$v^1(t_k) = 1 - e^{-r^2 \Delta} \frac{1 - G(t_k)}{1 - G(t_k^1)} v^2(t_k$$

$$v^2(t_k) = 1 - e^{-r^1 \Delta} \frac{1 - G(t_k^1)}{1 - G(t_k^2)} v^1(t_k^1)$$

Proceeding heuristically, by plugging in $v^2(t + \Delta)$ into the equation for $v^1(t)$, the following equality is obtained:

$$v^1(t) - e^{-r^2 \Delta} r^2 \Delta \frac{1 - G(t + 2\Delta)}{1 - G(t)} v^1(t + 2\Delta) = 1 - e^{-r^2 \Delta} \frac{1 - G(t + 2\Delta)}{1 - G(t)}$$

Dividing each side of this equation by $\Delta$ and taking the limit as offers are made frequently using l’Hôpital’s rule, defines a linear ODE for $t \in [0, T]$:

$$2v^1(t) = \left[ r^2 + r^1 + 2 \frac{g(t)}{1 - G(t)} \right] v^1(t) - \left[ r^2 + \frac{g(t)}{1 - G(t)} \right]$$

A similar expression can be obtained for agent 2. Using the boundary condition $\lim_{t \to T} v^1(t) \in [0, 1]$ gives:

$$v^1(t) = \frac{r^2}{r^2 + r^1} \left[ 1 - e^{b} \int_{t}^{T} g(s) e^{-bs} ds \right] + \frac{1}{2} \left[ e^{b} \int_{t}^{T} g(s) e^{-bs} ds \right]$$

Where:

$$b = \frac{r^2 + r^1}{2}$$

This is a convex combination of the stationary infinite horizon Rubinstein bargaining solution $\frac{r^2}{r^2 + r^1}$, and $\frac{1}{2}$. For any $T < \infty$ with $r^i \neq r^j$ this solution is non-stationary, and may even be non-monotonic depending on the nature of the deadline distribution.

Some intuition for the solution can be found from the following exercises. If most of the deadline distribution is centered around time $T$, and time $T$ is distant then the integrals in the above equation are approximately equal to 0, and so the solution is approximately equal to the infinite horizon demand $\frac{r^2}{r^2 + r^1}$. For $t$ close to $T$, however, the integrals are approximately $e^{-bt} (1 - G(t))$ ensuring $v^1(t)$ approaches $\frac{1}{2}$ as $t \to T$, which makes sense because agents should have approximately the same impatience for a deal close to $T$. Finally, extending the model slightly to allow for a constant deadline hazard $G(t) = 1 - e^{-t}$ with $T = \infty$, one recovers a constant solution
\[ \frac{r^2 + \lambda}{(r^2 + \lambda) + (r^4 + \lambda)}. \]

It may seem surprising that offers converge to \( \frac{1}{2} \) at time \( T \) when agent 2’s final demand is always 1. Insight can be gained by considering agent 1’s final demand. The conditional probability that the deadline hits before 2’s final demand is approximately \( \frac{1}{2} \), which means \( v^1(t^1_N) = 1 - e^{-r^2 \Delta (1 - G(T - \Delta T))} \) converges to \( \frac{1}{2} \). Further, consider agent 1’s \( k \)th to final demand, equation \( 22 \) shows that this converges to \( v^1(t^1_{N-k}) = \frac{k+1}{k} v^1(t^1_{N-k+1}) + \frac{1}{2k} \) which with recursive substitution gives exactly \( \frac{1}{2} \) for all fixed \( k \). Agent 2’s \( k \)th to final demand meanwhile converges to \( v^1(t^1_{N-k}) = \frac{k+1}{2k+1} \), which for large \( k \) is arbitrarily close to \( \frac{1}{2} \).

Notice also that with this Rubinstein demand one obtains the earlier observation made in reputational bargaining with simple commitment types: a relatively patient agent would always prefer that the deadline distribution be pushed further into the future.\(^8\)

### 5.3 Sophisticated types

Let \( \alpha^{1*}(t) = v^1(t) \) be defined by equation \( 24 \) above and \( \alpha^{2*}(t) = 1 - v^1(t) \). I henceforth define these as the Rubinstein bargaining positions. Returning to the reputational bargaining model, the paper’s second main result shows that if there is a sophisticated type who is committed to the Rubinstein bargaining position, then rational agents can guarantee the associated time zero payoff, when the probability of irrationality is small.

**Proposition 5.** Let \( E_n = (G,T,r', C', \pi', \zeta_n')_{i=1,2} \) be a sequence of bargaining games in which \( \zeta_n' \to 0 \) such that \( L > \frac{\zeta_n'}{2} > 1/L \) for some positive constant \( L \), and let \( V_n^i \) be agent \( i \)’s equilibrium payoff. Suppose there exists a Rubinstein bargaining type \( \alpha^{1*} \in C' \) then:

\[ \lim V_n^i \geq \alpha^{1*}(0) \]

**Proof.** Again considering a subsequence if necessary, so that limit probabilities are well defined, consider the problem for agent 2 (the proof for agent 1 is much the same). If agent 1 offers more than \( \alpha^{2*}(0) \), the bound is obtained by acceptance. If on the other hand \( \alpha^{2*}(0) + \alpha^{1}(0) > 1 \) with positive probability in the limit, then for any counterdemand there is a bound on the posterior probabilities of irrationality for the two agents, \( L_2 > \frac{\zeta_n'}{2} \) sufficiently close to the limit. In particular consider what happens if agent 2 always mimics type \( \alpha^{2*} \).

As previously, let \( T' \) be the minimum time such that \( \alpha^{1}(t) + \alpha^{2}(t) = 1 \), or T, if that minimum is not well defined. As defined in equation \( 6 \), \( \bar{F}(t) < 1 \) for all \( t < T' \), but \( \bar{F}(T') = 1 \). Given that \( \bar{\zeta}' \to 0 \) as \( \bar{\zeta} \to 0 \) and \( \bar{\zeta}' = 1 - \bar{F}(T) \), one must have \( T' \to T' \).

Let the difference between an agent’s demand and her opponent’s offer be \( l(t) = \alpha^{1}(t) + \alpha^{2}(t) - 1 \), and so \( (1 - \alpha^{1}(t)) = \alpha^{2}(t) - l(t) \) and \( \alpha^{1}(t) = l(t) - \alpha^{2}(t) \) then the equation \( 11 \) defining initial

---

\(^8\)One way to see this is to assume \( r^2 > r^1 \) and notice that \( 2v^{1}(-\varepsilon) = (r^2 + r^1)v^{1}(t) - r^2 < 0 \) given that \( v^1(t) < \frac{r^2}{r^2 + r^1} \)

\(^9\)In fact this result can easily be made more general. If there is a type \( \alpha^{\infty} \), such \( \alpha^{\infty} \leq \alpha^{1*} \) and \( \alpha^{\infty} \geq \alpha^{2*} \) then \( \lim V_n^i \geq \alpha^{\infty}(0) \) however, for simplicity of argument I present the exact Rubinstein case.
mass concession imples:

\[
\frac{c^1 \cdot 1 - \hat{F}^2(T^*)}{c^2 \cdot \frac{z^1}{z^2}} = 1 - \hat{F}^1(T^*)
\]

\[
\frac{c^1 \cdot 1 - \hat{F}^2(T^*)}{c^2 \cdot \frac{z^1}{z^2}} = \exp\left(-\int_0^{T^*} \frac{(1 - \alpha^2(s))(r^j + \frac{g(s)}{1 - G(s)}) + \alpha^2(s) - (1 - \alpha^1(t))(r^j + \frac{g(s)}{1 - G(s)}) - \alpha^1(s)ds}{\alpha^1(s) + \alpha^2(s) - 1} ds\right)
\]

\[
\frac{c^1 \cdot 1 - \hat{F}^2(T^*)}{c^2 \cdot \frac{z^1}{z^2}} = \exp\left(-\int_0^{T^*} 2\alpha^2(s) - \alpha^2(s)(r^j + r^j + \frac{g(s)}{1 - G(s)}) + r^j + \frac{g(s)}{1 - G(s)} - \frac{l(t)}{l(s)} ds\right)
\]

\[
\frac{c^1 \cdot 1 - \hat{F}^2(T^*)}{c^2 \cdot \frac{z^1}{z^2}} = \exp\left(-\int_0^{T^*} r^j + \frac{g(s)}{1 - G(s)} - \frac{l(t)}{l(s)} ds\right)
\]

\[
\leq \frac{L_2 e^{-r T^*} (1 - G(T^*)) l(T^*)}{l(t)}
\]

The second line merely uses the expressions for \(\hat{F}(t)\) defined in equation 6. The third line follows from substituting variables involving \(\alpha^1(s)\) for counterparts involving \(l(s)\) and rearranging terms. The fourth line follows from the definition of the ODE which determines the Rubinstein bargaining position, equation 23. The fifth follows from integration and the bound on \(\frac{z^1}{z^2}\). As \(T^* \to T\), either \(l(T^*) \to 0\) or \(1 - G(T^*) \to 0\) given how \(T\) is defined. This means the fifth line converges to 0 as the probability of commitment becomes small. But this entails \(c^1 \to 0\), meaning agent 2 wins the war of attrition, and receives initial mass concession with probability approaching 1.

The above proof appears almost too slick. To understand the reason that the Rubinstein demand is able to achieve this payoff bound one must consider concession rates. First impose equality of concession rates between agent i and j at all points in time:

\[
(1 - \alpha^j(t))(r^j + \frac{g(t)}{1 - G(t)}) + \alpha^j(t) = (1 - \alpha^i(t))(r^j + \frac{g(t)}{1 - G(t)}) + \alpha^i(t)
\]

Then impose compatibility of these demands, \(\alpha^i(t) = (1 - \alpha^j(t))\):

\[
\alpha^i(t)(r^j + \frac{g(t)}{1 - G(t)}) - \alpha^j(t) = (1 - \alpha^i(t))(r^j + \frac{g(t)}{1 - G(t)}) + \alpha^i(t)
\]

\[
2\alpha^i(t) = \alpha^i(t)(r^j + r^j + \frac{g(t)}{1 - G(t)}) - (r^j + \frac{g(t)}{1 - G(t)})
\]

The second line of equation 25 is merely a rearrangement of the first; it is also, however, the ODE that defines the Rubinstein bargaining position.

Given this construction, it becomes clear that in order to concede at a faster rate than the Rubinstein demand at time \(t\) (when facing an opponent’s Rubinstein demand) an agent must either be more generous than her Rubinstein demand, or increase her demand faster than her Rubinstein demand would. The former case would imply the two demands are compatible, while the latter must result in an even more aggressive demand in the future, which will slow down the agent’s future concession rate. The proof of Proposition 5 shows that this short term advantage and
long term disadvantage cannot pay.\textsuperscript{10}

As to why the Rubinstein demand is exactly that which equalizes agent’s concession rates, the answer lies in the alternating offers model (equal offer frequency) providing agents with equal innate bargaining power, while also incorporating a strong backward induction logic. AG first established this symmetry of treatment (regardless of relative offer frequency) in reputational bargaining models, while backward induction has been at the heart of reputational games since their earliest days.

5.4 Comparison of simple and sophisticated results

To illustrate the importance of non-stationarity for the Rubinstein commitment type, and to provide a point of comparison between Propositions 4 and 5, consider an example in which there is only one type for each agent and \( z^1 = z^2 = 5\% \). The deadline distribution is uniform on \([0, 1]\) with \( r^1 = 1 \) and \( r^2 = 3 \). This means the Rubinstein demand reduces to:

\[
\alpha^*(t) = \frac{3}{4} \left( 1 - \frac{1 - e^{-2(1-t)}}{2(1-t)} \right) + \frac{1}{2} \left( 1 - \frac{e^{-2(1-t)}}{2(1-t)} \right)
\]

And so the initial Rubinstein demand is \( \alpha^*(0) \approx 0.64 \). Numerical calculations reveal that if agent 1 imitates this Rubinstein type while agent 2 imitates the simple type \( \alpha^2 = 0.5 \) then agent 2 must make time zero concession with probability 74\%. If however, agent 2’s type remains \( \alpha^2 = 0.5 \), but agent 1’s type is now simple and committed to the initial Rubinstein demand \( \alpha^1 = \alpha^*(0) \), the situation is entirely reversed with agent 1 making time zero concession with probability 73\%.

In the Introduction, I suggested that one may be concerned about the plausibility of sophisticated behavioral types, and at the very least believe that the fixed demands of simple types seem much more likely. To build on this one can consider the implications of separating the distribution of simple and sophisticated types, and allowing the prior probability of sophisticated types to converge to zero faster than the probability of simple types.

In particular, consider \( r^i < r^j \) and simplify to one type for each agent, so that agent \( j \) is committed to \( \alpha^j = \frac{1}{2} \), and agent \( i \) is committed to her Rubinstein demand minus \( \varepsilon \), \( \alpha'(t) = \alpha^*(t) - \varepsilon \), for which \( \varepsilon = 0 \) is a special case. Let \( z^i_n \to 0 \), but now with \( L > \frac{\varepsilon}{(z^i)^p} > \frac{1}{L} \) for some positive constant \( L \), and \( p > 1 \). Notice that this riggs the game against agent \( i \), she must become \textit{infinitely} more

\textsuperscript{10}Nonetheless notice that for larger probabilities of irrationality, if a Rubinstein type faced a commitment type with an equal and large likelihood of rationality, which demanded slightly more than the Rubinstein demand initially and then rapidly increased its demand to 1, then \( \frac{N(T)}{N} \) could be much larger than 1, and so the Rubinstein agent might need to make initial mass concession.
irrational than her opponent in the limit. This implies that time zero concession is governed by:

\[
\frac{c^i}{(c^j)^p} = \frac{z^i}{(c^j)^p} \frac{(1 - \hat{F}(T^*))^p}{1 - \hat{F}(T^*)} \\
= \frac{z^i}{(c^j)^p} \exp \left( \int_{0}^{T^*} \frac{-p(1 - \alpha^i)(r^j + \frac{\mu_0}{1-\alpha_0}) + (1 - \alpha^i)(r^j + \frac{\mu_0}{1-\alpha_0}) + \alpha^i(t)}{\alpha^j + \alpha^i(t) - 1} \, dt \right) \\
= \frac{z^i}{(c^j)^p} \exp \left( \int_{0}^{T^*} \frac{-p\frac{1}{2}(r^j + \frac{\mu_0}{1-\alpha_0}) + (1 - \alpha^i)(r^j + \frac{\mu_0}{1-\alpha_0}) + \alpha^i(t) + \epsilon(\frac{r^i}{2} + \frac{\mu_0}{1-\alpha_0}) - \frac{1}{2}(r^j + \frac{\mu_0}{1-\alpha_0})}{\alpha^j + \alpha^i(t) - 1} \, dt \right)
\]

Where the final line imposes \( \alpha^j = \frac{1}{2} \) and \( \alpha^i(t) = \alpha^{\mu_0}(t) \). As \( z^i \rightarrow 0 \) one must again have \( T^* \rightarrow T' \) where \( T' \leq T \) is defined as the first time such that \( \alpha^i(t) + \alpha^j(t) = 1 \). The numerator of the integrand in the above equation must therefore converge to:

\[
-p \frac{1}{2} \left( r^i + \frac{G(T')}{1 - G(T')} \right) + \frac{1 + 2\epsilon}{2} \left( \frac{r^j + r^i}{2} + \frac{G(T')}{1 - G(T')} \right)
\]

If \( \epsilon = 0 \) then \( T' = T \) and so for \( T^* \) close to \( T \) this is negative and of arbitrarily large magnitude given \( p > 1 \). This ensures that the entire integral will converge to \(-\infty\) and consequently \( c^i \rightarrow 0 \).\(^{11}\) This means the Rubinstein demand agent must concede with probability approaching 1 to the simple demand of \( \alpha^j = 0.5 \), if her probability of irrationality converges to zero even slightly faster.

This is by no means the final word on the matter, however, for instance if \( \epsilon > \frac{p-1}{2} \) the above numerator of the integrand at \( T' < T \) is certainly positive, remembering that \( r^j > r^i \), and this implies \( c^i \rightarrow 0 \). Of course, this in turn does not necessarily mean that agent \( i \) can guarantee \( \alpha^{\mu_0}(0) - \epsilon \) for arbitrary simple types of agent \( j \), as a demand \( \alpha^j < \frac{1}{2} \) would change the equation once again. Finding ‘optimal’ limit demands for simple and stationary types in this setting would seem a very difficult task. I expect that demand optimization with multiple types will lead to a result that depends heavily on the identity of agent 1 and 2, with a significant second mover advantage.

### 5.5 Extension to discontinuous types

The sophisticated types considered thus far have been restricted to continuously differentiable demands. This section provides a robustness check to the Rubinstein bargaining result (Proposition 5) when allowing for types making discontinuous history contingent demands.

To do this, I adapt the continuous time framework above to discrete-continuous time. This is an invention of AP, which in essence allows for continuous time bargaining with discontinuous demands, without the usual openness problems associated with continuous time. In this, there is a sense in which the setup more truthfully approximates the limit of a discrete time game as the time between offers become small. In discrete time there is always a ‘last time’ that an agent

\(^{11}\)To see this choose \( r^i \) close to \( T \) such that the integrand is negative. On \((0, r^i]\) the integral is bounded. On the remaining interval \((r^i, T^*)\) bound the denominator on the integrand from above \( \alpha^i(t) + \alpha^j(t) - 1 < \epsilon \), and similarly bound from above the terms multiplying \( r^j \) and \( r^i \) in the numerator. Observe that the sum of the terms multiplying \( \frac{\mu_0}{1-\alpha_0} \) is bounded above by \( -\frac{p-1}{2} \). Integrating with such bounds imposed and taking the limit as \( T^* \rightarrow T \) then gives the desired result.
could concede to her opponent’s old demand, and a ‘first time’ that she can concede to a new demand.

The discrete part of the discrete-continuous time comes from the existence of a finite set of times $t_1, t_2, ..., t_m$ within the interval $(0, T)$ which have a special structure. The single time $t_k$ is divided into four discrete times $t_{k-2} < t_{k-1} < t_{k+1} < t_{k+2}$. If bargaining is not concluded before time $t_{k-2}$ both agents can concede to their opponent’s existing demand. At $t_{k-1}$ one agent alone, say $i$, has the choice of conceding to $j$’s existing demand, not conceding and reaffirming her bargaining position, or changing her bargaining position to some new $\alpha'$ describing her demand on the interval $[t_{k+2}, t_{k+1}, - 2]$. At $t_{k+1}$ the other agent $j$ similarly has the opportunity to (alone) concede to her opponent’s (possibly new) demand, not concede and reaffirm her position, or change her bargaining position. At $t_{k+2}$ both agents can again concede to their opponent’s (possibly new) demand.

The identity of the agent who can move at $t_{k-1}$ rather than $t_{k+1}$ is arbitrary for each $k$, but is known to agents. There is no discounting across the times $t_{k-2}, ..., t_{k+2}$, their structure is similar to the assumption that agent 1 announces her initial demand marginally before agent 2, even though both events are recorded at time zero. The existence of these discrete times also reintroduces the possibility that an agent will initially imitate a behavioral type, before revealing rationality at $t_{k+1}$ without conceding.

Using this new time structure, one can represent the behavioral type’s bargaining position on the interval $[t_{k+2}, t_{k+1}, - 2]$ as a function $(0, 1)^{|t_{k+2} - t_{k+1}, - 2|}$ with the choice of function possibly depending on the entire history of play prior to the time she changes her demand (at $t_{k+1}$), where I interpret $0 = t_0, t_2$ and $T = t_{m+1}, - 2$. I maintain the assumption that $(1 - \alpha'(t))\left(r^j + \frac{g(t)}{G(t)}\right) > -\alpha'(t)$ on each interval $[t_{k+2}, t_{k+1}, - 2]$. Whichever agent can move at $t_{k-1}$ can concede to her opponent’s $t_{k-2}$ demand; the agent who can move at $t_{k+1}$ can concede to her opponent’s $t_{k+2}$ demand. If a behavioral agent would change her demand at $t_{k+1}$ in such a way that it would be compatible with her opponent’s existing demand, I assume she instead immediately concedes. I also now assume that at time zero each player announces their their entire contingent bargaining function, effectively their type.

To simplify matters suppose that agent R always imitates the Rubinstein demand, but agent NR need not. This will mean Lemma 1 applies to agent NR: she must concede immediately if revealed to be rational when R is possibly irrational. The same need not be true of agent R, however. In particular if R is revealed to be rational, but irrational NR’s demand is expected to discontinuously jump downwards at $t_{k+1}$ (assuming R continues to behave as expected) R may not concede to NR immediately, but instead wait for NR to lower her demand. This implies that equilibrium must involve either both agents reaching probability 1 of irrationality at the same time or agent NR reaching this level strictly before R (with R then waiting for a downwards expected demand jump).

Not only will R not necessarily immediately concede after revealing rationality at $t_{k+1}$, but revealing rationality may induce a lower future demand from her opponent (compared to if she had maintained her Rubinstein position) given history contingent demands by an irrational agent NR. It may also mean that a rational agent NR will optimally reveal rationality without conceding after R has done so rather than continuing to imitate her irrational type, which is set to lower its demand in the future.
Continuation payoffs after both agents reveal rationality without conceding depend on the complete information, discrete-continuous time game. This has multiple equilibria that can depend arbitrarily on the past history of play, creating some indeterminacy in the reputational game.\footnote{This is a clear difference between the discrete continuous time game and considering the limit of a discrete time as the period length becomes small.} Nonetheless, in any given equilibrium of the reputational game if agent R reveals rationality first at $t_{k,+/-1}$ (without conceding) one can treat her continuation payoff as if exogenous to the model $v^R(t_{k,+/-1})$. The expected continuation payoff for a rational agent NR, when R does this, is similarly given by $v^{NR}(t_{k,+/-1})$. Given this protocol, although irrational NR’s demand can depend on R’s behavior I refer to $\alpha^{NR}(t)$ as NR’s demand at $t$ when R has always imitated the Rubinstein type prior to $t$, without confusion.

With discontinuous demands equilibrium may also now result in mass concession at times other than 0. In the simplest case, when irrational agent NR’s demand is expected to jump upwards after, at $t_{k,+/-1}$ and has no other discontinuities one can have mass concession by NR at $t_{k,+2}$ with probability $(1 - c^{NR}_{k,+2}) > 0$ such that:

$$(1 - \alpha^{NR}(t_{k,-2})) = (1 - c^{NR}_{k,+2})\alpha^R(t_k) + c^{NR}_{k,+2}\alpha^{NR}(t_{k,+2})$$

This concession behavior makes agent R indifferent between concession at $t_{k,-2}$ or an instant after, at $t_{k,+2}$.

Equilibrium may also involve gaps of waiting by both parties followed by probabilistic concession at a later date. The simplest variety of this occurs when agent NR’s demand discontinuously drops only once at $t_{k,+/-1}$ such that $\alpha^R(t_k) + \alpha^{NR}(t_{k,+2}) > 1$. In this case equilibrium can involve both agents ceasing to concede on the interval $[t', t_{k,-2}]$ where $t' \geq 0$ is determined by:\footnote{If $(1 - \alpha^{NR}(0)) < e^{-R(t_k)}(1 - G(t_k))(1 - \alpha^{NR}(t_{k,+2}))$ then agents wait on the interval $(0, t_{k,-2}]$} \footnote{I assume $(1 - \alpha^{NR}(t')) > e^{-R(t_k)}\left(\frac{1 - G(t_k)}{1 - G(t')}\right)^R(t_{k,+/-1})$ for all $h < k$.}

$$(1 - \alpha^{NR}(t')) = e^{-R(t_k-t')} \frac{1 - G(t_k)}{1 - G(t')} (1 - \alpha^{NR}(t_{k,+2}))$$

To compensate agent NR for waiting on this interval, agent R then concedes to the demand $\alpha^{NR}(t_{k,+2})$ at $t_{k,+2}$ with probability $(1 - c^{R}_{k,+2})$ so that:

$$(1 - \alpha^R(t')) = e^{-R(t_k-t')} \frac{1 - G(t_k)}{1 - G(t')} (1 - c^{NR}_{k,+2})\alpha^{NR}(t_{k,+2}) + c^{R}_{k,+2}(1 - \alpha^R(t_k))$$

Even stranger things can happen when $(1 - \alpha^{NR}(t_{k,-2})), (1 - \alpha^{NR}(t_{k,+2})) < v^R(t_{k,+/-1})$, in that both agents wait on some interval $[t', t_{k,-2}]$ before agent R reveals rationality without conceding with some probability, at $t_{k,+/-1}$ and if she does not, then NR concedes with positive probability to the demand $\alpha^R(t_k)$ at time $t_{k,+2}$ to ensure R was indifferent between revealing rationality or not at $t_{k,+/-1}$. The upshot is that both agents reveal rationality (first) at time $t_k$ with strictly positive probability.

The nature of these strange occurrences is explored more thoroughly in AP. Here I simply press on to state the robustness result: as the probability of irrationality becomes small agents can still guarantee their Rubinstein bargaining payoff.
Proposition 6. Let \( E_n = \{G, T, r', C', \pi', z'_n\}_{i=1,2} \) be a sequence of bargaining games (with possibly discontinuous types) in which \( z'_n \rightarrow 0 \) such that \( L > \frac{z'_1}{z'_2} > 1/L \) for some positive constant \( L \), and let \( V^i_n \) be some equilibrium payoff for agent \( i \). Suppose there exists a Rubinstein bargaining type \( \alpha^r_i \in C' \) then:

\[
\lim_{n \to \infty} V^i_n \geq \alpha^r_i(0)
\]

The proof involves partly characterizing an equilibrium with discontinuous types, before arguing that the conditional probability of concession by agent NR at the times \( t_1, ..., t_m \), is bounded away from 1 when agent R imitates her Rubinstein type. The total concession probability by NR at this finite set of points is therefore bounded below 1. To reach her exhaustion time, as the probability of commitment becomes (vanishingly) small, however, requires the total probability of concession to approach 1. Any concession on the intervals \((t_k, t_{k+1}-2), (t_k+2, t_{k+1})\) must necessarily be at the concession rates given in equation 5, where it was previously shown that agent R an advantage. Indeed, Proposition 5 showed this advantage was complete as \( z'^R_N \rightarrow 0 \) so long as \( \frac{z'^R_N}{z'^R_R} < K \) for any fixed \( K \). But the concession by NR at times \( t_1, ..., t_m \), being bounded away from 1, can only improve agent NR’s reputation by a bounded amount, and so the logic of Proposition 5 still goes through.

This is considerably simpler than AP’s proof when they allowing for discontinuous history contingent demands. AP allow agents to change demands at integer times, which in an infinite horizon model implies a countably infinite number of potential discontinuities. This necessitates ‘adding up’ concession probabilities by each agent at those points. In a finite horizon, however, this timing protocol leads to only a finite number of demand discontinuities, allowing the uniform bound on concession at those points. Adapting the model to allow an infinite number of discontinuities within the finite horizon would present considerable additional challenges.

6 Conclusion

The results of the current model do not suggest a way out of partisan brinkmanship in Washington DC in the face of fiscal deadlines. Indeed, if politicians impatience for a deal is relatively small, the results suggest some sort of deadline effect is inevitable, with a positive probability of an 11th hour deal, and a risk of failure by rational agents. Clearly the analogy should be treated with great caution, however, as the model is a drastic and imperfect simplification of forces that may be at work in such an environment, with no role for ‘blame’ or elections. Also, the model does not address reputational bargaining in repeated game framework. This suggests an avenue for further work, which must consider how agents might recover reputation after revealing rationality.

In addition to illustrating deadline effects, the model gives more general insights into reputational bargaining, in particular when the likelihood of irrationality is small. The optimality of the Rubinstein demand when behavioral types may change their demands over time is the first characterization in the literature of canonical bargaining types in a non-stationary environment. Indeed, the non-stationarity of the bargaining environment draws out key features of the optimal demand that are partially obscured in a stationary infinite horizon model. The link between reputational bargaining and an alternating offers model arises because of two key features, sym-
metry and backward induction, that are common to both. Building on this paper, Fanning [2013a] and Fanning [2013b] find the Rubinstein demand to be the canonical sophisticated type in other non-stationary bargaining settings, illustrating the robustness of the result.

If one takes the reputational bargaining model seriously, and views the limit results as a route to gain insights with larger probabilities of commitment, however, then the markedly different predictions with simple and sophisticated types also highlight a need to consider the plausibility of different behavioral types. If sophisticated commitment types seem inherently implausible then pretending to commit to a fixed demand of approximately half the surplus would seem good advice for agents when bargaining in the face of a deadline, regardless of their discount rates.

More generally this difference suggests a need to explicitly motivate commitment behavior from preferences instead of imposing it exogenously. This paper has already motivated further work in Fanning [2013c], which shows how (non-stationary) commitment demands can be motivated in a Coasean bargaining model when rational sellers have preferences for equal material payoffs.

7 Appendix

7.1 Delayed implementation model

This section presents the outlines of a model in which there is a known hard deadline at time $T$ but agreements take an unknown time to implement. The model is closely related to the one in the paper and is formally identical when $r^i = r^j = 0$.

I assume that after agreement on the principles of a deal at time $t$ the time it takes to implement a deal is distributed according to $G$ on the interval $[t, T + t]$ with the continuous density. The density for a deal to be implemented at time $t + w$ given agreement at $t$ is $g(T - w)$, where I continue to assume that $g(T) > 0$ although this can be easily generalized to allow the delay from implementation to always be strictly positive. This means that the probability of implementing a deal before time $T$ is:

$$\int_0^{T-t} g(T - w)dw = 1 - G(t)$$

Agent i’s expected utility from a deal agreed at time $t$ that gives her $\alpha^i(t)$ of the surplus is:

$$u^i(t, \alpha^i(t)) = \alpha^i(t) \int_0^{T-t} g(T - w)e^{-r^i(w + t)}dw$$

Under this utility formulation, Lemma 1 will go through without change: a rational agent must concede immediately to an agent who might possibly be committed to her demand. This reduces strategies to the choice of commitment type to mimic and concession times.

Given demands and agent $j$’s strategy, agent $i$’s expected utility from concession at time $t$ is given by:

$$U^i(t, \alpha^j|\alpha) = \int_0^{T} u^i(t, \alpha^i(t))dF^j(s) + (1 - F^j(t))u^i(t, 1 - \alpha^i(t))$$

$$= \int_0^{T} \alpha^i(s) \int_0^{T-s} g(T - w)e^{-r^i(w + s)}dwdF^j(s) + (1 - F^j(t))(1 - \alpha^i(t)) \int_0^{T-t} g(T - w)e^{-r^i(w + t)}dw$$

The arguments of Lemma 2 again go through requiring agents to reach probability 1 of rationality at the same time $T^* < T$ and to be indifferent to conceding on the interval $(0, T^*)$. Differentiating the above equation gives the
The Rubinstein bargaining demand in this setting is defined by:

$$\frac{f^i(t)}{1 - F^i(t)} = \left( r^i + \frac{e^{r^i t} g(t)}{\alpha^i(t) + \alpha^i(t) - 1} \right) \frac{1 - \alpha^i(t) + \alpha^i(t)}{\alpha^i(t) + \alpha^i(t) - 1}$$

This allows us to define $\hat{F}^i(t)$ as before as the concession up to time $t$ assuming no time zero concession.

$$1 - \hat{F}^i(t) = \exp \left( - \int_0^t \frac{dt}{\alpha^i(s) + \alpha^i(s) - 1} \left( 1 - \alpha^i(s) + \alpha^i(s) \right) ds \right)$$

In the particular case of simple types this reduces to:

$$1 - \hat{F}^i(t) = \left( \int_0^{T-} g(T - w) e^{-rT(t)dw} \right)^{K^i}$$

Where $K^i$ is defined as before. With this new $\hat{F}^i$ exhaustion times and time zero concession are defined exactly as before, characterizing the unique equilibrium.

After noting that:

$$\int_0^{T-} g(T - w) e^{-rT(t)dw} \in [e^{-rT(1 - G(t))}, 1 - G(t)]$$

It is clear that the proof of Proposition 4 goes through exactly as before with this new technology. With only simple types, rational agents can guarantee half the surplus as the probability of commitment becomes small.

The Rubinstein bargaining demand in this setting is defined by:

$$2\alpha^r(t) = -\alpha^i(t) \left( \frac{dl}{dt} \left( \int_0^{T-} g(T - w) e^{-rT(t)dw} \right) \right) + \frac{dl}{dt} \left( \int_0^{T-} g(T - w) e^{-rT(t)dw} \right) + \frac{dl}{dt} \left( \int_0^{T-} g(T - w) e^{-rT(t)dw} \right)$$

Using the same trick as in Proposition 5 of defining $l(t) = \alpha^i(t) + \alpha^i(t) - 1$, the proof that each agent can secure her Rubinstein payoff when such a sophisticated type exists then works exactly as before. As $T - \rightarrow T^*$, time zero concession by agent 1 must converge to probability 1 when agent 2 imitates the Rubinstein type:

$$\frac{c^1}{c^2} \geq \frac{\pi^1}{\pi^2} \left( \frac{\int_0^{T-} g(T - w) e^{-rT(t)dw}}{\int_0^{T-} g(T - w) e^{-rT(t)dw}} \right) \leq L_2 e^{2rT(1 - G(T^*))} \frac{l(T^*)}{l(0)}$$

### 7.2 Proofs

#### Proof of Lemma 1

I first prove that there exists a $t' < T$ such that i must have agreed to j’s demand by time $t'$, assuming j has always behaved in a way consistent with the commitment type.

Define $\pi^i = \max_{t \in [0, T]} \alpha^i(t)$; this is well defined by continuity and strictly less than 1 by assumptions made about irrational types demands. Let $\bar{\pi}(t)$ record agent j’s conditional probability of irrationality at time $t$ when she has always previously behaved like the irrational type, and finally let $\pi(t, l)$ be the total probability that agent j does not change her demand from that of the irrational type on the interval $[l_1, l_2]$ conditional on not having done so before $l_1$.

At $l_1 \in [t', T)$ such that j has always behaved like the irrational type, a rational agent can obtain utility of at least $(1 - \bar{\pi})$ conditional on the deadline not having passed. Given the probability that agent i does not change her demand on $[l_1, l_2)$, one also has a bound on the the maximum expected payoff from waiting until $l_2$ (given that the
deadline has not passed by \( \bar{t}_1 \). For it to be optimal for agent i not to concede to agent i prior to \( \bar{t}_2 \) then requires:

\[
1 - \alpha^i \leq 1 - \pi^i(\bar{t}_1, \bar{t}_2) + \pi^i(\bar{t}_1, \bar{t}_2)e^{-r(\bar{t}_2 - \bar{t}_1)} \frac{1 - G(\bar{t}_2)}{1 - G(\bar{t}_1)}
\]

This gives a bound on the size of \( \pi^i(\bar{t}_1, \bar{t}_2) \):

\[
\pi^i(\bar{t}_1, \bar{t}_2) \leq \frac{\alpha^i}{1 - e^{-r(\bar{t}_2 - \bar{t}_1)} \frac{1 - G(\bar{t}_2)}{1 - G(\bar{t}_1)}}
\]

Fixing \( \delta \in (\bar{t}^i, 1) \), for any \( \bar{t}_1 \), it is clear that one can always find a \( \bar{t}_2 < T \) such that \( \pi^i(\bar{t}_1, \bar{t}_2) < \delta \). For such a fixed \( \delta \) there clearly exists a \( K \) such that \( \delta^K < \frac{\epsilon}{2}(t') \). Given \( K \) iterations of the above argument, letting \( t' = t_1 \), one has \( \bar{t}_1, ..., \bar{t}_{K+1} < T \) such that the above equation defines \( \pi^i(\bar{t}_i, \bar{t}_{i+1}) \leq \delta \). This means that the probability that agent i has continued to behave like an irrational type until time \( \bar{t}_{K+1} \) must be less than \( \delta^K < \frac{\epsilon}{2}(t') \). However, given that the probability of behaving like an irrational type until \( \bar{t}_{K+1} \) must be weakly greater than \( \frac{\epsilon}{2}(t') \) this is a contradiction. And so rational agent i cannot possibly wait until time \( \bar{t}_{K+1} \) to concede. This proves the initial claim (that i must concede by some \( t' < T \)).

I shall now prove that the latest time that agent i can concede to i’s demand is in fact \( t' \). Suppose not.

Let, \( \bar{t}_1 \in (t', T) \), be supremum of times such that i has not agreed to the irrational j’s demand. Consider the last \( \epsilon \) units of time prior to \( \bar{t}_1 \) if j has continued to behave like the irrational type, and i has not conceded. Let x be i’s expected payoff at time \( \bar{t}_1 - \epsilon \) (conditional on the deadline not having passed) if j agrees to a deal on terms worse than \( \alpha^i(t) \) at some \( t \in [\bar{t}_1 - \epsilon, \bar{t}_1 - (1 - \beta)\epsilon] \), where \( \beta \in (0, 1) \). Let y be i’s expected payoff (at time \( \bar{t}_1 - \epsilon \)) if j does not agree to such an offer prior to \( \bar{t}_1 - (1 - \beta)\epsilon \). And finally let \( \xi \) be the probability that i assigns to the event that j will not agree to such an offer prior to \( \bar{t}_1 - (1 - \beta)\epsilon \).

If agent i optimally rejects \( 1 - \alpha^i(\bar{t}_1 - \epsilon) \) this implies that:

\[
1 - \alpha^i(\bar{t}_1 - \epsilon) \leq (1 - \xi)x + \xi y
\]

And so:

\[
\xi \leq \frac{x - (1 - \alpha^i(\bar{t}_1 - \epsilon))}{x - y}
\]

whenever \( x - y > 0 \)

At time \( t \in [\bar{t}_1 - \epsilon, \bar{t}_1] \) conditional on the deadline not having passed by \( t \), a rational agent j can choose to imitate the irrational type, knowing that i must concede before time \( \bar{t}_1 \) to obtain a payoff of at least:

\[
\min_{s \in [\bar{t}_1]} \alpha^i(s)e^{-r(s - \bar{t}_1)} \frac{1 - G(s)}{1 - G(\bar{t}_1)}
\]

The continuity of the derivative of the above argument means j’s minimum payoff is either uniformly increasing or uniformly decreasing close to \( \bar{t}_1 \). This means in particular that for sufficiently small \( \epsilon \) that the minimum payoffs at \( \bar{t}_1 - \epsilon \) and \( \bar{t}_1 - (1 - \beta)\epsilon \) are either both achieved at those times or at \( \bar{t}_1 \). Suppose the former. These minimum payoffs for j imply bounds on the payoffs i can expect:

\[
x \leq (1 - \alpha^i(\bar{t}_1 - \epsilon))
\]

\[
y \leq (1 - \alpha^i(\bar{t}_1 - (1 - \beta)\epsilon))e^{-r\beta\epsilon} \frac{1 - G(\bar{t}_1 - (1 - \beta)\epsilon)}{1 - G(\bar{t}_1 - \epsilon)}
\]

Combining these bounds with equation 26 gives:

\[
1 - \alpha^i(\bar{t}_1 - \epsilon) \leq (1 - \alpha^i(\bar{t}_1 - (1 - \beta)\epsilon))e^{-r\beta\epsilon} \frac{1 - G(\bar{t}_1 - (1 - \beta)\epsilon)}{1 - G(\bar{t}_1 - \epsilon)}
\]

However, this in turn creates a contradiction to the assumption of optimal concession against a known irrational agent. Hence one can assume that the minimum expected payoffs for agent j at \( \bar{t}_1 - \epsilon \) and \( \bar{t}_1 - (1 - \beta)\epsilon \) are found
when i does not concede until $\bar{t}_1$. This creates new bounds on agent i’s payoffs:

$$x \leq 1 - \alpha(\bar{t}_i)e^{-r/\beta} \frac{1 - G(\bar{t}_i)}{1 - G(\bar{t}_1 - \epsilon)}$$

$$y \leq e^{-r/\beta} \frac{1 - G(\bar{t}_i) - (1 - \beta)e}{1 - G(\bar{t}_1 - \epsilon)} \left( 1 - e^{-r/\beta} (1 - G(\bar{t}_i)) \alpha(\bar{t}_i) \right)$$

The second of these bounds implies that $y < 1 - \alpha(\bar{t}_1 - \epsilon)$ whenever:

$$e^{-r/\beta} \frac{1 - G(\bar{t}_i) - (1 - \beta)e}{1 - G(\bar{t}_1 - \epsilon)} \left( 1 - e^{-r/\beta} (1 - G(\bar{t}_i)) \alpha(\bar{t}_i) \right) < 1 - \alpha(\bar{t}_1 - \epsilon)$$

Rearranging this gives:

$$\frac{1 - G(\bar{t}_1 - \epsilon)e}{1 - G(\bar{t}_i) - (1 - \beta)e} \left( 1 - e^{-r/\beta} (1 - G(\bar{t}_i)) \alpha(\bar{t}_i) \right) < 1$$

Evaluating the limit of the left hand side using l’Hospital’s rule as $\epsilon \to 0$ gives:

$$\frac{(1 - \beta)g(\bar{t}_i) + r'(1 - \beta)(1 - G(\bar{t}_i))\alpha(\bar{t}_i) - \alpha'(\bar{t}_1)(1 - G(\bar{t}_i)) - g(\bar{t}_i)(1 - \alpha'(\bar{t}_1))}{r(1 - \beta)(1 - G(\bar{t}_i))(1 - \alpha'(\bar{t}_1))}$$

This is strictly less than 1 whenever:

$$\beta \in \left\{ \alpha'(\bar{t}_1) \left( \frac{r}{1 - G(\bar{t}_1)} + \frac{g(\bar{t}_1)}{1 - G(\bar{t}_1)} \right) - \alpha'(\bar{t}_1) \left( \frac{r'}{1 - G(\bar{t}_1)} + \frac{\alpha'(\bar{t}_1) + \alpha'(\bar{t}_1) r'}{1 - \alpha'(\bar{t}_1)} \right), 1 \right\}$$

By the assumption of optimal concession against an irrational agent one has $(1 - \alpha'(\bar{t}_1)) \left( \frac{r'}{1 - G(\bar{t}_1)} + \frac{\alpha'(\bar{t}_1) + \alpha'(\bar{t}_1) r'}{1 - \alpha'(\bar{t}_1)} \right) > -\alpha'(\bar{t}_1)$, which means that the above interval is non-empty. This means for a fixed $\beta$ in the above interval and for sufficiently small $\epsilon$, one has $x \geq (1 - \alpha'(\bar{t}_1)) > y$. Translating this information into equation 27 implies:

$$\xi \leq \frac{(1 - G(\bar{t}_1 - \epsilon)e\alpha'(\bar{t}_1) - e^{-r/\beta} (1 - G(\bar{t}_1)) \alpha'(\bar{t}_1))}{1 - G(\bar{t}_1 - \epsilon) - e^{-r/\beta} (1 - G(\bar{t}_1)) \alpha'(\bar{t}_1) - e^{-r/\beta} [1 - G(\bar{t}_1 - \epsilon)e\alpha'(\bar{t}_1) - e^{-r/\beta} (1 - G(\bar{t}_1)) \alpha'(\bar{t}_1)]]}$$

Again taking the limit of the right hand side as $\epsilon \to 0$ using l’Hospital’s rule gives:

$$\frac{1}{\beta} \frac{\alpha'(\bar{t}_1) \left( \frac{g(\bar{t}_1)}{1 - G(\bar{t}_1)} + r' \right) - \alpha'(\bar{t}_1) \left( \frac{g(\bar{t}_1)}{1 - G(\bar{t}_1)} + \alpha'(\bar{t}_1) r' + (1 - \alpha'(\bar{t}_1)) r' \right)}{\alpha'(\bar{t}_1) \left( \frac{r}{1 - G(\bar{t}_1)} + \frac{g(\bar{t}_1)}{1 - G(\bar{t}_1)} \right) - \alpha'(\bar{t}_1) \left( \frac{r'}{1 - G(\bar{t}_1)} + \frac{\alpha'(\bar{t}_1) + \alpha'(\bar{t}_1) r'}{1 - \alpha'(\bar{t}_1)} \right)}$$

And so for any fixed $\beta$ in the interval given by equation 28, there exists some $\delta_\beta < 1$, such the right hand side of this equation is less than $\delta_\beta$ for sufficiently small $\epsilon$.

Given this, the probability that agent $j$ continues to play irrationally on the interval $[\bar{t}_1 - \epsilon, \bar{t}_1 - (1 - \beta)e]$ must be less than $\delta_\beta$. However, at time $\bar{t}_1 - \beta\epsilon$ the same argument can be repeated on the interval $[\bar{t}_1 - (1 - \beta)e, \bar{t}_1 - (1 - \beta^2)e]$. And so the the probability of agent $j$ continuing to behave like the irrational type until time $\bar{t}_1 - (1 - \beta^k)e$ must be less than $\delta^k$. For sufficiently large $k$ this means $\delta^k < \xi(\tau^*)$. However, the probability of acting like the irrational type until $\bar{t}_1 - (1 - \beta^k)e$ must be at least $\xi(\tau^*)$ given that the irrational type never deviates from her demand. This contradiction means that $\bar{t}_1 > \tau^*$ and proves that agent $i$ must concede to agent $j$ immediately at $\tau^*$.

**Proof of Lemma 2**

I argue that any equilibrium must satisfy the conditions i)-iii). Let $\sigma = (F^1, F^2)$ define an equilibrium. Let $u^t_i$ be the expected utility of a rational agent $i$ who concedes at time $t$. Define $A^t = \{t : u^t_i = \max_t u^t_i\}$. Since an equilibrium is presumed, $A^t \neq \emptyset$. Also let $T^*$ be defined as the minimum time such that agents’ bargaining positions are compatible $(\alpha(t) + \alpha(t) = 1)$ or T otherwise. Finally define $\tau^* = \inf \{t: F^t(t) \geq 1 - \epsilon \}$. Notice that the first part proof of Lemma 1 shows that one must have $\tau^* < T$. As argued in the text of section 2 if $T^* < T$ then all agents
(rational and irrational) must concede at $T'$ and so in all cases $t' \leq T'$, that is rational agents must concede weakly before committed ones.

I first make a series of subclaims:

(a) $\tau' = \tau$. Given that rational agents concede before committed agents this simply says that a rational agent will not delay once she knows that her opponent is committed, which is true by assumption. This proves condition ii).

(b) If $F'$ jumps at $t < T'$ then $F'$ does not jump at $t$. If $F'$ had a jump at $t$, then because bargaining positions are incompatible and continuous and the deadline distribution is continuous at $t < T'$, agent $j$ would receive a strictly higher utility by conceding an instant after $t$ than by conceding at exactly $t$. This proves condition iii). Furthermore if $F'$ jumps at $t \in (0, T')$ then $j$ cannot concede with positive probability on the interval $(t - \varepsilon, t]$ for some $\varepsilon > 0$, because doing so would again forgo an expected profit bump from waiting until just after $t$.

(c) If a rational agent $j$ does not concede with positive probability on an interval $(t_i, t_2)$, with $t_2 < T'$ then neither does agent $i$. Conceding at any point on this interval would give agent $j$ a strictly lower payoff than conceding momentarily before that time given the assumption of optimal concession against a know irrational agent $(1 - \alpha'(t))(1 - G'(s)) > -\alpha'(t)$. This in turn implies that for $t \in (0, T')$ there can be no mass concession by either agent, and so any concession on this interval must be continuous. To see this final claim notice that mass concession by $j$ at $t$ induces an interval of non-concession by $j$ on $(t - \varepsilon, t]$, with $\varepsilon > 0$ by (c), and I have just argued this means there cannot be concession by agent $i$ on this interval.

(d) If $T' = \tau'$ then both agents must concede with positive probability on an interval $(T' - \varepsilon, T')$, for all sufficiently small $\varepsilon > 0$. Being conceded to at time $T'$ is payoff equivalent to conceding, as given continuous bargaining positions demands must be exactly compatible at $T'$. Thus, if the claim were not true, without loss of generality one may assume that agent $j$ does not concede with positive probability on the interval $(T' - \varepsilon, T')$, which by (c) must induce agent $i$ not to concede on this interval either. More generally even if there is mass concession by $j$ at $T'$ from the perspective of agent $i$’s payoffs one can assume there is not.

(e) If $t' \leq \tau'$ then $F'(t') > F'(t')$. Suppose not, then let $t'$ be the supremum of times such that $F'(t) = F'(t')$, that is $t' = \sup \{t : F'(t) = F'(t')\}$. Given (d) one has $t' < \tau'$ and so there is no mass concession by $i$ at $t'$. This must imply $F'$ is constant on $(t' - \varepsilon, t')$, for some $\varepsilon > 0$. By (c) this must also be true for $F'$. But with no mass concession by agent $j$ at any $t \in [t', \tau')$ conceding at or momentarily after $t'$ must give agent $i$ a strictly lower utility than conceding at $t' - \varepsilon < t'$, and so $t'$ cannot be the supremum.

(f) If $F'$ is continuous at $t$, then $u'_i$ is continuous at $s = t$. This follows immediately from the definition of $u'_i$.

Given (e) it follows that $A'$ is dense in $[0, \tau']$. From (c) and (d) one may assume $F'$ is continuous on $(0, \tau')$ and hence by (f) $u'_i$ is continuous on $(0, \tau')$. This implies that $A' \supseteq (0, \tau')$, which in turn implies that $u'_i$ is differentiable with derivative equal to 0 on $(0, \tau')$. This is exactly what is needed for condition i) of the Lemma.

Proof of Lemma 3

The equilibrium density of agreements at time $t \in (0, T']$ conditional on the deadline not having hit is $f(t)(1 - F(t)) + f(t)(1 - F(t))$. This means that the total probability of an agreement before time $t \in [0, T]$ is:

$$
M(t) = (1 - c'c) + \int_0^{\min(t, T')} \left( f(s)(1 - F(s)) + f(s)(1 - F'(s)) \right) (1 - G(s)) ds \\
= (1 - c'c) + c'c \int_0^{\min(t, T')} \left( f(t) + \frac{f(t)}{1 - F(t)} \right) (1 - F(t))(1 - F(t))(1 - G(s)) ds \\
= (1 - c'c) + c'c \int_0^{\min(t, T')} \left( rK' + rK' + (K' + K) \frac{g(t)}{1 - G(t)} e^{-(rK' + rK')(t - 1)} (1 - G(s))^{K' + K'} ds \\
$$

Where this used the facts $(1 - F(t)) = c'(1 - F'(t))$ and $(1 - F'(t)) = (c'(1 - G(t)))^{K'}$. For sufficiently small $r$ one has $(1 - F'(t))$ arbitrarily close $(1 - G(t))^{K'}$, uniformly for all $t$. This in turn has implications for exhaustion times, meaning $(1 - F'(T')) = c' = 0$ is arbitrarily close to $(1 - G(T'))^{K'}$. Time zero
concession is then defined by:

$$c^i = \frac{\mu^i(\alpha^i)}{1 - \bar{F}(T^*)} = \frac{\mu^i}{(e^{-\alpha^i T} (1 - G(T^*)))^{K^i}} \approx \min \{1, \mu^i(\alpha^i) K^i\} = \bar{c}^i$$

Where the approximation is arbitrarily good for $r^i$ small. Given the assumption that $\frac{\mu(\alpha)}{1 - \bar{F}(T)} \leq \frac{\mu(\alpha)}{1 - \bar{F}(T)}$, this means agent 1 makes (weakly) positive initial mass concession in the limit.

All this ensures that for sufficiently small $r^i$, one can approximate $M(t)$ arbitrarily closely, and uniformly all $t$, by $\bar{M}(t)$ which is defined as follows:

$$\bar{M}(t) = (1 - c^i \bar{c}^i) + c^i \bar{c}^i \int_0^{\min(t,T^*)} (K^i + K^j)g(t)(1 - G(s))^{K^i + K^j} ds$$

$$= (1 - c^i \bar{c}^i) + c^i \bar{c}^i K^i + K^j (1 - (1 - G(\min(t,T^*))))^{K^i + K^j}$$

$$= (1 - z^i(z^2) K^i) + z^i(z^2) K^i + K^j (1 - (1 - G(\min(t,T^*))))^{K^i + K^j}$$

This in particular ensures the first converge claim in the proposition for $M(0)$, equation 14. For the second claim regarding $M(T - e) - M(0)$, equation 15, notice that if $G(T - e)$ is arbitrarily close to zero, then $M(T - e)$ must be arbitrarily close to $\bar{M}(0)$. Given that the deadline distribution converges weakly to the point distribution at time $T$, one must indeed then have $M(T - e) - M(0)$ arbitrarily small for all sufficiently large $n$.

The total probability of agreement before the deadline is $M(T) = M(T^*)$. As implied above for $r^i$ sufficiently small $(1 - G(T^*))$ is arbitrarily close to $(z^2)^{K^i}$, which means:

$$\bar{M}(T) = \bar{M}(T^*) = (1 - z^i(z^2)^{K^i}) + z^i(z^2)^{K^i} K^i + K^j (1 - (z^2)^{K^i + K^j})$$

With the approximation arbitrarily good for small $r^i$. Notice in particular, that the right hand side is independent of the particular deadline distribution. This proves the fourth claim in the proposition, equation 17, regarding the convergence of the total probability of disagreement $1 - M(T)$. The third claim, equation 16, regarding $M(T) - M(T - e)$ is then implied by equations 14, 15, and 17. The subordinate claims $\lim M(0) < 1 - z^2$, $\lim (M(T) - M(T - e)) = 0$, and $\lim (1 - M(T)) > z^2$, are easily verified given $z^i < 1$.

**Proof of Proposition 3**

Given a finite set of types, consider any subsequence of rational agent demand choice probabilities $\mu^i(\alpha^i), \mu^j(\alpha^2)$ which converge to a limit for each demand combination with $a^1 + a^2 > 1$. These define limiting posterior probabilities of irrationality for $z^i(\alpha^i)$ and $z^j(\alpha^2)$. The results of Lemma 3 then immediately follow through for each subgame with $a^1 + a^2 > 1$ with $\lim z^i \in (0, 1)$ for both agents, with $z^i$ replaced by $\lim z^i$ in the limiting solutions. In particular this immediately proves claim 19, that no agreements are reached on the interval $(0, T - e)$ in the limit.

Given this, all that remains to be proved is that there is at least one subgame with $a^1 + a^2 > 1$, in which $\bar{M}(T^*) < 1$ for both agents (one always has $z^i \geq z^i(\alpha^i)$). If that is true then there exists some subgame is reached with a probability bounded above $\bar{z}^i(\alpha^i)z^2 \pi(\alpha^2)$ for large $n$, and such that the total probability of disagreement in that subgame is bounded above $\bar{z}^i(\alpha^i)z^2 \pi(\alpha^2)$ with the probability of an $11^{th}$ hour deal in that subgame bounded above $0$.

In particular, I shall prove this for the case of $a^i = \text{max} C^i$. Suppose agent 1 did not imitate such type with strictly positive limit probability along some subsequence, then $z^1 \rightarrow 1$. If rational agent 2, made a counterdemand $\alpha^2 > 1 - \text{max} C^1$ with strictly positive limit probability, so that $\lim z^2 < 1$, then for each such counterdemand agent 2 must concede in this subgame with probability approaching $1 - \lim z^2$. To see this not that agent 1’s exhaustion time $T^*$ must satisfy $\bar{z}^i = (1 - G(T^*))e^{-\alpha^i T^*}k^i$ which converges to 1, and so the denominator of $\frac{\bar{z}^i}{1 - G(T^*)e^{-\alpha^i T^*}k^i} = \bar{c}^i$ necessarily also converges to 1, giving the result. This in turn means that agent 1’s limiting payoff from (the
deviation of) always demanding $\max C^1$ and conceding an instant after time zero is:

$$(1 - \varepsilon^2)\max C^1 + \varepsilon^2 \sum_{\alpha > 1 - \max C^1} \pi(\alpha^2)(1 - \alpha^2)$$

This is however, strictly greater than her best possible payoff from imitating any demand below $\max C^1$, given that behavioral types are committed to their demands. This is a contradiction and so 1 must imitate $\max C^1$ with strictly positive limit probability, so that $\lim z^1 < 1$ for this demand.

Similarly, suppose rational agent 2 did not imitate $\max C^2$ with positive limit probability in response to the demand $\max C^1$. This would imply $\lim z^2 \to 1$ in the subgame with $\max C^1$, $\max C^2$, which in turn ensures agent 1 would concede with probability approaching $1 - \lim z^1 = 0$ in the limit, following the same reasoning given in the previous paragraph about the exhaustion time $T^2$. This means agent 2’s limiting (deviation) payoff from always demanding $\max C^2$ and conceding an instant after zero is:

$$(1 - \lim z^1)\max C^2 + \lim z^2(1 - \max C^1)$$

Which is strictly better than she could obtain by making any other counterdemand. This is a contradiction implying $\lim z^2 < 1$, which completes the proof.

**Proof of Proposition 6**

The proof proceeds in 3 steps. First, I characterize some of the features of an equilibrium with a given (possibly discontinuous) type for agent NR and a Rubinstein type for agent R. Next, I create bounds on the probability of concession by agent NR in that equilibrium at the discrete times $t_{k+2}$. Finally, I proceed to examine the limit of the game as $\varepsilon_n \to 0$ allowing multiple types.

**Step 1:**

Consider an equilibrium when agent NR imitates a given type $\alpha^R$ while R imitates the Rubinstein demand $\alpha^R = \alpha^R_*$. As before let $F'(t)$ record the total probability that agent 1 has conceded or revealed rationality prior to time $t$ conditional on $j$ not having done so. Let $u'_i$ be the expected utility of a rational agent 1 who reveals rationality first at time $t$. Define $A' = \{t : u'_i = \max, u'_i\}$. Since an equilibrium is conjectured $A' \neq \emptyset$. Again let $T'$ be defined as the minimum time such that irrational bargaining positions are compatible ($\alpha'(t) + \alpha'(t) = 1$) if such a time exists and $T$ otherwise. Again define $\tau' = \inf \{t : F'(t) \geq 1 - \varepsilon^2\}$. Notice that, given continuous demands on the final interval $[t_{m+2}, T]$, the first part of the proof of Lemma 1 shows that $\tau' < T$. Moreover, if $T' < T$ then as previously it is without loss of generality to assume that bargaining ends immediately at $T'$ even if $T' = t_{k+1-1}$, and so in particular $T^{NR} \leq T'$. One can also assume $\tau' = T'$ so long as $T' = t_{k+1-1}$.

Let $d'(t) = 1$ if agent 1 can change her demand at time $t$, and $d'(t) = 0$ otherwise. This is used to define:

$$\phi^R(t) = \max\left\{e^{-\tau^R(t)-t} \frac{1}{1 - G(t)} \left[\max_{\alpha^R(t)} \min[1 - \alpha^{NR}(t_{k+2})], \alpha^R(t), \alpha^R(t_{k-1})]\right] + d^R(t_{k+1}) \left[\frac{1}{\max_{\alpha^R(t)} \min[1 - \alpha^{NR}(t_{k+2})], \alpha^R(t_{k-1})]\right] : t_{k+1} \geq t\right\}$$

This is effectively the maximum expected utility that agent R could obtain by waiting from $t$ until some $t_k \geq t$ when agent NR does not concede between $t$ and $t_{k-2}$. Notice that this function is continuous and indeed if $\phi^R(t') > (1 - \alpha^{NR}(t'))$ for some $t' \in [t_{k+2}, t_{k+1-2}]$ one must have $\phi^R(t) > (1 - \alpha^{NR}(t'))$ for $t \in [t', t_{k+1-2}]$ given the assumption $(1 - \alpha'(t)) \left(\frac{\varepsilon}{1 - G(t)} + r'\right) > -\alpha'(t)$ on that interval. Next, define the region $X = \{t \neq t_k : \phi^R(t) < (1 - \alpha^{NR}(t))\}$. I now prove some facts about the nature of the equilibrium.

(a) $\tau^R \geq T^{NR}$. Given that rational agents concede before committed agents this simply says that a rational agent NR will not delay once she knows that her opponent R is committed. A slight modification of this statement concerns when agent R reaches probability 1 of rationality at $t_{k-2}$, $t_{k-1}$, or $t_{k+1}$ in which case NR must certainly have reached probability 1 of irrationality before $t_{k+2}$.

(b) If agent 1 does not concede with positive probability on some interval $(r' - \varepsilon, r') \subseteq (t_{k+2}, t_{k+1-2}]$ with $r' < T'$
and $\varepsilon > 0$, then neither does agent j. Given demands satisfy $(1 - a'(t)) \left( \frac{g(t)}{1 - a(t)} + r' \right) > -a''(t)$ on this interval, it is always preferable for j to concede earlier on this interval rather than at any given time on it.

(c) There can be no mass concession by agent i at any $t' \in (t_k + t_{k+1}, t_{k+1} + 2)$ such that $t' < T$. Suppose there was, then given continuous demands agent j would not concede on the interval $(t' - \varepsilon, t']$ for some $\varepsilon > 0$, preferring to wait and receive such concession, which by (b) would then make it optimal for i not to concede at $t'$.

(d) There can be no joint mass concession at any $t_k + 2 < T'$, as given incompatible positions agents would strictly prefer to wait an instant longer to receive such concession.

(e) Agent R cannot concede at any time such that $\phi^R(t) > (1 - a^NR(t))$ or reveal rationality at $t_k + 1$ when $\phi^R(t_k + 1) > \phi^R(t_k + 1)$. If the inequality holds on some interval $(t' - \varepsilon, t'] \subseteq (t_k + 2, t_{k+1} + 2)$ agent NR cannot concede on that interval either by (b).

(f) There can be no mass concession by agent i at $t_k + 2 < T$. Suppose there was, then agent j must optimally not concede on some interval $(t_k + 2 - \varepsilon, t_k + 2)$ for some $\varepsilon > 0$ as conceding at $t_k + 2$ would otherwise give a strictly higher payoff. Hence there must certainly be mass concession by agent j at $t_k + 2$ as well or else by (b) i would not concede at $t_k + 2$. But then whichever agent has an opportunity to change their demand at $t_k + 2$ would then rather wait to concede at $t_k + 1$, rather than concede at $t_k + 2$.

(g) If $t_k + 2 < T'$ then it is without loss of generality to assume that agent NR does not reveal rationality at times $t_{k-1}$ or $t_{k+1}$. Any first revelation by agent NR at these time must amount to concession to the demand $a^R(t_k)$. Hence any revelation at time $t_{k+1}$ can be instead regarded as happening at time $t_k + 2$ (if R ever conceded with positive mass at $t_k + 2$ NR’s revelation would not be optimal). Suppose however NR conceded with positive mass at time $t_{k+1}$. If $\phi^R(t_{k+1}) > (1 - a^NR(t_{k+1}))$ then such mass concession will certainly ensure that R does not concede on the interval $(t_k + 2 - \varepsilon, t_k + 2)$ for some $\varepsilon > 0$, which by the logic of (c) ensures rational NR would prefer to concede strictly before $t_k + 1$. Alternatively if $\phi^R(t_{k+1}) > (1 - a^NR(t_{k+2}))$ then agent R will certainly reveal rationality (without conceding) at $t_{k+1}$ and so NR’s mass concussion can again be viewed as occurring at $t_k + 2$ (as if R conceded at $t_k + 1$ or $t_k + 2$, concession by NR at $t_k + 1$ would be suboptimal).

(h) There can be no mass concession by NR at $t_k + 2 < T'$ unless either $\phi^R(t_{k+1} - 1) > (1 - a^NR(t_{k+2}))$ or $\phi^R(t_{k+2}) > a^NR(t_{k+2})$. Suppose otherwise $\phi^R(t_{k+2}) < (1 - a^NR(t_{k+2}))$ and $\phi^R(t_{k+1}) < a^NR(t_{k+2})$, then mass concession by agent NR at $t_k + 2$ is inconsistent with revelation of rationality by R at $t_k + 1$ or on $t_k + 2 - \varepsilon, t_k + 2$ for some $\varepsilon > 0$ (she would prefer to wait). But in which case agent NR would prefer to concede strictly before $t_k + 2$ instead.

(i) Let $[r', r''] \in [t_k + 2, t_{k+1} + 2]$ with $r' < r''$ be such that $\phi^R(t') > (1 - a^NR(t'))$ then $F''(r') > F(r')$. This states that $F''$ is strictly increasing on X. Suppose not then let $r' = sup \{ t : F'(t) = F'(r') \}$ and $r'' = min \{ t' > r' \}$. Clearly $r' > r''$ by (b). Suppose $r''' = r'' \in (t_{k+1} + 2, t_{k+2} + 2)$, given no mass concession at $t \in (t_k + 2, t_{k+1})$ and $(1 - a^R(t)) \left( \frac{g(t)}{1 - a(t)} + r' \right) > -a''(t)$ conceding at $t' - \varepsilon$ (for small $\varepsilon > 0$) gives i a strictly higher utility than conceding at or just after $t'$. If $r'' = r''' \in (t_{k+2} + 2, t_{k+3} + 2)$ then i must make probabilistically reveal rationality at $t_{k+2}$ (given no mass concession at $t_{k+2}$). But concession at $t_{k+1}$ can give agent i a utility of $(1 - a^R(t_{k+2}))$ which must be strictly worse than conceding at $t_{k+2} - \varepsilon$, for small $\varepsilon > 0$. Revealing rationality without concession can give agent R a utility of $\phi^R(t_{k+2})$, but given $\phi^R(t''') < (1 - a^NR(t''))$ this is strictly worse than conceding at $t' + \varepsilon$, for small $\varepsilon > 0$. Entirely similar logic rules out $r'' = r_''' \in (t_{k+1} + 2, t_{k+2} + 2)$ (probabilistic revelation at $t_{k+1}$). Suppose then $r''' = r' = t_{k+2}$ by (d) at least one agent, say j does not concede with positive probability at $t_{k+2}$. In this case agent i would again strictly prefer to concede at $t' + \varepsilon$, for small $\varepsilon > 0$, than waiting to concede at $t_{k+2}$ (again given $\phi^R(t'') < (1 - a^R(t''))$).

(j) If $T' = t_1 \in (t_k + 2, t_{k+1} + 2)$ then both agents must concede with positive probability on an interval $(T' - \varepsilon, T')$, for all sufficiently small $\varepsilon > 0$. This follows for the same reason as in Lemma 2. Being conceded to at time $T'$ must be payoff equivalent to conceding, given that continuous bargaining positions on this interval mean that demands must be exactly compatible at $T'$.

(k) If $T' = t_{k+1}$ or $T' = t_{k+2}$ then $r'' = T'$. Given mutually compatible positions at $T'$ one must have $\phi^R(t_{k+2}) \geq \phi^R(t_k) > (1 - a^NR(t_{k+2}))$, ensuring agent R cannot concede on $(t_k + 2 - \varepsilon, t_{k+2})$ for some $\varepsilon > 0$. Equally, rational agent NR’s continuation payoff in the game conditional at $t_{k+2}$ must in all cases be less than $(1 - a^R(t_k))$ which means that conceding strictly prior to $t_{k+2}$ must be better than waiting until that time.
Given (i) it follows that \( \mathcal{A}' \) is dense in \([0, \tau^\text{NR}] \cap X\). From (c) and (j) one may assume \( F' \) is continuous on \((0, \tau^\text{NR}] \cap X\) and hence by (f) \( u'_t \) is continuous on \((0, \tau^\text{NR}] \cap X\). This implies that \( \mathcal{A}' \supseteq (0, \tau^\text{NR}] \cap X\), which in turn implies that \( u'_t \) is differentiable with derivative equal to 0 on \((0, \tau^\text{NR}] \cap X\). This implies rates of concession given by equation (5).

And in turn means that one must have agent NR’s reputation for \( t \leq \tau^\text{NR} \) given by:

\[
\mathcal{z}_{NR}^R(t) = \frac{z_{NR}^R}{\prod_{k \geq h_{n_2}, \tau_{k}=t}^R} \exp \left( \int_0^t \left[ 1_{\{s \in X\}} (1 - a_{NR}(s)) \left( r^R + \frac{g}{1 - e^s} \right) + a^R(s) - 1 \right] ds \right)
\]

Where \((1 - c_{h_{n_2}}^\text{NR})\) is the probability of concession by NR at \( t_{h_{n_2}} \). A similar expression can be given for agent R additionally incorporating revelation of rationality at \( t_{h_{n_2}} \), however, in fact it is sufficient to consider the following bound on agent R’s probability of rationality at time \( t \leq \tau^R \):

\[
\mathcal{z}_R^R(t) = \mathcal{z}_R^R \exp \left( \int_0^t \left[ 1_{\{s \in X\}} (1 - a^R(s)) \left( r^R + \frac{g}{1 - e^s} \right) + a^R(s) - 1 \right] ds \right)
\]

**Step 2:**

The next task is creating bounds on the probability of concession by NR at \( t_{h_{n_2}} \in (0, \tau^\text{NR}] \). Suppose that \((t_{k-2} - \epsilon, t_{k-2}) \subseteq X \cap (0, \tau^\text{NR}] \) for some \( \epsilon > 0 \) and so also by continuity \( \mathcal{y}^R(t_{k-2}) \leq (1 - a^R(t_{k-2})) \). The first task is to show:

\[
\mathcal{y}^R_{k-2} = \frac{a^R(t_{k-2})}{a^R(t_{k-2}) + a^R(t_{k-1})} \geq 1
\]

The second inequality must certainly hold given incompatible demands at \( t_{h_{n_2}} \) (which is implied by \((t_{k-2} - \epsilon, t_{k-2}) \subseteq X \cap (0, \tau^\text{NR}] \) ). Suppose that the first inequality did not hold, then agent R would strictly prefer to wait and concede an instant after \( t_{h_{n_2}} \) rather than concede herself on the interval \((t_{k-2} - \epsilon, t_{k-2}) \) for sufficiently small \( \epsilon > 0 \). This contradicts that \( F' \) is increasing on that interval.

Next suppose that \( t_{h_{n_2}} - \epsilon \notin X \) (for small \( \epsilon > 0 \)), and let \( \bar{t} = \inf \{ t : F^R(t) > F^R(t_{k-2} - \epsilon) \} > t_{h_{n_2}} \). Necessarily, \( \bar{t} \in \{h_{n_1}, h_{n_2}, h_{n_2} \} \), for some \( h \geq k \) and there must be probabilistic revelation of rationality by agent R in some form at \( \bar{t} \). Moreover, \( F^R(t_{k-2}) = F^R(\bar{t}) \) as rational NR would otherwise strictly prefer to concede at \( t_{h_{n_2}} - \epsilon \) for some \( \epsilon > 0 \). Furthermore if \( \bar{t} = t_{h_{n_2}} \) then clearly \( c_{h_{n_2}}^\text{NR} = 1 \), as NR would then strictly prefer to wait to concede an instant after \( t_{h_{n_2}} \) in order to receive R’s mass concession.

Alternatively if \( \bar{t} \in \{h_{n_1}, h_{n_2} \} \) one can assume \( \mathcal{y}(\bar{t}) > (1 - a^R(t_{h_{n_2}})) \) as if not then certainly \( \mathcal{y}^R_{h_{n_2}} = 1 \), otherwise R would not reveal rationality at \( \bar{t} \) but wait until just after \( t_{h_{n_2}} \). In this case:

\[
\mathcal{y}^R_{h_{n_2}} = \frac{a^R(\bar{t}) - 1}{a^R(t_{h_{n_2}}) + a^R(\bar{t})} > 0
\]  \hspace{1cm} (29)

Suppose the first inequality did not hold, in this case agent R would not find it optimal to concede at \( \bar{t} \) but would instead wait until an instant after \( t_{h_{n_2}} \) in the hope of receiving such mass concession, contradicting the definition of \( \bar{t} \) as the supremum.

To show the second inequality, suppose it does not hold so that \( a^R(\bar{t}) > a^R(\bar{t}) \). This must imply \( \mathcal{y}^R(\bar{t}) \leq (1 - a^R(\bar{t})) \). But for NR to find it optimal to concede at \( t_{h_{n_2}} \) to the demand \((1 - a^R(\bar{t})) \) instead of at \( t_{h_{n_2}} - \epsilon \) for small \( \epsilon > 0 \) implies:

\[
(1 - a^R(t_{h_{n_2}} - \epsilon)) \leq (1 - a^R(\bar{t}))e^{-\mathcal{A}'(t_{h_{n_2}} - \epsilon)} \frac{1 - G(\bar{t})}{1 - G(t_{h_{n_2}} - \epsilon)}
\]

But this cannot hold, given that the Rubinstein demand satisfies \((1 - a^R(t)) \left( \frac{g(0)}{1 - e^s} \right) + r^R) > -a^R(t) \). The above argument can also be extended to show that for a sequence of equilibria (taking subsequences if necessary) that \( \lim_{h_{n_2}} \geq \mathcal{t} > 0 \). Suppose not, then consider the sequence \( t_{h_{n_2}} - \epsilon_n \notin X \) (for small \( \epsilon_n > 0 \) with mass
concession bounded by equation 29. This will lead to a uniform positive bound on $c_{NR2}^R$ unless $l_{i}^R(i) = a^R(i)$, which in turn means that $lim_{i}^NR(i) = (1 - a^R(i))$, where $i \in \{t_{h-1}, t_{h+1}\}$. Given $a^R(i) + a^NR(t_{h-2}) > 1$ there exists an $\bar{\epsilon} > 0$ such that for all sufficiently large $n$, agent $R$ will not concede on $[t_{h-2} - \bar{\epsilon}, t_{h-1}]$. Using the same argument as above then shows that for sufficiently large $n$, agent $NR$ would then strictly prefer to concede at $t_{h-2} - \bar{\epsilon}$ rather than wait until $t_{h-2}$ to receive a payoff arbitrarily close to $(1 - a^R(i))$, given the bound on the rate of decline of agent $R$’s demand. This is a contradiction.

Step 3:

Consider a sequence of equilibria as $\varepsilon_n \to 0$ when rational agent $NR$ imitates some type such that $a^NR(0) + a^R(0) > 1$ with positive limit probability (taking a subsequence if necessary), this implies that there is a bound $\bar{\varepsilon}_n < L_2$ on agents initial probabilities of irrationality even when a rational agent $R$ imitates her Rubinstein demand with probability 1. Again considering a subsequence if necessary, let $lim_{i}^NR$ exist. Along this (sub)sequence:

$$\varepsilon_{NR}^R(\tau^R) = \frac{z_{NR}^R}{\prod_{k_{R}^2 > k_{R}^2}^c_{NR2}^R} \sum_{k_{0}^2}^c_{NR2}^R \exp \left( \int_{0}^{\tau^R} \frac{(1 - a^NR(s))(R^R + \frac{g(s)}{1 - G(s)}) + a^NR(s)(R^R + \frac{g(s)}{1 - G(s)}) - a^R(s)}{a^NR(s) + a^R(s) - 1} ds \right) = 1$$

Given that $\varepsilon_{\varepsilon_n}^R \to 0$ and the bound $c_{k_{0}^2}^R \geq \bar{\epsilon}$ for some $\bar{\epsilon} > 0$ and all $k > 0$ such that $c_{k_{0}^2}^R \leq \bar{\epsilon}^R$ one must have that either $c_{k_{0}^2}^R \to 0$ or $\tau^R \to T' \in (t_{k_{2}}, t_{k_{2}+1}]$ for some $k'$, where $k' = m$ if $T' = T$. The first of these would mean that agent $NR$ conceded to $R$ at time zero with probability approaching 1 in the limit.

Suppose otherwise, for sufficiently large $n$ one must then have $\tau^R \in (t_{k_{2}}, t_{k_{2}+1}]$ and given agent $NR$’s demand on this interval satisfies $(1 - a^NR(i))(R^R + \frac{g(s)}{1 - G(s)}) - a^NR(i)$ a rational agent $R$ would certainly concede against an opponent she knew to be irrational and so $\tau^R = \varepsilon_{NR}^R = T^*$. 

Let region $X$ be described as $X = \bigcup_{i=0}^{m}(t_{h_{0}, h_{2}+1}, x_{h})$, a series of discrete intervals where $x_{h} \in [t_{h_{0}, h_{2}+1}, t_{h_{0}+1}]$. For large $n$, one can then proceed in a similar fashion to Proposition 5, again using $(l(i) = a^R(i) + a^NR(i) - 1$ to give:

$$c_{k_{0}^2}^R \leq \frac{z_{\varepsilon_n}^R}{z_{\varepsilon_n}^R(\tau^R)} \leq \frac{\varepsilon_n^R}{\prod_{k_{0}^2 \geq k_{0}^2}^c_{NR2}^R} \sum_{k_{0}^2}^c_{NR2}^R \exp \left( \int_{0}^{\tau^R} \frac{(1 - a^NR(s))(R^R + \frac{g(s)}{1 - G(s)}) + a^NR(s)(R^R + \frac{g(s)}{1 - G(s)}) - a^R(s)}{a^NR(s) + a^R(s) - 1} ds \right)$$

$$= \frac{L_2}{\bar{\varepsilon}_n^R} \prod_{k_{0}^2}^c_{NR2}^R \exp \left( \int_{0}^{\tau^R} \frac{(1 - a^NR(s))(R^R + \frac{g(s)}{1 - G(s)}) + a^NR(s)(R^R + \frac{g(s)}{1 - G(s)}) - a^R(s)}{a^NR(s) + a^R(s) - 1} ds \right)$$

$$\leq \frac{L_2}{\bar{\varepsilon}_n^R} \prod_{k_{0}^2}^c_{NR2}^R \exp \left( \int_{0}^{\tau^R} \frac{(1 - a^NR(s))(R^R + \frac{g(s)}{1 - G(s)}) + a^NR(s)(R^R + \frac{g(s)}{1 - G(s)}) - a^R(s)}{a^NR(s) + a^R(s) - 1} ds \right)$$

$$= \frac{L_2}{\bar{\varepsilon}_n^R} \prod_{k_{0}^2}^c_{NR2}^R \exp \left( \int_{0}^{\tau^R} \frac{(1 - a^NR(s))(R^R + \frac{g(s)}{1 - G(s)}) + a^NR(s)(R^R + \frac{g(s)}{1 - G(s)}) - a^R(s)}{a^NR(s) + a^R(s) - 1} ds \right)$$

The first proceeds from the fact $\varepsilon_{NR}^R(\tau^R) \leq 1$ and the second inequality from substituting in the equation defining $c_{k_{0}^2}^R$ and the bound developed for $z_{\varepsilon_n}^R(\tau^R)$. The third line uses the same trick as in Proposition 5 to simplify the exponential term, the bound on $z_{\varepsilon_n}^R < L_2$ (given that type $NR$ is imitated with positive limit probability), and the bound $c_{k_{0}^2}^R \geq \bar{\epsilon}$ for all $k > 0$ such that $c_{k_{0}^2}^R \leq \bar{\epsilon}^R$. The final line simply follows from integration of the integral. Given that $G(t) < 1$ and $l(t) > 0$ for all for all $t < T'$, all terms on the final line are bounded away from zero, apart from $l(T')$ and $(1 - G(T'))$. One of these terms must converge to 0 as $T' \to T'$ given the definition of $T'$, which means $c_{k_{0}^2}^R \to 0$, the result sought.

References


