A THEORY OF PLEDGE-AND-REVIEW BARGAINING*

Bård Harstad

bard.harstad@econ.uio.no

June 18, 2021

Abstract

This paper presents a novel bargaining game where every party is proposing only its own contribution, before the set of pledges must be unanimously approved. I show that, with uncertain tolerance for delay, each equilibrium pledge maximizes an asymmetric Nash product. The weights on others' payoffs increase in the uncertainty, but decrease in the correlation of the shocks. The weights vary pledge to pledge, and this implies that the outcome is generically inefficient. The Nash demand game and its mapping to the Nash bargaining solution follow as a limiting case. The model sheds light on the Paris climate change agreement, but it also applies to negotiations between policymakers or business partners that have differentiated responsibilities or expertise.

Keywords: Bargaining games, the Nash program.

JEL codes: C78, D78.

* I am grateful to the editor, Tilman Borgers, an associate editor, and a referee. I also thank Geir Asheim, Ernesto Dal Bo, Hülya Eraslan, Amanda Friedenberg, Faruk Gul, Paolo Piacquadio, Leo Simon, Jean Tirole, Joel Watson, Asher Wolinsky, and audiences in Adelaide (AARES pre-conference), U. Autonoma de Barcelona, U. of Barcelona, the BEET workshop at BI, UC Berkeley, UC3M, UC San Diego, U. of Chicago, CREST-Ecole Polytechnique, EIEF, ESEM 2018, University of Essex, HEC Paris, Hong Kong Baptist University, Ifo institute, London School of Economics, Manchester University, University of Melbourne, MIT, National Taiwan University, National University of Singapore, Northwestern University, University of Oslo, UPF, Princeton University, University of Notre Dame, Queen Mary University, Singapore Management University, Stanford GSB, SURED 2018, Toulouse School of Economics, the 2019 Wallis conference, and WCERE 2018. Johannes Hveem Alsvik, Marie Karlsen, Valer-Olimpiu Suteu, and Kristen Vamsæter provided excellent research assistance. Frank Azevedo helped with the editing.

-The Paris talks were a bit like a potluck dinner, where guests bring what they can.

The New Yorker, December 14, 2015

1. Introduction

Real-world negotiations differ substantially from how we typically model them. Standard bargaining models permit a proposer to propose a specific point in the space of alternatives. In business as well as in politics, however, a party is often emphasizing – or limiting attention to – its own individual contribution or demand.

For instance, the negotiations leading up to the 2015 Paris Agreement on climate change have been characterized as "pledge and review" (P&R). Before the agreement was signed, each party was asked to submit an intended nationally determined contribution. For most developed countries, the pledge specified an unconditional cut in the emissions of greenhouse gases in the years following 2020. Also in other situations – ranging from legislative bargaining to negotiations among business partners and experts – it is frequently the case that each party proposes only its own contribution or dimension, even though everyone must accept the entire vector of contributions.

The novel feature of P&R bargaining, the way I formalize it, is that each party is permitted to propose the outcome of only one single dimension of the vector describing the outcome. This dimension can be interpreted as the party's individual contribution. I assume, for simplicity, that parties propose their pledges simultaneously. If some parties find the vector of pledges unacceptable, the procedure can start again.

At first, the procedure appears rather nonsensical. In the absence of any uncertainty, a country can always obtain approval for a contribution level that is slightly less than what the other parties expect in the future, since disapprovals lead to costly delays. Therefore, a trivial equilibrium of this game coincides with the noncooperative outcome, where every party simply maximizes one's own utility (Theorem 0).

There will, realistically, be uncertainty regarding the other parties' willingness to reject and delay an agreement. In politics, for example, a negotiator's willingness to reject and delay depends on the media picture and the set of other issues urgently needing one's attention. A simple formalization of this uncertainty is to let each party's discount rate be influenced by shocks that are distributed i.i.d. over time. With such uncertainty, I show that the parties may be willing to contribute significant amounts, since each party may fear that a less attractive pledge can lead to rejections and delays.

I first present a folk theorem (Theorem 1) stating that every strictly Pareto-optimal vector of pledges can be supported in some subgame-perfect equilibrium (SPE) if the time lag between offers is sufficiently small. To make sharper predictions, I gradually narrow the set of equilibria by considering standard refinements. Suppose, as a start, that the strategies are stationary, Markov-perfect, or robust to a finite time horizon. With any such a refinement, we face an upper boundary for what a contribution can be (Theorem 2). This upper boundary turns out to be the only equilibrium outcome that survives the additional refinement to trembling-hand perfection (Theorem 3).

In this equilibrium, each party's equilibrium contribution level coincides with the quantity that maximizes an asymmetric Nash product. The weight that a party places on the payoff of another party depends on differences in the (expected) discount rates. This result confirms findings in the existing literature (Footnote 3), but the analysis also uncovers four novel results.

First, uncertainty helps. If it is more difficult to pin down a party's minimum discount rate, then that party's preference will more strongly influence the others' equilibrium pledges. Intuitively, a marginally less attractive offer will be rejected with a larger chance, and, to avoid a delay, other parties are willing to make more attractive pledges.[1]

Second, correlation hurts. If the discount rate shocks are positively correlated, then one party's cost of delay is likely to be small exactly when another party is willing to delay by rejecting the offers on the table. In this situations, therefore, a party does not find it necessary to reduce the risk. Consequently, the weights on other parties' payoffs are smaller when the willingness to delay is correlated across the parties.

Third, the weight a party places on the payoff of another party is independent of how many other parties there are. Consequently, if many parties benefit from the contributions,

[1]The result that uncertainty improves the bargaining outcome is in contrast to much of the literature (Rubinstein, 1985; Watson, 1998, and many others). Most recently, Friedenberg (2019) derives inefficient equilibria simply with off-path strategic uncertainty.

then each party contributes more. This result might appear to be efficient, altruistic, and in line with other bargaining outcomes, but here the intuition is that when there are many other parties, there is a larger risk that one of them will reject. The larger risk motivates each party to make a more attractive pledge.

Finally, and in sharp contrast to the literature discussed below, each party maximizes its own Nash product. The weights vary pledge-to-pledge, and the equilibrium vector is thus not Pareto optimal. If, for example, the variances in the shocks are small, then every party pays most attention to its own payoff.

Literature: By showing that each contribution maximizes an asymmetric Nash product, I contribute to the "Nash program," aimed at finding noncooperative games implementing cooperative solution concepts (Serrano, 2020). The Nash bargaining solution (NBS), axiomatized by Nash (1950), is implemented by the alternating-offer bargaining game of Rubinstein (1982); see Binmore et al. (1986).[2] The asymmetric NBS characterizes the outcome if there are asymmetric discount rates, recognition probabilities, or voting rules.[3] My contribution to the Nash program is to show that, with P&R, each equilibrium pledge maximizes an asymmetric Nash product where the weights not only reflect differences in the discount rates (in line with this literature), but also the extent of uncertainty in shocks and the correlation in shocks across the parties. More fundamentally, and in contrast to these articles, with P&R the equilibrium weights vary from one party's pledge to another's, implying that the bargaining outcome is not Pareto optimal.

The Nash demand game (NDG) was designed by Nash (1953) to implement the NBS. There is now a large literature investigating the extent to which the NDG implements the NBS.[4] Even though I assume that every utility function is continuous in every pledge, the NDG is a special case of my model if we in that game permit non-vanishing uncertainty regarding whether demands are compatible. When this uncertainty does vanish and the

[2] Although there can be multiple equilibria with more than two players (Sutton, 1986; Osborne and Rubinstein, 1990), the NBS is the unique equilibrium if we impose stationarity or reasonable consistency conditions (Asheim, 1992; Chae and Yang, 1994; Krishna and Serrano, 1996).

[3] See Miyakawa (2008), Okada (2010), Britz et al. (2010), and Laurelle and Valenciano (2008). Osborne and Rubinstein (1990:310) note that the asymmetric NBS satisfies all axioms in Nash (1950) except for symmetry.

[4] See Binmore et al. (1992), Abreu and Gul (2000), or Kambe (2000). Some contributions allow for strategic uncertainty in the NDG (Binmore, 1987; Carlsson, 1991; Andersson et al., 2018). Chatterjee and Samuelson (1990) study the repeated version.

utility functions become discontinuous, then, in the limit, my results generalize Nash's mapping from the NDG to the NBS. This mapping is generalized in that P&R bargaining game allows for many parties, multiple rounds, veto-rights, and uncertainty regarding the willingness to delay. With heterogeneous discount rates or shock distributions, the NDG implements an *asymmetric* NBS, I show, and my characterization of the weights is novel. If uncertainty does *not* vanish, then the outcome is *inefficient.* For the NDG, the interpretation is that each party takes too much risk.[5]

The literature on limited specifiability is small. Yildiz (2003) finds an efficient allocation when a proposer can only propose a price, while the other party can subsequently select any traded quantity given the price. More recently, Fukuda and Kamada (2020) present a general bargaining game where a party can propose a subset rather than a singleton in the set of alternatives. In contrast to my paper, negotiations must continue on the intersection of subsets and their focus is on the difference between asynchronous and synchronous moves. They show that asynchronicity of proposal announcements, and the existence of a common-interest alternative, lead to sharper predictions.

Outline: The next section discusses applications to climate agreements, legislative bargaining, issue linkages, and haggling among business partners or experts. Section 3 formalizes P&R bargaining and presents benchmark results before uncertainty is introduced. Section 4 starts with a folk theorem, before the set of equilibria is gradually reduced by referring to standard refinements. Section 5 shows that the Nash demand game, and the mapping from that game to the NBS, can be both generalized and proven in a special (limiting) case of the model. A number of generalizations are discussed in Section 6. Section 7 concludes. Appendix A contains all proofs. Additional generalizations are investigated in Appendix B (for online publication only).

[5] Relative to the literature on contribution games, the interpretation of the inefficiency result is, instead, that each party contributes too little. There is a large literature on the private provision of public goods. Because of the free-rider problem, which predicts small contributions, scholars have suggested that contributions may be larger because of threshold effects (Palfrey and Rosenthal, 1984; Marx and Matthews, 2000; Compte and Jehiel, 2004), refunds (Bagnoli and Lipman, 1989; Admati and Perry, 1991), voting (Ledyard and Palfrey, 2002), side payments (Jackson and Wilkie, 2005), or irreversibility (Battaglini et al., 2014) and there may be a large set of equilibria (Matthews, 2013). The standard equilibrium in the basic model below also predicts small contributions, and thus I complement the papers above by explaining when *uncertainty* can make the results consistent with larger contributions. The simultaneous offers and the need for unanimity, however, make my model quite different from this literature.

2. Applications: Climate treaties, legislative bargaining, and business linkages

The bargaining game is quite general and, as I will now explain, it might be applied to negotiations among countries attempting to agree on climate change policies, among political representatives who request public funds and share the total burden of the expenses, and among business partners that have different expertise or responsibilities.

Climate negotiations: The model and its assumptions are inspired by the pledge-and-review procedure associated with the Paris Agreement on climate change. P&R has been referred to as a "bottom-up" approach since countries themselves determine how much to cut nationally, without making these cuts conditional on other countries' emissions cuts.[6] In the absence of a world government, the set of contributions must be acceptable by everyone that contributes.[7] The need for consensus motivates the *review*: "By subjecting domestically determined mitigation pledges to the international review mechanism, the Paris Agreement ensures that the gap between the required level of action and the total sum of national measures becomes the subject of international policy deliberation and coordination" (Falkner, 2016:1120). Although the pledges were not made simultaneously in the Paris talks, the countries faced a common deadline and each of them was free to revise its pledge before that deadline.[8] Thus, it seems more reasonable to assume simultaneous pledges than assuming that there is a fixed sequential order.

Leading scientists and political scientists, such as Keohane and Oppenheimer (2016:142), have feared that "many governments will be tempted to use the vagueness of the Paris

[6] *The New York Times* (Nov. 28, 2015) wrote that: "Instead of pursuing a top-down agreement with mandated targets, [the organizers] have asked every country to submit a national plan that lays out how and by how much they plan to reduce emissions in the years ahead." Indeed, the Paris Agreement (Art. 4.2) states: "Each Party shall prepare, communicate and maintain successive nationally determined contributions that it intends to achieve." The official list of pledges is here: http://www4.unfccc.int/ndcregistry but for an overview see http://cait.wri.org/indc/#/.

[7] Global climate treaties require consensus and individual countries can indeed veto them: Before the 2009 Copenhagen negotiations, when P&R was first attempted, many countries had submitted pledges. However: "Objections by a small group of countries (led by Bolivia, Sudan, and Venezuela) prevented the Copenhagen conference from 'adopting' the Accord ... as a COP decision, which requires consensus (usually defined as the absence of formal objection)" (Bodansky, 2010:231; 238). As a consequence, negotiations were delayed for years.

[8] According to the Paris Agreement (Art. 4.2), the treaty: "Invites Parties to communicate their first nationally determined contribution no later than when the Party submits its respective instrument of ratification, accession, or approval of the Paris Agreement. If a Party has communicated an intended nationally determined contribution prior to joining the Agreement, that Party shall be considered to have satisfied this provision unless that Party decides otherwise."

Agreement, and the discretion that it permits, to limit the scope or intensity of their proposed actions." They continue (p. 149): "What is less clear is whether the resulting deals will [help] the world limit climate change. We can imagine high-level equilibria of these games that would do so. These equilibria would induce substantial cuts in emissions [but] we can also imagine low-level equilibria [that enable] both sides to pursue essentially business as usual." The theoretical results in this paper are very much in line with the various scenarios imagined by Keohane and Oppenheimer.

One lesson from this paper is that uncertain willingness to object and delay can motivate contributions that might be larger than without the uncertainty. In fact, pledge-and-review can be relatively attractive in climate negotiations: Predicted contributions are larger when there is a large number of parties, when the parties are very different and associated with unexpected shocks that are not highly correlated, and when the negotiations proceed so slowly that the future willingness to object and delay is hard to forecast. All these characteristics are familiar to climate negotiators.

Furthermore, the modesty can deter free riding (Finus and Maus, 2008). My follow-up paper (Harstad, 2021) embeds the pledge-and-review bargaining outcome in a dynamic climate policy with endogenous emissions, technologies, participation, and compliance, and argues that the P&R game can rationalize five facts regarding how the Paris Agreement differs from the Kyoto Protocol of 1997.

Domestic politics: There is a large literature in political economy where each district, or "spending minister," specifies one's own level of spending although the sum of expenses is a public bad that raises federal taxes, deficits, or debt (see the survey by Eraslan and Evdokimov, 2019). The model comes in two extreme variants: (i) In the *common-pool* setting (beginning with Weingast et al., 1981), there are no checks or balances, and no one can veto others' spending decisions. (ii) In analyses of procedural rules or bargaining situations, the ministers negotiate *efficiently* (Baron and Ferejohn, 1989; von Hagen and Harden, 1995). The model in this paper is an intermediate case that might be more realistic than the two extremes: each party is indeed permitted to decide on its own level of spending or, equivalently, spending cut, but the party risks delays if the spending levels are unacceptable to the others. One lesson is that the inefficiency is larger when the ministers are familiar to one another and face correlated shocks.

In Morelli (1999), parties make competitive demands, but he focuses on the sequence (determined by the head of state) and coalition formation (unanimity is not required), and there is no relation to the NBS.

Business and issue linkages: The P&R game can describe a situation in which multiple business partners must negotiate a package, and where each partner is recognized as an expert in, or as being responsible for, only a single dimension of the package: one partner describes the product quality, another offers a strategy for advertisements, while a third manages a set of retailers, for instance. In such meetings, it might be unrealistic to assume that a single partner is capable of proposing and describing a specific terminal outcome, as is normally assumed in bargaining theory. Instead, it can be more reasonable that each partner emphasizes what or how it can contribute, simply. After all, only the engineer is endowed with the vocabulary to describe technical solutions, the advertiser with the imagination to draw creative advertisements, and the manager sits on the alternative retailers' names and track records.

Because the parties make proposal on different things, I contribute to our understanding of issue linkages (see the survey by Maggi, 2016). Fershtman (1989) and In and Serrano (2004) consider the case in which the parties can only negotiate on one issue at the time. Fershtman analyzes disagreements over alternative fixed sequences. In and Serrano allow the proposer to propose a solution on any (but only one) of the issues. In other games, Horstmann et al. (2005) and Chen and Eraslan (2013) characterize the gains from linking various issues. This paper, in contrast, does not compare the timing or whether issues should be linked or not. The novel lesson from this paper is that when the parties simultaneously make proposals on their individual offers, then each equilibrium pledge is not only inefficient, but it also maximizes its own asymmetric Nash product where the weighs on others' payoffs are determined by factors that are new to the literature. Somewhat surprisingly, I show that business partners that are unfamiliar to one another may contribute more, because they are more concerned with the possibility that the opponents may otherwise reject.

3. A theory of pledge-and-review bargaining

3.1. A benchmark game

There are n parties, each endowed with a payoff function $U_i : \mathbb{R}^n \to \mathbb{R}$, $i \in N \equiv \{1, ..., n\}$. A typical terminal outcome is referred to as $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$. I assume, for tractability, U_i to be concave and continuously differentiable. Concavity is natural when x_i measures contributions to a public good, for example, since party i would then begin with the most cost-effective types of contributions. Both U_i and x_i are measured relative to the default outcome.[9]

Furthermore, I begin by making the *additional assumptions* $\partial U_i(\cdot)/\partial x_i < (>) 0$ for $x_i > (<) 0$, and $\partial U_j(\cdot)/\partial x_i > 0$, $\forall i, j \neq i$, so that the x_i's can be interpreted as additional contributions to a public good *above* the individually rational level. Consequently, the trivial Nash equilibrium in the one-stage game in which every i sets x_i noncooperatively is normalized at $\mathbf{x} = \mathbf{0}$. Appendix A proves the main result, Theorem 3, and a generalization of Theorem 2 *without* these additional assumptions. The additional assumptions are not needed for Theorems 0 and 1.

The set of $\mathbf{x}$'s such that everyone obtains a strictly positive payoff is the open set

$$\mho_{\mathbf{x}} \equiv \{\mathbf{x} \in \mathbb{R}^n \colon U_i(\mathbf{x}) > 0 \ \forall i\}, \text{ and}$$
$$\mho_{\mathbf{U}} \equiv \mathbf{U}(\mho_{\mathbf{x}}) \equiv \{\mathbf{U} \in \mathbb{R}^{\mathbf{n}} : \exists \mathbf{x} \in \mho_{\mathbf{x}} \text{ s.t. } U_i(\mathbf{x}) = U_i \forall i \in N\}.$$

I will assume that the set $\mho_{\mathbf{U}}$ is bounded and convex.

Example E. Suppose $n = 2$ and

$$U_i(\mathbf{x}) = x_j - x_i^2/2, \text{ where } j \in N \backslash i. \tag{E}$$

The set $\mho_{\mathbf{x}}$ is shaded in the left panel of Figure 1, while $\mho_{\mathbf{U}}$ is in the right panel.

[9] This is a normalization in the following sense: If the contributions were $\widetilde{\mathbf{x}} = (\widetilde{x}_1, ..., \widetilde{x}_n) \in \mathbb{R}^n$, with payoffs payoff $\widetilde{U}_i(\widetilde{\mathbf{x}})$ and default outcome $\widetilde{\mathbf{x}}^d$, then we can define $x_i \equiv \widetilde{x}_i - \widetilde{x}_i^d$ and $U_i(\mathbf{x}) \equiv \widetilde{U}_i(\widetilde{\mathbf{x}}^d + \mathbf{x}) - \widetilde{U}_i(\widetilde{\mathbf{x}}^d) = \widetilde{U}_i(\widetilde{\mathbf{x}}) - \widetilde{U}_i(\widetilde{\mathbf{x}}^d)$. It follows that the default is $\mathbf{x} = 0 \Rightarrow U_i(\mathbf{0}) = 0$.

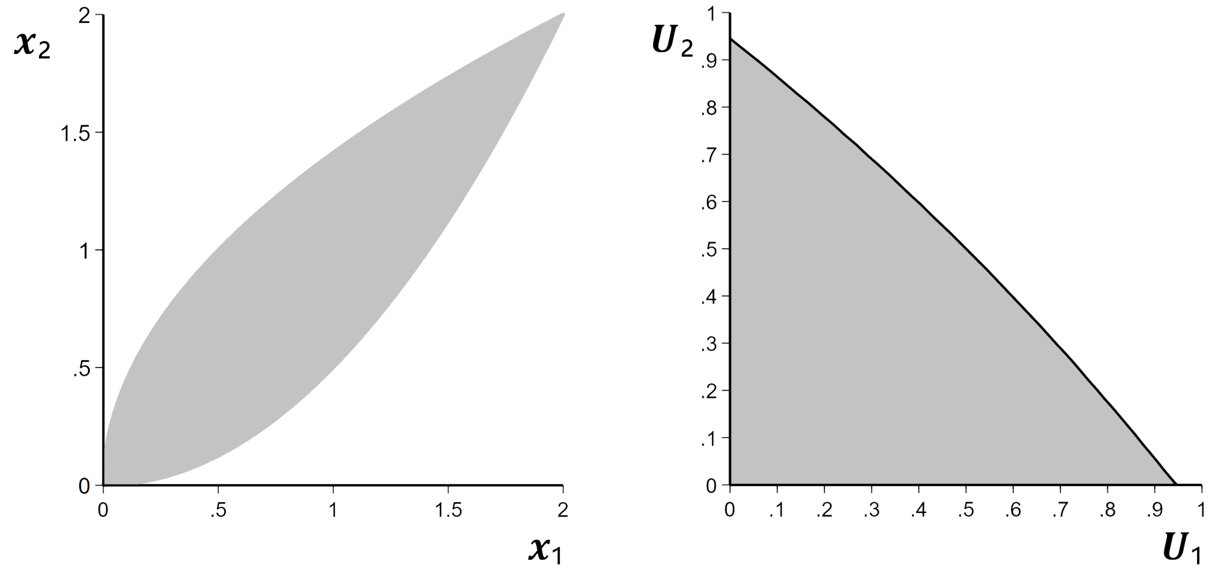

Figure 1: *For Example E, the left panel illustrates the open set* $\mho_{\mathbf{x}}$ *of pairs* (x_1, x_2) *s.t.* $U_1 > 0$ *and* $U_2 > 0$*. The right panel illustrates the corresponding set of utility pairs,* $\mho_{\mathbf{U}}$*.*

The bargaining game starts when every party i simultaneously proposes its own dimension, or contribution, $x_i \in \mathbb{R}$. After they observe $\mathbf{x} = (x_1, ..., x_n)$, each party must decide whether to accept. (It will not matter whether the acceptance decisions are simultaneous or not.) If everyone accepts, every $i \in N$ receives payoff $U_i(\mathbf{x})$ and the game ends. If one or more parties decline $\mathbf{x}$, the game continues in the following period where the players interact again in the same way. An indifferent party is assumed to accept.[10]

The lag between one acceptance stage and the next acceptance stage is $\Delta > 0$. With continuous-time discount rate $r_j > 0$, the discount factor between time t and $t + \Delta$ is

$$e^{-r_j\Delta} \approx 1 - r_j\Delta \Leftrightarrow r_j \approx \rho_j \equiv \left(1 - e^{-r_j\Delta}\right)/\Delta,$$

where the approximation holds when $\Delta \to 0$. Although I will not require Δ to be small, it will be convenient to refer to $\rho_j \equiv \left(1 - e^{-r_j\Delta}\right)/\Delta$ as the discount rate.

Thus, if party j declines an offer and expects the outcome $\mathbf{x}^*$ in the next period, then j's present-discounted payoff is $(1 - \rho_j\Delta) U_j(\mathbf{x}^*)$. Anticipating $\mathbf{x}^*$ and $U_j(\mathbf{x}^*) > 0$, j rejects $\mathbf{x}$ now if and only if:

$$U_j(\mathbf{x}) < (1 - \rho_j\Delta) U_j(\mathbf{x}^*) \Leftrightarrow \frac{U_j(\mathbf{x}^*) - U_j(\mathbf{x})}{\rho_j\Delta U_j(\mathbf{x}^*)} > 1. \tag{1}$$

[10] This assumption rules out uninteresting equilibria in which everyone rejects everything because no-one is pivotal.

3.2. A benchmark result

As I will explain in Section 4.1, there is typically a large number of subgame-perfect equilibria in games with infinite time horizon. Section 4.2 follows much of the literature by considering stationarity as a refinement. To appreciate the result in that section, consider the stationary subgame-perfect equilibria (SSPEs) in the game developed so far.

There clearly exists a "trivial" SSPE consisting of the acceptance strategies (1) and a vector $\mathbf{x}^* = \mathbf{0}$, so that the payoffs are $U_j(\mathbf{x}^*) = 0 \forall j$. If this outcome is always expected, there is no reason for any individual party to offer anything else. Unfortunately, no $\mathbf{x} \in \mho_{\mathbf{x}}$ or, equivalently, $\mathbf{U} \equiv (U_1, ..., U_n) \in \mho_{\mathbf{U}}$, can be supported as an SSPE outcome: For any equilibrium candidate in which $U_j(\mathbf{x}^*) > 0 \forall j$, contributing party i can suggest x_i slightly different from x_i^* without satisfying (1). Thus, x_i^* must coincide with i's preferred level, $x_i^* = \arg\max_{x_i} U_i\left(x_i, \mathbf{x}_{-i}^*\right)$, which is zero under the above additional assumptions.[11]

Theorem 0. *There is no SSPE with* $\mathbf{x} \in \mho_{\mathbf{x}}$ *or payoffs* $\mathbf{U} \in \mho_{\mathbf{U}}$.

3.3. Relaxing the "no uncertainty" assumption

From (1), we obtain that j rejects $\mathbf{x}$, when $\mathbf{x}^*$ can be expected in the next period, with a probability, $F_j(\cdot)$, that is either 0 or 1:

$$F_j\left(\frac{U_j(\mathbf{x}^*) - U_j(\mathbf{x})}{\rho_j \Delta U_j(\mathbf{x}^*)}\right) = \left\{\begin{array}{l} 1 \text{ if } \frac{U_j(\mathbf{x}^*) - U_j(\mathbf{x})}{\rho_j \Delta U_j(\mathbf{x}^*)} > 1 \\ 0 \text{ if } \frac{U_j(\mathbf{x}^*) - U_j(\mathbf{x})}{\rho_j \Delta U_j(\mathbf{x}^*)} \leq 1 \end{array}\right\} \in \{0, 1\}. \tag{2}$$

In reality, the parties cannot be certain of what opponents will accept. Therefore, Bastianello and LiCalzi (2019:837), in their probability-based interpretation of the NBS, "introduce uncertainty over which alternatives bargainers are willing to accept."

For similar reasons, I henceforth assume F_j to be a continuous function for which $F_j(0) = 0$, while $F_j > 0$ if and only if its argument is strictly positive. In other words, j certainly accepts the allocation that is expected in the next period, $\mathbf{x}^*$, but there is always a chance that j declines $x_i < x_i^*$.

[11] There can be other equilibria in the game than the trivial equilibrium $\mathbf{x}^* = \mathbf{U} = \mathbf{0}$. With the additional assumptions, every $\mathbf{x}$ such that $U_i(\mathbf{x}) = 0$ for at least two parties can be supported as an SSPE, but no other $\mathbf{x}$ can be supported as an SSPE. In Example E, there is an equilibrium in which $\mathbf{x} = (2, 2)$ and both payoffs are zero: if party i reduces x_i, then U_j turns negative and j rejects.

Note that if we define a shock $\theta_{j,t}$ to be distributed as F_j, i.i.d. over time, then we can equivalently say that j rejects $\mathbf{x}$ if and only if

$$\frac{U_j(\mathbf{x}^*) - U_j(\mathbf{x})}{\rho_j \Delta U_j(\mathbf{x}^*)} > \theta_{j,t}, \tag{3}$$

since this event arises with probability $\Pr\left(\theta_{j,t} < \frac{U_j(\mathbf{x}^*) - U_j(\mathbf{x})}{\rho_j \Delta U_j(\mathbf{x}^*)}\right) = F_j\left(\frac{U_j(\mathbf{x}^*) - U_j(\mathbf{x})}{\rho_j \Delta U_j(\mathbf{x}^*)}\right)$.

A microfoundation: It is worth mentioning that this uncertainty can be derived from shocks in the utility functions, from the subjective beliefs over the delay following rejections, or from the impatience. The results do not hinge on any particular form of uncertainty, and Appendix B shows that all the mentioned uncertainties can generate similar results.

In bargaining, uncertainty was first introduced though the discount rate (Rubinstein, 1985).[12] After all, estimates of discount rates "differ dramatically across studies, and within studies across individuals. There is no convergence toward an agreed-on or unique rate of impatience" (Gollier and Zeckhauser, 2005:879). There are conflicting views on what the discount rate ought to be (Arrow et al., 2014), how it varies with the time horizon (Frederick et al., 2002), across individuals (Andersen et al., 2008), how it should be aggregated (Chambers and Echenique, 2018), and what form it takes: Consider the cases for hyperbolic (Angeletos et al., 2001), quasi-hyperbolic (Laibson, 1997), beta (Dietz et al., 2018), or gamma discounting (Weitzman, 2001). The discount rate can be smaller when decisions are collective (Jackson and Yariv, 2014; Adams et al., 2014) or influence others (for theory and evidence, see Dreber et al., 2016; Rong et al., 2019). Ramsey (1928) argued the discount rate should simply be zero.

The discount rate can also be viewed as the Poisson rate of a bargaining breakdown. Different parties may have fluctuating opinions regarding the level of this rate.

Given these controversies and debates, it seems unreasonable to assume the discount rate to be common, deterministic, and known for every future period.

In international negotiations, it is reasonable that a policymaker's tolerance for delay

[12] See also Watson (1998) and Abreu et al. (2015). While I follow these scholars by letting the discount rate be stochastic, our approaches are complementary in that they consider persistent shocks while I consider temporary shocks.

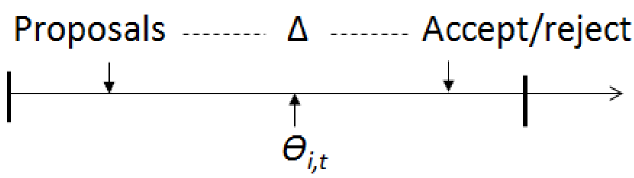

Figure 2: *Proposals are made before the shocks are observed.*

is influenced by a number of (con)temporary domestic policy or economic issues that compete for the policymaker's attention. The impatience may also depend on today's probability of remaining in office (Ortner, 2017; Harstad, 2020). Since no one can foresee all these issues when the pledges are made, perhaps several months in advance, Figure 2 illustrates how the shocks may be realized and observed by everyone after the offers but before acceptance decisions are made. This timing seems quite reasonable.[13]

Formally, write the discount rate as $\rho_{i,t} = \theta_{i,t}\rho_i$, where $\rho_i \equiv \mathrm{E}\rho_{i,t}$ is the expected discount rate of $i \in N$, so that $\theta_{i,t}$ captures a shock with mean 1. When $\theta_{i,t}\rho_i$ replaces the discount rate in (1), we obtain that if $U_j(\mathbf{x}^*) > 0$, then $j \in N \backslash i$ rejects $\mathbf{x}$, after learning $\theta_{j,t}$, if and only if (3) holds:

$$U_j(\mathbf{x}) < (1 - \theta_{j,t}\rho_j\Delta)\, U_j(\mathbf{x}^*) \Leftrightarrow \ (3).$$

The shocks are assumed to be distributed according to a continuous probability density function (pdf) $f(\theta_{1,t}, ..., \theta_{n,t}) \in (0, \infty)$ on support $\prod_{i \in N} \left[0, \overline{\theta}_i\right]$. The marginal distribution of $\theta_{i,t}$ is $f_i(\theta_{i,t}) \equiv \int_{\Theta_{-i}} f(\theta_{1,t}, ..., \theta_{n,t})\, d\Theta_{-i}$, where $\Theta_{-i} \equiv \prod_{j \neq i} \left[0, \overline{\theta}_j\right]$.

If the shocks were correlated over time, the game would be nonstationary and there could be delay on the equilibrium path when one period's shocks indicated that the pledges could be more attractive in the future. If, in addition, $\theta_{i,t}$ were privately observed by i, then i might reject to signal a small $\mathrm{E}\theta_{i,t+1}$. These issues are interesting, but they are partly analyzed already (see, for instance, Chen and Eraslan, 2014) and they are orthogonal to the results that are novel in this paper. To isolate them, I assume that the shocks are i.i.d. at each time t. In the real world, after learning about one another's time

[13] Because there can be a substantial lag between offers and acceptance decisions, it is natural that policymakers in the meantime learn about how urgent it is for them to conclude the negotiations, or about the attention they instead must give to other policy and economic issues.

preference has converged, some uncertainty will remain and it is this uncertainty that is captured by f.

4. The pledge-and-review bargaining solution

4.1. A folk theorem

There are often many SPEs in games with infinite time horizon.

Theorem 1. *There exists* $\underline{\Delta} \in (0, \infty)$ *such that for every* $\Delta \in (0, \underline{\Delta}]$, *every outcome* $\mathbf{x} \in \mho_{\mathbf{x}}$ *and payoffs* $\mathbf{U} \in \mho_{\mathbf{U}}$ *can be supported as an SPE.*

The additional assumptions are not needed for this result. To support any $\mathbf{U}^* \in \mho_{\mathbf{U}}$ as an SPE, the proof (in Appendix A) considers the possibility that if i deviates, then the continuation payoff vector is $\mathbf{U}^i$, where $U_j^i = k_j^i U_j^*$ with $k_i^i \in (0, 1)$ and $k_j^i = 1$, $j \neq i$. (I assume there is free disposal, so that if $\mathbf{U}^* \in \mho_{\mathbf{U}}$, then $\mathbf{U}^i \in \mho_{\mathbf{U}}$.) The idea is that if i deviates, then the parties "punish" i by switching from $\mathbf{U}^*$ to $\mathbf{U}^i$.

4.2. Stationary equilibria: Justifications

There are several reasons for why we may want to study the stationary equilibria in this game. First, the set of equilibria permitted by folk theorems is too large to make sharp predictions. Second, history-contingent strategies, as those permitted above, may not be renegotiation proof and, third, they can be quite complicated to coordinate on. Baron and Kalai (1993:292) explain that "simplicity is likely to be a major consideration ... when an equilibrium is being selected." For the bargaining game by Baron and Ferejohn (1989), they define and prove that there is a unique "simplest equilibrium" – namely the stationary one.

Bhaskar et al. (2013:925) write that arguments for focusing on stationary Markov equilibria "include (i) their simplicity; (ii) their sharp predictions; (iii) their role in highlighting the key payoff relevant dynamic incentives; and (iv) their descriptive accuracy in settings where the coordination implicit in payoff irrelevant history dependence does not seem to occur." They prove that yet another foundation for stationarity arises when social memory is bounded.

These arguments are especially relevant for negotiations among political representatives. Bowen et al. (2014:2947) explain that imposing stationarity "is reasonable in dynamic political economy models where there is turnover within parties since stationary Markov equilibria are simple and do not require coordination." This argument is particularly relevant for the climate negotiations that inspire the present game: Because a country's chief negotiator is often and frequently replaced, it is quite attractive to rely on strategies that can be played by new representatives who may not remember the history of the game.

Finally, Bhaskar et al. (2013:926) note that "the common assumption ... that players are infinitely lived is an approximation." Games with finite time horizons can be solved by backward induction and, then, stationarity can be derived instead of assumed: See the below discussion in Section 6, and the proof in Appendix B.

For all these reasons, I will now search for stationary subgame-perfect equilibria (SSPEs) in pure strategies.

4.3. Stationary equilibria: Characterization

An SSPE is a vector, $\mathbf{x}^*$, combined with a set of strategies for the acceptance stage.

A characterization of the SSPEs is especially interesting in light of Theorem 0, stating that no $\mathbf{U} \in \mho_{\mathbf{U}}$ can be supported as an SSPE in the game without uncertainty. With the shocks introduced in Section 3.3, there is no $x_i < x_i^*$ that is entirely "safe" in that it will be accepted with probability one. A deviating party may always face some risk.

As derived already, the optimal acceptance strategies are given by (3). Since $\theta_{j,t}$ is drawn from a continuous distribution, the probability that j accepts will be continuous in x_i. On the one hand, this continuity can motivate positive contributions: $\mathbf{x}^* \in \mho_{\mathbf{x}}$ can be supported as a "nontrivial" SSPE if the marginal benefit for i, by slightly reducing x_i, is outweighed by the risk that at least one party might be sufficiently patient to decline the offer and wait for $\mathbf{x}^*$. On the other hand, the punishment for trying to get $x_i < x_i^*$ accepted is simply the risk of delay. (In contrast, Section 4.1., which considered SPEs, permitted the parties to move to another equilibrium outcome if i deviated.) Party i may thus be quite tempted to take *some* risk and reduce x_i, especially when x_i^* is large and costly to i. This temptation will limit how large the equilibrium x_i^* can be.

Note that there cannot be delay on the equilibrium path: If i finds it optimal to offer less than what j would prefer today, i will find this to be optimal later, as well. After all, opponents cannot gain from rejecting a stationary offer. An equilibrium offer will thus *not* be risky at the equilibrium path: $\mathbf{x}^*$ will be proposed and (3) implies that, as a result, the equilibrium proposal will be accepted without delay with probability 1. The temptation to take (further) risks is merely generating an upper boundary for how large x_i^* can be, as shown in the following theorem.

Theorem 2. *Consider an SSPE with $\mathbf{x}^* \in \mho_{\mathbf{x}}$ and $\mathbf{U} \in \mho_{\mathbf{U}}$. For every $i \in N$:*

$$x_i^* \leq x_i^\circ \text{ if } x_i^\circ = \arg\max_{x_i} \prod_{j \in N} \left(U_j\left(x_i, \mathbf{x}_{-i}^*\right)\right)^{w_j^i}, \text{ where} \tag{4}$$

$$\frac{w_j^i}{w_i^i} = \frac{\rho_i}{\rho_j} \cdot f_j(0) \cdot \mathrm{E}\left(\theta_{i,t} \mid \theta_{j,t} = 0\right), \forall j \neq i.$$

The upper boundary on x_i^* has a remarkably simple characterization: When (4) binds, x_i^* maximizes an asymmetric Nash product, where the payoff of every party j is associated with some weight, w_j^i. The theorem endogenizes the weights and shows how they depend on three factors.[14]

First, the weight on j's utility is larger if j is expected to be patient relative to i. This is intuitive (and in line with existing papers on the Nash program, mentioned in Footnote 3): When j is patient, j is more tempted to reject an offer that is worse than what one can expect in the next period, and thus i finds it too risky to reduce x_i, especially when i is likely to be impatient.

Second, the weight on j's payoff is larger when there is more uncertainty regarding j's shock. The intuition is that when it is uncertain whether j will accept, then i is willing to offer more in order to reduce the risk of delay. As the shocks vanish, however, the equilibrium payoff set converges to the origin. That is, if $f_j(0) \to 0$, or if the shock $\theta_{j,t}$

[14] Theorem 2 endogenizes only the relative weights, w_j^i / w_i^i, but this is sufficient since $\arg\max_{x_i} \prod_{j \in N} \left(U_j\left(x_i, \mathbf{x}_{-i}^*\right)\right)^{w_j^i}$ stays unchanged if every weight w_j^i is multiplied by the same positive number.

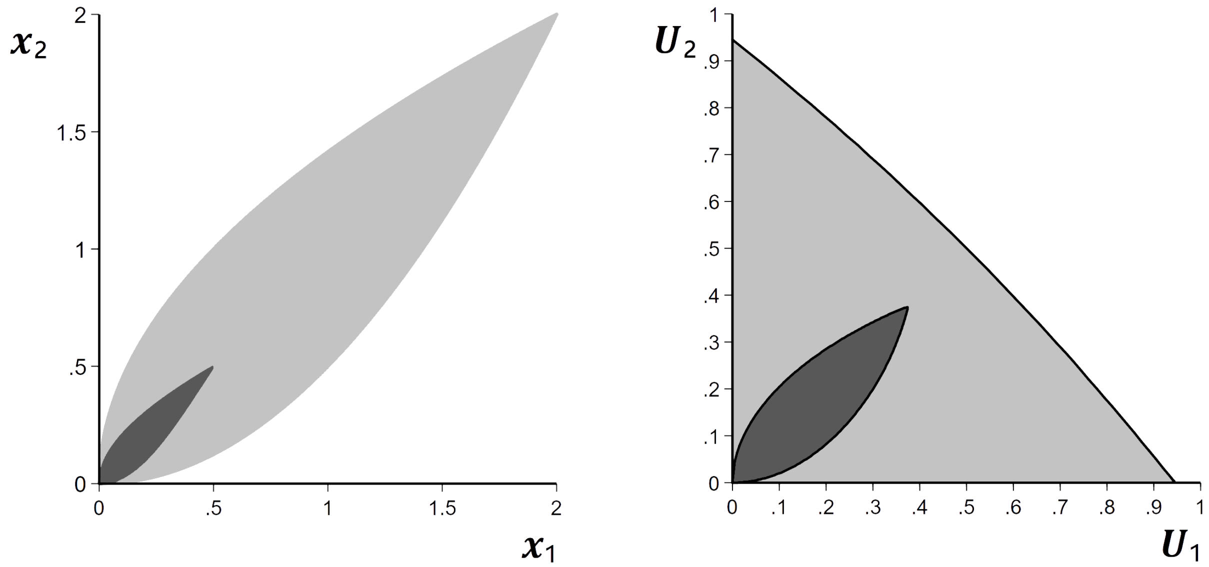

Figure 3: *For Example E, the dark area in the left panel illustrates the set of SSPE pairs* (x_1, x_2)*, while the dark area in the right panel illustrates the corresponding set of utilities.*

were bounded away from zero, then $w_j^i \to 0$, and (4) converges to the trivial equilibrium $\mathbf{x}^* = \mathbf{0}$. Appendix B shows how a version of the results survives even if $f_j(0) \to 0$.

Third, the weight on j's payoff is small in the presence of a small $\mathrm{E}(\theta_{i,t} \mid \theta_{j,t} = 0)$, which measures i's expected shock given that j's $\theta_{j,t}$ is small. The intuition is that if i can be expected to have a small discount rate exactly when j has a small discount rate, then it matters less that j declines an offer in this circumstance. When the delay matters less, i does not find it necessary to offer a lot. This result predicts that a party i may pay less attention to the payoffs of those who face shocks that are positively correlated with i's shock.

Interestingly, the set of SSPEs does not depend on the level of Δ or on any requirement that Δ is small (for a fixed discount rate). The intuition is that a larger Δ is increasing j cost of rejecting (and delaying) any given offer by the same amount as it is increasing i's cost of making an unattractive offer.

As a comparison to Theorem 2, in the asymmetric NBS, each x_i maximizes the *same* asymmetric Nash product:

$$x_i^A = \arg\max_{x_i} \prod_{j \in N} \left(U_j\left(x_i, \mathbf{x}_{-i}^A\right)\right)^{w_j}, \tag{5}$$

for some fixed weights, $(w_1, ..., w_n)$. In this case, the vector $\mathbf{x}^A$ will be Pareto optimal.

Also when (4) binds, the equilibrium x_i^* maximizes an asymmetric Nash product, but, in stark contrast to the asymmetric NBS, with P&R *different parties* apply *different weights* (f.ex., $w_j^i/w_i^i \neq w_j^j/w_i^j$). The vector $\mathbf{x}^*$ is, for that reason, not Pareto optimal. In particular, if $w_j^i/w_i^i < 1$ for every (i,j), $j \neq i$, then it is possible to make every party better off by increasing all contributions relative to $\mathbf{x}^*$.

The dark region in Figure 3 illustrates the set of equilibria permitted by Theorem 2 when $w_j^i/w_i^i = w = \frac{1}{2} \; \forall (i,j), j \neq i$, in Example E. The multiplicity of SSPEs arises from the inequality in (4). The logic leading up to Theorem 2 limits how large the x_i's can be, but not how *small* the pledges can be. After all, there is no point for i to contribute more than the equilibrium quantity, whatever the equilibrium is. (As noted, j always accepts an SSPE vector given that $\theta_{j,t} \geq 0$ and $U_j(\mathbf{x}^*) \geq 0$.)

4.4. Locally perfect equilibria

There are two reasons for why we may want to refine the set of equilibria further.

First, the multiplicity of equilibria makes it difficult to establish sharp predictions.

Second, some of the equilibria permitted by Theorem 2 are not very robust. To see this, note that when x_i^* is so small that (4) is nonbinding, then i is *not* indifferent to a marginal reduction in x_i, relative to x_i^*. A marginal reduction is strictly worse for i, because of the risks that are involved. Thus, in the presence of small trembles, where not even $\mathbf{x}^*$ is guaranteed acceptance, i might prefer to raise x_i slightly above x_i^* to reduce the risk. With trembles, party i benefits from increasing x_i^* as long as (4) is nonbinding. This is the intuition for why a trembling-hand perfect SSPE will require (4) to hold with equality.

Selten (1975:35) argued that "a satisfactory interpretation of equilibrium points in extensive games seems to require that the possibility of mistakes is not completely excluded." With this reasoning, Selten introduced trembling-hand perfection in finite games. When the action space is continuous, Myerson (1978) argued that the trembles should be smaller for costlier errors. This reasoning is captured by the notion of "local perfection," defined by Simon (1987). The following definition of local perfection is a simplification of the definition provided by Simon (1987).[15]

Definition of Local Perfection: *Consider a perturbed game in which, when the vector of submitted offers is* $\mathbf{x}$, *then* $\mathbf{x} + \epsilon s_t$ *is realized and observed, where* s_t *is a vector of* n *trembles distributed i.i.d. over time, with bounded support, and with strictly positive density on a neighborhood of* $\mathbf{0}$. $\mathbf{x}^*$ *is* a locally perfect equilibrium *if* $x_i^* = \lim_{\epsilon \to 0} x_i^*(\epsilon) \, \forall i \in N$, *where* $\mathbf{x}^*(\epsilon)$ *is an equilibrium of the perturbed game.*

For our purposes, *equilibrium* refers to an SSPE.[16]

Theorem 3. *Consider a locally perfect SSPE. Inequality (4) binds for every* $i \in N$:

$$x_i^* = \arg\max_{x_i} \prod_{j \in N} \left(U_j\left(x_i, \mathbf{x}_{-i}^*\right)\right)^{w_j^i}, \text{ where} \tag{6}$$
$$\frac{w_j^i}{w_i^i} = \frac{\rho_i}{\rho_j} \cdot f_j(0) \cdot \mathrm{E}\left(\theta_{i,t} \mid \theta_{j,t} = 0\right), \forall j \in N \backslash i.$$

The condition is necessary; the theorem claims neither sufficiency nor uniqueness. Nevertheless, local perfection allows us to make sharper predictions and to justify the emphasis on the weights, w_j^i / w_i^i, and what they depend on, and where the intuition for the terms are discussed in Section 4.3.

The intuition for Theorem 3 is as described at the beginning of this subsection: With trembles, party i is not confident that x_i^* will be approved and thus i finds it beneficial to raise x_i as long as x_i^* is small.[17]

Although the equilibrium is not Pareto optimal, it is interesting to note that uncertainty is beneficial for the parties in two ways in this model. First, it is the presence of the $\theta_{i,t}$'s that motivates the parties to pledge sufficiently much so that everyone can be strictly better off relative to the default outcome. Second, of all the SSPEs permitted by Theorem 2, trembles rule out the SSPEs with the smallest contributions, that is, contributions that

[15] I am grateful to Leo Simon for discussions on how the definitions relate. Carlsson (1991) and Simon and Stinchcombe (1995) are also motivated by the need to generalize trembling-hand perfection to infinite games.

[16] The definition follows Selten (1975) in that the trembles are uncorrelated over time (for further justifications on this, see the final paragraph in Section 3). As with the $\theta_{i,t}$'s, allowing the trembles to be correlated across parties comes at no costs for the analysis.

[17] Chatterjee and Samuelson (1990) show that there continue to be multiple perfect equilibria in the dynamic version of the NDG unless the stationarity assumption is maintained.

are so small that (*4*) is nonbinding.[18]

For Example E, Theorem 3 predicts the top-right corners in the dark-grey regions in Figure 3. One can show that this point Pareto dominates all other SSPEs in Example 3 if $w < \sqrt{3} - 1 \approx 0.73$, as is assumed in Figure 3. Thus, focusing on equilibria that are not Pareto dominated might in some cases replace the restriction to local perfection.[19]

4.5. Simplifications and corollaries

Theorem 3 has several important consequences: It describes how the equilibrium is influenced by the different parties' utility functions, mean discount rates, shock distributions, and the correlation of the shocks (i.e., the $\theta_{i,t}$'s). We can learn still more from the theorem if we simplify to special cases.

Corollary 1. *Suppose all parties share the same mean discount rate and shocks are independent.*
(i) If f_i is single-peaked, then $w_i^j/w_j^j < 1$, $\forall i \in N, j \in N \backslash i$.
(ii) If f_i is single-peaked and symmetric, then $w_i^j/w_j^j \leq \frac{1}{2}$, $\forall i \in N, j \in N \backslash i$.
(iii) If f_i is constant (uniform), then $w_i^j/w_j^j = \frac{1}{2}$, $\forall i \in N, j \in N \backslash i$.

Intuitively, the corollary illustrates that each party is likely to weight the value of others' payoffs less than the party weights its own payoff. Technically, the three parts of the corollary follow from the formula for the weights combined with the fact that, for a pdf, $\int_0^{\overline{\theta}_i} f_i(\theta_{i,t})\, d\theta_{i,t} = 1$.[20]

If the shock correlations and payoff functions are the same for all parties, then the characterization can be simplified even further.

Corollary 2. *Suppose all parties have the same payoff functions and marginal shock distributions, f_i, so that $w_j^i/w_i^i = f_j(0) \cdot \mathrm{E}(\theta_{i,t} \mid \theta_{j,t} = 0) = w$ for all $i \in N$, $j \in N \backslash i$. In a*

[18] As an additional benefit of uncertainty, note that if we also introduce perturbations in the accept-vs-reject decision, then we no longer need to *assume* that each party votes as if pivotal (see Footnote 10), since that will be part of the optimal strategy.

[19] I thank Asher Wolinsky for making this observation. Section 6 explains why local perfection can also be replaced by trembles in the support of the shocks.

[20] To see part (ii), for example, note that if $f_i(0) > 1/2$, then, when $f_i(\cdot)$ is single-peaked and symmetric around the mean of one, $\int_0^2 f_i(\theta_{i,t})\, d\theta_{i,t} > 1$, violating the definition of a pdf. If the shocks are not correlated, then $E(\theta_{i,t} \mid \theta_{j,t} = 0) = 1$.

symmetric locally perfect SSPE, the equilibrium offers can be written as:

$$x_i^* = \arg\max_{x_i}[U_i\left(x_i, \mathbf{x}_{-i}^*\right) + w\sum_{j\neq i} U_j\left(x_i, \mathbf{x}_{-i}^*\right)]. \tag{7}$$

It is straightforward to check that the first-order condition of (7) coincides with the first-order condition of (6) when $w_j^i/w_i^i = w$ and $U_i(\mathbf{x}^*) = U_j(\mathbf{x}^*)$ for every $i, j \in N$. In Example E, we simply get $x_i = w \forall i$.

If $n > 2$ in Example E, we get $x_i = (n-1)w$. The fact that contributions are increasing in n holds more generally: This can be seen from Theorem 3, as well, under the additional assumptions, so that j benefits when $i \neq j$ contributes.

Corollary 3. *Equilibrium contributions are larger when n is large.*

The intuition for this result is that when n is large, it is more likely that at least one of the other parties will decline $x_i < x_i^*$. The larger risk motivates i to contribute more.

Note that the intuition for why a larger n raises contributions in standard conditional-offer bargaining games is remarkably different: In those games, i is willing to contribute more because it can then, simultaneously, ask other parties to contribute more.

5. Relationship to Nash's demand game and bargaining solution

The P&R bargaining outcome is in stark contrast to the Nash bargaining solution, predicting that the x_i's would follow from (5) with $w_j^i/w_i^i = 1\forall(i,j) \in N^2$. The NBS is frequently used to describe multilateral bargaining outcomes partly because the NBS results from noncooperative bargaining games. Nash (1953) introduced his "demand game" (NDG) exactly because he could show that it implemented the NBS. Despite this contrast, Nash's result can be derived from Theorem 3 because the NDG can be shown to be a limiting case of the P&R bargaining game.[21]

In the NDG, each player is demanding an ex post payoff level or, equivalently, a variable (x_i) that dictates i's ex post demanded payoff, $d_i(x_i)$. The vector of demands is feasible with probability $p(\mathbf{x})$. If the vector is not feasible, everyone receives zero. Party

[21] I am grateful to Jean Tirole for the motivation for this subsection.

i's expected utility is:

$$U_i(x_i, \mathbf{x}_{-i}) = d_i(x_i)\, p(\mathbf{x}). \tag{8}$$

This utility function is permitted in the above analyses if the d_i's and p are continuous functions. As in Nash (1953:132), the continuity of p "should be thought of as representing the probability of the compatibility of the demands d_1 and d_2. It can be thought of as representing uncertainties in the information structure of the game, the utility scales, etc." Note that this uncertainty comes on the top of the shocks (the $\theta_{i,t}$'s) and the trembles (the s_t's) considered in Section 4. In the special case of (8), (6) can be rewritten as follows.

Theorem 4. *Consider pledge-and-review bargaining and suppose i's expected utility is given by (8). If $\mathbf{x}^*$ is a locally perfect SSPE, then*:

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} \prod_{i\in N} d_i(x_i)^{\varrho_i}\, p(\mathbf{x})^{\varpi} \quad \textit{and} \tag{9}$$

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} \prod_{i\in N} d_i(x_i)^{\varrho_i} \quad \textit{s.t. } p(\mathbf{x}) = p(\mathbf{x}^*), \textit{ where} \tag{10}$$

$$\varrho_i = \frac{w_i^i / \sum_{j\in N} w_j^i}{\sum_{k\in N}\left(w_k^k / \sum_{j\in N} w_j^k\right)} \quad \textit{and } \varpi = \frac{1}{\sum_{k\in N}\left(w_k^k / \sum_{j\in N} w_j^k\right)}. \tag{11}$$

Note that if the parties face the same distribution of the discount rates, then $w_i^i / \sum_{j\in N} w_j^i$ is the same for every $i \in N$, and, therefore, $\varrho_i = 1 \forall i$. Note also that if $f_i(0)$ approaches or equals 0 for every i, then $w_i^i / \sum_{j\in N} w_j^i = \varrho_i = 1 \forall i$. In both cases, $\mathbf{x}^*$ coincides with the NBS.

Corollary 4. *Suppose all parties face identical expected discount rates and shock distributions, or that $f_i(0) \to 0 \forall i \in N$. In either case, $\varrho_i = 1 \forall i \in N$, and $\mathbf{x}^*$ implements the NBS:*

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} \prod_{i\in N} d_i(x_i) \quad \text{s.t.} \quad p(\mathbf{x}) = p(\mathbf{x}^*).$$

The condition $p(\mathbf{x}) = p(\mathbf{x}^*)$ fixes the total risk. If the uncertainty on the feasibility constraint vanishes, in the sense that $p(\mathbf{x})$ is close to 0 or 1 for almost every $\mathbf{x}$, then it is intuitive that $\mathbf{x}^*$ must be close to an $\mathbf{x}$ that ensures $p(\mathbf{x}) \approx 1$. In this case, the constraint $p(\mathbf{x}) = p(\mathbf{x}^*)$ simply requires $\mathbf{x}$ to be feasible (see Binmore, 1987).

The result (by Nash, 1953) that the NDG implements the NBS is generalized by Corollary 4 in several respects, since the corollary builds on the P&R bargaining model:

(i) According to Corollary 4, the mapping from the NDG to the NBS continues to hold if, as with P&R, any party can veto the allocation $\mathbf{x}$ after which there will be a finite delay before the demand game can be played again.

(ii) There can be $n \geq 2$ parties, and not only 2 as in Nash (1953) and in the dynamic version analyzed by Chatterjee and Samuelson (1990).

(iii) The parties can have stochastic discount rates. This uncertainty influences w, but both the uncertainty and the common w are irrelevant for the mapping to the NBS if the parties are symmetric. The intuition for why w is irrelevant is that, given the sharp feasibility constraint characterized by $p(\mathbf{x})$ when uncertainty vanishes, i's preferred x_i coincides with the efficient level, given the other x_j's.

(iv) When the weights (w_j^i/w_i^i) are heterogeneous, (10) shows that $\mathbf{x}^*$ characterizes an *asymmetric* NBS: The bargaining power index (ϱ_i) is larger for those parties who are likely to be patient or who face less uncertainty regarding the opponents' discount rates. This finding is consistent with analyses of alternating-offer bargaining games (see Footnote 3); Theorem 4 shows that it holds also in this generalization of the NDG.

(v) Theorem 4 also uncovers the limitation of the mapping from the NDG to the NBS. When the uncertainty on the feasibility constraint vanishes, U_i becomes discontinuous in x_j, technically violating the assumption in Section 3.1. If instead each $U_i(\mathbf{x})$ is continuous in every x_j, as when the uncertainty on the feasibility constraint is *not* vanishing, then Theorem 4 shows that $\mathbf{x}^*$ is technically different from the NBS, thanks to the condition $p(\mathbf{x}) = p(\mathbf{x}^*)$ in eq. (10). Eq. (9) shows that i places less weight on the (collective) risk if w_j^i/w_i^i is small for every $j \neq i$ because, then, ϖ is small, as well, according to (11). Therefore, when the feasibility of $\mathbf{x}$ is uncertain, as reflected by the continuous function p, then each party i takes too much risk in that i sets x_i, and demands $d_i(x_i)$, without internalizing that the risk may be large for everyone.

(vi) More generally, the NDG, leading to the expected payoff (8), is only one of many cases permitted in the analysis of Section 4. If the parties do not demand utility levels, but pledge contribution levels, as in Example E, then $U_i(\mathbf{x})$ is likely to be continuous in all the x_j's. This continuity makes the P&R outcome inefficient in that each x_i^* maximizes its own asymmetric Nash product, as described by Theorem 3.

6. Robustness and generalizations

To proceed with the analysis above in a tractable and pedagogical way, several assumptions were introduced. This section contains a brief discussion of how some of them can be relaxed (some details are available in the Appendices; further details on request).

1. Relaxing stationarity: In the model with uncertainty, Theorems 2-4 continue to hold if, instead of restricting attention to stationary SPEs, there is a finite time horizon, $T < \infty$, $T \to \infty$, and there is a terminal outcome $\mathbf{x}^T$, interpreted as the outcome that will be implemented unless the parties complete the negotiations before the time expires. In this case, the set of SPEs can be derived with backward induction.

Theorem 5. *Suppose $T - t < \infty$.*
(i) Consider a unique SPE with $\mathbf{x}_t^*$ and $\lim_{T-t\to\infty} \mathbf{x}_t^* = \mathbf{x}^*$. Then, (4) holds, $\forall i \in N$.
(ii) Suppose the equilibrium in (i) is locally perfect, as well. Then, (4) binds, $\forall i \in N$.

This result states that if, when we solve the game by backward induction, the SPE converges, then it converges to a stationary equilibrium, characterized in Section 4. Such an equilibrium is therefore robust to the introduction of a finite time horizon.

2. Relaxing local perfection: Trembles and local perfection were introduced in Section 4.4 in order to refine the set of equilibria. The same refinement can be obtained if, instead of trembles, we relax the assumption that $\Theta_{-i} \equiv \prod_{j \neq i} \left[0, \overline{\theta}_j\right]$. Suppose that $\Theta_{-i} \equiv \prod_{j \neq i} \left[\underline{\theta}_j, \overline{\theta}_j\right]$ with $\underline{\theta}_j < 0 \forall j$. A negative $\theta_{j,t}$ would imply that j would prefer to agree on $\mathbf{x}$ next period rather than immediately. The interpretation of a negative discount rate may be that, in some circumstances, a party prefers to delay signing agreements because of other urgent economic/policy issues that require the decision makers' attention. If $\underline{\theta}_j \uparrow 0$, the claims in Theorem 3 continue to hold without imposing local perfection.

Theorem 6. *Suppose $f(\theta_{j,t}) > 0 \Leftrightarrow \theta_{j,t} \in \left[\epsilon \underline{\theta}_j, \overline{\theta}_j\right]$, where $\underline{\theta}_j < 0$, $\epsilon > 0$. Consider an SSPE with contributions $\mathbf{x}^*(\epsilon)$. For $\mathbf{x}^* \equiv \lim_{\epsilon \to 0} \mathbf{x}^*(\epsilon)$, (4) binds, $\forall i \in N$.*

If $\underline{\theta}_j < 0$ is bounded below zero, then there will be delay on the equilibrium path with some probability, but otherwise the results above will essentially continue to hold. (The proof is available upon request.)

3. Uncertainty other than on the discount rate: For the results above, it is important that the acceptance criterion, (2), is uncertain. As mentioned in Section 3.3, the shock does not need to be related to the discount rate. Equation (3), and thus the subsequent results, would continue to hold if $\theta_{j,t}$ represented a shock on j's subjective belief regarding the lag (Δ) before the next acceptance stage, rather than a shock regarding j's discount rate. Appendix B permits such an alternative shock, and also the possibility that the shock can represent a shock on j's utility and/or marginal utility. In lab experiments, Lippert and Tremewan (2020) find that my results hold, qualitatively, without my exact specification of the uncertainty.

4. Relaxing the assumption $f_i(0) > 0$: Appendix B also contains a discussion of how the results would be modified if there were uncertainty but the density of $\theta_{i,t}$ at zero were zero, that is, if $f_i(0) = 0 \forall i$. In the model above, this case would imply that $w_j^i / w_i^i = 0 \forall j \neq i$. However, these weights (and thus contributions) can be positive even if $f_i(0) = 0 \forall i$, if the model is modified in another direction. To be specific, suppose the pledge x_i must be a discrete number, implying that if i wanted to reduce x_i, i would have to reduce x_i by the magnitude $\Delta_x > 0$, or more. For example, we may require the pledge to be written with a finite number of decimals. If Δ/Δ_x is a finite and strictly positive number, then one can sustain equilibria with strictly positive contributions even if $f_i(0) = 0 \forall i$, and even if $\Delta_x \to 0$, if just $\Delta \to 0$ at the same time, so that Δ/Δ_x continues to be a finite and strictly positive number. Appendix B shows that this modification of the model can permit equilibria with strictly positive contributions, generalizing the main insight of this paper. These equilibria cannot be formulated as neatly as in Theorems 2 and 3, however. Thus, $f_i(0) > 0$ is assumed for tractability.

5. Relaxing the "additional assumptions:" Section 3.1 made the "additional assumptions" that $\partial U_i(\cdot)/\partial x_i < (>) 0$ for $x_i > (<) 0$, and $\partial U_j(\cdot)/\partial x_i > 0$, $j \neq i$. These assumptions are not needed for Theorems 3 and 4, and a generalization of Theorem 2 is proven in Appendix A without these additional assumptions.

6. Remark on existence: Condition (4) is necessary for $\mathbf{x}^*$ to be an SSPE, but it may not be sufficient. Whether the second-order condition for an optimal deviation for i holds, globally, depends on the f_j's. If $n = 2$, a sufficient condition for the second-order condition to hold is that f_j be weakly increasing, as when $\theta_{j,t}$ is uniformly distributed, for example.[22] In this case, a locally perfect SSPE exists, as determined by (6).

If, instead, $\partial f_j(0)/\partial x_i < 0$, then the second-order condition might fail, and a pure-strategy locally perfect SSPE might not exist. In this case, a small decrease in x_i might be unattractive to i because of the risk, but a large decrease in x_i can be attractive because the additional risk, in this case, might not outweigh i's benefit from a much lower x_i. In such a situation, the best-response, x_i, given $\mathbf{x}^*$, can be cyclic, and only mixed-strategy equilibria might survive. When $n > 2$, $\partial f_j(0)/\partial x_i < 0$ is neither sufficient, nor necessary, for the second-order condition to hold; the payoff functions' concavity will also matter.

7. Future research

This paper presents a model and an analysis of pledge-and-review bargaining. The novelty of this bargaining game is that each party proposes how much to contribute independently – not conditional on what other parties pledge – before the parties agree to the vector of pledges. If there is some uncertainty regarding what other parties are willing to accept, for example due to shocks on the short-term discount rate, then contributions can be larger if there is a substantial variance in these shocks. With standard equilibrium refinements, each party's contribution level maximizes an asymmetric Nash bargaining solution, where the weights on others' payoffs reflect the distribution and correlation of shocks. Since the weights vary from pledge to pledge, the bargaining outcome is not Pareto optimal.

The model is simple and can be extended in several directions. Future research should relax the unanimity requirement, allow for persistent shocks, or study alternative equilibrium refinements that stationarity, for example. On the applied side, the model can be more tightly connected to the situations where bargaining takes place between business partners or policymakers, to mention two applications discussed in Section 2.

[22] It is then easy to see from the first-order conditions in the Appendix that the second-order condition holds.

References

Abreu, D., and F. Gul (2000): "Bargaining and Reputation," *Econometrica* 68(1): 85-117.

Abreu, D., D. Pearce, and E. Stacchetti (2015): "One-sided uncertainty and delay in reputational bargaining," *Theoretical Economics* 10(3): 19-773.

Adams, A., L. Cherchye, B. De Rock, and E. Verriest (2014): "Consume now or later? Time inconsistency, collective choice, and revealed preference," *American Economic Review* 104(12): 4147-83.

Admati, A., and M. Perry (1991): "Joint projects without commitment," *Review of Economic Studies* 58(2): 259-76.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2008): "Eliciting risk and time preferences," *Econometrica* 76(3): 583-618.

Andersson, O., C. Argenton, and J. W. Weibull (2018): "Robustness to Strategic Uncertainty in the Nash Demand Game," *Mathematical Social Sciences* 91:1-5.

Angeletos, G-M., D. Laibson, A. Repetto, J. Tobacman, and S. Weinberg (2001): "The Hyperbolic Consumption Model: Calibration, Simulation and Empirical Evaluation," *Journal of Economic Perspectives* 15: 47-68.

Arrow, K. J., M. L. Cropper, C. Gollier, B. Groom, G. Heal, R. G. Newell, W. Nordhaus, R. Pindyck, W. Pizer, P. Portney, T. Sterner, R. Tol, and M. Weitzman (2014): "Should governments use a declining discount rate in project analysis?" *Review of Environmental Economics and Policy* 8(2): 145-63.

Asheim, G. (1992): "A Unique Solution to n-Person Sequential Bargaining," *Games and Economic Behavior* 4: 169-81.

Bagnoli, M., and B. L. Lipman (1989): "Provision of public goods: Fully implementing the core through private contributions," *Review of Economic Studies* 56(4): 583-01.

Baron, D. P., and J. A. Ferejohn (1989): "Bargaining in legislatures," *American Political Sience Review* 83(4): 1181-1206.

Baron, D. P., and E. Kalai (1993): "The Simplest Equilibrium of a Majority-Rule Division Game," *Journal of Economic Theory* 61(2): 290-301.

Bastianello, L., and M. LiCalzi (2019): "The probability to reach an agreement as a foundation for axiomatic bargaining," *Econometrica* 87(3): 837-65.

Battaglini, M., S. Nunnari, and T. R. Palfrey (2014): "Dynamic free riding with irreversible investments," *American Economic Review* 104(9): 2858-71.

Bhaskar, V., G. J. Mailath, and S. Morris (2013): "A Foundation for Markov Equilibria in Sequential Games with Finite Social Memory." *Review of Economic Studies* 8(3): 925–48.

Binmore, K. (1987): "Nash Bargaining Theory (II)," *The Economics of Bargaining*, ed. by K. Binmore and P. Dasgupta. Cambridge: Basil Blackwell.

Binmore, K., M. J. Osborne, and A. Rubinstein (1992): "Non-Cooperative Models of Bargaining," *Handbook of Game Theory* 1, ed. by R. J. Aumann and S. Hart. Elsevier.

Binmore, K., A. Rubinstein, and A. Wolinsky (1986): "The Nash Bargaining Solutin in Economic Modelling," *The RAND Journal of Economics* 17(2): 176-88.

Bodansky, D. (2010): "The Copenhagen Climate Change Conference: A Postmortem," *American Journal of International Law* 104(2): 230-40.

Bowen, R., Y. Chen, and H. Eraslan (2014): "Mandatory versus Discretionary Spending: The Status Quo Effect," *American Economic Review* 104(10): 2941-74.

Britz, V., P. J. Herings, and A. Predtetchinski (2010): "Non-cooperative Support for the Asymmetric Nash Bargaining Solution," *Journal of Economic Theory* 145: 1951-67.

Carlsson, H. (1991): "A Bargaining Model Where Parties Make Errors," *Econometrica* 59(5): 1487-96.

Chambers, C., and F. Echenique (2018): "On multiple discount rates," *Econometrica* 86(4):1325-46.

Chatterjee, K., and L. Samuelson (1990): "Perfect Equilibria in Simultaneous Offers Bargaining," *International Journal of Game Theory* 19: 237-267.

Chen, Y., and H. Eraslan (2013): "Informational loss in bundled bargaining," *Journal of Theoretical Politics* 25(3): 338-62.

Chen, Y., and H. Eraslan (2014): "Rhetoric in legislative bargaining with asymmetric information," *Theoretical Economics* 9(2): 483-513.

Compte, O., and P. Jehiel (2004): "Gradualism in bargaining and contribution games," *Review of Economic Studies* 71(4): 975-1000.

Dietz, S., C. Gollier, and L. Kessler (2018): "The climate beta," *Journal of Environmental Economics and Management* 87: 258-74.

Dreber, A., D. Fudenberg, D. K. Levine, and D. G. Rand (2016): "Self-control, social preferences and the effect of delayed payments," mimeo, MIT.

Eraslan, H., and K. Evdokimov (2919): "Legislative and Multilateral Bargaining," *Annual Review of Economics* 11: 443-72.

Falkner, R. (2016): "The Paris Agreement and the new logic of international climate politics," *International Affairs* 92(5): 1107-25.

Fershtman, C. (1990): "The importance of the agenda in bargaining," *Games and Economic Behavior* 2(3): 224-38.

Finus, M., and S. Maus (2008): "Modesty may pay!" *Journal of Public Economic Theory* 10: 801–26.

Frederick, S., G. Loewenstein, and T. O'Donoghue (2002): "Time Discounting and Time Preference," *Journal of Economic Literature* 40(2): 351-401.

Friedenberg, A. (2019): "Bargaining under strategic uncertainty: The role of second-order optimism," *Econometrica* 87(6): 1835–65.

Fukuda, S., and Y. Kamada (2020): "Negotiations with Limited Specifiability," *American Economic Journal: Microeconomics*, forthcoming.

Gollier, C., and R. Zeckhauser (2005): "Aggregation of heterogeneous time preferences," *Journal of Political Economy* 113(4): 878-96.

Harstad, B. (2020): "Technology and Time Inconsinstency," *Journal of Political Economy* 128(7): 2653-89.

Harstad, B. (2021): "Pledge-and-Review Bargaining: From Kyoto to Paris," mimeo.

Horstmann, I., J. R. Markusen, and J. Robles (2005): "Issue Linking in Trade Negotiations: Ricardo Revisited or No Pain No Gain," *Review of International Economics* 13(2): 185-204.

In, Y., and R. Serrano (2004): "Agenda restrictions in multi-issue bargaining," *Journal of Economic Behavior & Organization* 53(3): 385-99.

Jackson, M. O., and S. Wilkie (2005): "Endogenous Games and Mechanisms: Side Payments Among Players," *Review of Economic Studies* 72(2): 543-66.
Jackson, M. O., and L. Yariv (2014): "Present bias and collective dynamic choice in the lab," *American Economic Review* 104(12): 4184-4204.
Kambe, S. (2000): "Bargaining with Imperfect Commitment," *Games and Economic Behavior* 28: 217-37.
Keohane, R. O., and M. Oppenheimer (2016): "Paris: Beyond the Climate Dead End through Pledge and Review," *Politics and Governance* 4(3): 42-51.
Krishna, V., and R. Serrano (1996): "Multilateral Bargaining", *Review of Economic Studies* 63(1): 61-80.
Laibson, D. (1997): "Golden eggs and hyperbolic discounting," *The Quarterly Journal of Economics* 112(2): 443-78.
Laurelle, A., and F. Valenciano (2008): "Non-Cooperative Foundations of Bargaining Power in Committees and the Shapley-Shubik Index," *Games and Economic Behavior* 63: 341-53.
Ledyard, J. O., and T. R. Palfrey (2002): "The approximation of efficient public good mechanisms by simple voting schemes," *Journal of Public Economics* 83(2): 153-71.
Lippert, S., and J. Tremewan (2020): "Pledge-and-Review in the Laboratory," mimeo.
Maggi, G. (2016): "Issue linkage," *Handbook of Commercial Policy* 1-B, ed. by K. Bagwell and R. W. Staiger. Elsevier.
Marx, L. M., and S. A. Matthews (2000): "Dynamic voluntary contribution to a public project," *Review of Economic Studies* 67(2): 327-58.
Matthews, S. A. (2013): "Achievable outcomes of dynamic contribution games," *Theoretical Economics* 8(2): 365-403
Miyakawa, T. (2008): "Note on the Equal Split Solution in an n-Person Non-Cooperative Bargaining Game," *Mathematical Social Sciences* 55(3): 281-91.
Morelli M. (1999): "Demand competition and policy compromise in legislative bargaining," *American Political Science Review* 93:809-20.
Myerson, R. (1978): "Refinement of the Nash Equilibrium Concept," *International Journal of Game Theory* 7: 73-80.
Nash, J. F. (1950): "The Bargaining Problem," *Econometrica* 18: 155-62.
Nash, J. F. (1953): "Two-Person Cooperative Games," *Econometrica* 21(1): 128-40.
Okada, A. (2010): "The Nash bargaining solution in general n-person cooperative games," *Journal of Economic Theory* 145(6): 2356-79.
Ortner, J. (2017): "A theory of political gridlock," *Theoretical Economics* 12(2): 555–86.
Osborne, M. J., and A. Rubinstein (1990): *Bargaining and Markets*, Academic Press.
Palfrey, T. R., and H. Rosenthal (1984): "Participation and the provision of discrete public goods: a strategic analysis," *Journal of Public Economics* 24(2): 171-93.
Ramsey, F. P. (1928): "A mathematical theory of saving," *The Economic Journal* 38(152): 543-59.
Rong, R., T. C. Grijalva, J. Lusk, and W. D. Shaw (2019): "Interpersonal discounting," *Journal of Risk and Uncertainty* 58(1): 17-42.
Rubinstein, A. (1982): "Perfect Equilibrium in a Bargaining Model," *Econometrica* 50(1): 97-109.

Rubinstein, A. (1985): "A bargaining model with incomplete information about time preferences," *Econometrica* 53(5): 1151-72.

Selten, R. (1975): "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory* 4(1): 25-55.

Serrano, R. (2020): "Sixty-Seven Years of the Nash Program: Time for Retirement?" *Journal of the Spanish Economic Association*, forthcoming.

Simon, L. K. (1987): "Local Perfection," *Journal of Economic Theory* 43: 134-56..

Sutton, J. (1986): "Non-Cooperative Bargaining Theory: An Introduction," *The Review of Economic Studies* 53(5): 709-24.

von Hagen, J., and I. J. Harden (1995): "Budget processes and commitment to fiscal discipline," *European Economic Review* 39(3-4): 771-79.

Watson, J. (1998): "Alternating-Offer Bargaining with Two-Sided Incomplete Information," *Review of Economic Studies* 65(3): 573–94.

Weingast, B. R., K. Shepsle, and C. Johnson (1981): "The political economy of benefits and costs: A neoclassical approach to distributive policies," *Journal of Political Economy* 89: 642-64.

Weitzman, M. L. (2001): "Gamma discounting," *American Economic Review* 91(1): 260-71.

Yildiz, M. (2003): "Walrasian bargaining," *Games and Economic Behavior* 45(2): 465-87.

APPENDIX A: PROOFS

Proof of Theorem 1

The first part of this proof follows the same steps as the proof of Theorem 2. To economize on space, the additional steps, required for Theorem 1, are introduced and discussed at the end of the proof of Theorem 2. ‖

Proof of Theorem 2

As advertised in Section 4, the following generalization of Theorem 2 is here proven without the additional assumptions $\partial U_i(\cdot)/\partial x_i < (>)\, 0$ for $x_i > (<)\, 0$, and $\partial U_j(\cdot)/\partial x_i > 0$, $\forall j \neq i$.

Theorem A-2. *If* $\mathbf{x}^*$ *is an SSPE in which* $U_i(\mathbf{x}^*) > 0 \forall i$*, then, for every* $i \in N$*, we have:*
(a) if $\partial U_i(\mathbf{x}^*)/\partial x_i \leq 0$*,*

$$-\frac{\partial U_i(\mathbf{x}^*)/\partial x_i}{\rho_i U_i(\mathbf{x}^*)} \leq \sum_{j\neq i} \max\left\{0, \frac{\partial U_j(\mathbf{x}^*)/\partial x_i}{\rho_j U_j(\mathbf{x}^*)}\right\} f_j(0)\,\mathrm{E}(\theta_{i,t} \mid \theta_{j,t} = 0)\,; \tag{12}$$

(b) if $\partial U_i(\mathbf{x}^*)/\partial x_i > 0$*,*

$$\frac{\partial U_i(\mathbf{x}^*)/\partial x_i}{\rho_i U_i(\mathbf{x}^*)} \leq \sum_{j\neq i} \max\left\{0, -\frac{\partial U_j(\mathbf{x}^*)/\partial x_i}{\rho_j U_j(\mathbf{x}^*)}\right\} f_j(0)\,\mathrm{E}(\theta_{i,t} \mid \theta_{j,t} = 0)\,.$$

With the constraint $x_i \geq 0 \forall i \in N$, and the additional assumptions $\partial U_i(\cdot)/\partial x_i < 0$, $\partial U_j(\cdot)/\partial x_i > 0$, $\forall j \in N \backslash i$, (12) corresponds to the first-order condition of the right-hand side of (4).

Proof of part (a): First, note that in any SSPE we must have $U_i(\mathbf{x}^*) \geq 0 \forall i$, since otherwise a party with $U_i(\mathbf{x}^*) < 0$ would always reject $\mathbf{x}^*$ in order to obtain the default payoff, normalized to zero. I will search for equilibria in which $U_i(\mathbf{x}^*) > 0 \forall i$.

Consider an equilibrium $\mathbf{x}^*$, satisfying $U_j(\mathbf{x}^*) > 0 \forall j$. When $\mathbf{x}^*$ is proposed, it will be accepted with probability 1 since $\rho_{j,t} \geq 0$. Therefore, i will never offer $x_i > x_i^*$ when $\frac{\partial U_i(\mathbf{x}^*)}{\partial x_i} \leq 0$, so to check when $\mathbf{x}^*$ is an equilibrium, it is sufficient to consider a deviation by i, $\mathbf{x}^i$, such that $x_i^i < x_i^*$ while $x_j^i = x_j^*$, $j \neq i$.

Acceptable offers: Let $P\left(\mathbf{x}^i;\mathbf{x}^*\right)$ be the probability that at least one $j \neq i$ rejects $\mathbf{x}^i$, and $P_{-j}\left(\mathbf{x}^i;\mathbf{x}^*\right)$ the probability that at least one party other than j and i rejects $\mathbf{x}^i$, given an equilibrium $\mathbf{x}^*$.

Since party j's discount factor can be written as $1-\rho_{j,t}\Delta = 1-\theta_{j,t}\rho_j\Delta$, $j \neq i$ rejects $\mathbf{x}^i$ if and only if:

$$\left(1-P_{-j}\left(\mathbf{x}^i\right)\right)U_j\left(\mathbf{x}^i\right)+P_{-j}\left(\mathbf{x}^i\right)\left(1-\rho_{j,t}\Delta\right)U_j\left(\mathbf{x}^*\right)<\left(1-\rho_{j,t}\Delta\right)U_j\left(\mathbf{x}^*\right) \Longleftrightarrow$$
$$\theta_{j,t}<\widetilde{\theta}_j\left(\mathbf{x}^i\right)\equiv\max\left\{0,\frac{U_j\left(\mathbf{x}^*\right)-U_j\left(\mathbf{x}^i\right)}{\rho_j\Delta U_j\left(\mathbf{x}^*\right)}\right\}. \tag{13}$$

Note on the derivative: Since we only need to consider $x_i \leq x_i^*$ and U_j is a function concave, $U_j\left(\mathbf{x}^*\right) \leq U_j\left(\mathbf{x}^i\right)$ holds if $\partial U_j\left(\mathbf{x}^i\right)/\partial x_i \leq 0$. In this case, j benefits from the deviation so j accepts $\mathbf{x}^i$ with probability 1, $\widetilde{\theta}_j\left(\mathbf{x}^i\right)=0$, and $\partial\widetilde{\theta}_j\left(\mathbf{x}^i\right)/\partial x_i=0$. If, instead, $U_j\left(\mathbf{x}^*\right)>U_j\left(\mathbf{x}^i\right)$, $\partial\widetilde{\theta}_j\left(\mathbf{x}^i\right)/\partial x_i=\left[-\partial U_j\left(\mathbf{x}^i\right)/\partial x_i\right]/\rho_j\Delta U_j\left(\mathbf{x}^*\right)<0$. For both cases, it holds that:

$$\frac{\partial\widetilde{\theta}_j\left(\mathbf{x}^i\right)}{\partial x_i}=-\max\left\{0,\frac{\partial U_j\left(\mathbf{x}^i\right)/\partial x_i}{\rho_j\Delta U_j\left(\mathbf{x}^*\right)}\right\}\leq 0.$$

When the joint pdf of shocks $\theta_t=\left(\theta_{1,t},...,\theta_{n,t}\right)$ is represented by $f\left(\theta_t\right)$, the probability that every $j\neq i$ accepts $\mathbf{x}^i$ can be written as follows if every $\widetilde{\theta}_j\left(\mathbf{x}^i\right)\leq\overline{\theta}_j$ (when dx_i is small, then $\widetilde{\theta}_j\left(\mathbf{x}^i\right)$ is proportional to dx_i):

$$1-P\left(\mathbf{x}^i\right)=G\left(\widetilde{\theta}_1\left(\mathbf{x}^i\right),...,\widetilde{\theta}_{i-1}\left(\mathbf{x}^i\right),\widetilde{\theta}_{i+1}\left(\mathbf{x}^i\right),...\widetilde{\theta}_n\left(\mathbf{x}^i\right)\right) \tag{14}$$
$$\equiv\int_0^{\overline{\theta}_i}\left[\int_{\widetilde{\theta}_1\left(\mathbf{x}^i\right)}^{\overline{\theta}_1}...\int_{\widetilde{\theta}_{i-1}\left(\mathbf{x}^i\right)}^{\overline{\theta}_{i-1}}\int_{\widetilde{\theta}_{i+1}\left(\mathbf{x}^i\right)}^{\overline{\theta}_{i+1}}...\int_{\widetilde{\theta}_n\left(\mathbf{x}^i\right)}^{\overline{\theta}_n}f\left(\theta_t\right)d\theta_{-i,t}\right]d\theta_i,$$

note that the identity defines G as a function of $n-1$ thresholds, each given by (13). If we take the (right) derivative of (14) w.r.t. x_i^i and use the chain rule, we get:

$$-\frac{\partial P\left(\mathbf{x}^i\right)}{\partial x_i}=\sum_{j\neq i}-\max\left\{0,\frac{\partial U_j\left(\mathbf{x}^i\right)/\partial x_i}{\rho_j\Delta U_j\left(\mathbf{x}^*\right)}\right\}G_j'\left(\widetilde{\theta}_1\left(\mathbf{x}^i\right),...,\widetilde{\theta}_{i-1}\left(\mathbf{x}^i\right),\widetilde{\theta}_{i+1}\left(\mathbf{x}^i\right),...\widetilde{\theta}_n\left(\mathbf{x}^i\right)\right).$$

So, at the equilibrium, $\mathbf{x}^i = \mathbf{x}^*$, we have:

$$\frac{\partial P(\mathbf{x}^*)}{\partial x_i} = \sum_{j \neq i} \max\left\{0, \frac{\partial U_j(\mathbf{x}^*)/\partial x_i}{\rho_j \Delta U_j(\mathbf{x}^*)}\right\} G'_j(\mathbf{0}) = -\sum_{j \neq i} \max\left\{0, \frac{\partial U_j(\mathbf{x}^*)/\partial x_i}{\rho_j \Delta U_j(\mathbf{x}^*)}\right\} f_j(0),$$

where, as written in the text already, $f_j(0)$ is defined as the marginal distribution of $\theta_{j,t}$ at $\theta_{j,t} = 0$.

Equilibrium offers: When proposing x_i, party i's problem is to choose $x_i \leq x_i^*$ so as to maximize

$$\left(1 - P\left(\mathbf{x}^i\right)\right) U_i\left(\mathbf{x}^i\right) + P\left(\mathbf{x}^i\right)\left(1 - \mathrm{E}\theta_{i,t}^R \rho_i \Delta\right) U_i(\mathbf{x}^*), \tag{15}$$

where $\mathrm{E}\theta_{i,t}^R$ is the expected $\theta_{i,t}$ conditional on being rejected (this will be more precise in eq. (18)).

To derive a necessary condition for when it is optimal to propose the equilibrium pledge, x_i^*, suppose i considers a small (marginal) reduction in x_i relative to x_i^*, given by $dx_i = x_i^i - x_i^* < 0$. If accepted, this gives i utility

$$U_i\left(\mathbf{x}^i\right) \approx U_i(\mathbf{x}^*) + dx_i \partial U_i(\mathbf{x}^*)/\partial x_i > U_i(\mathbf{x}^*), \tag{16}$$

but $\mathbf{x}^i$ is rejected with probability

$$P\left(\mathbf{x}^i\right) \approx P(\mathbf{x}^*) + \frac{\partial P(\mathbf{x}^*)}{\partial x_i} dx_i = 0 - \sum_{j \neq i} \max\left\{0, \frac{\partial U_j(\mathbf{x}^*)/\partial x_i}{\rho_j \Delta U_j(\mathbf{x}^*)}\right\} dx_i f_j(0), \tag{17}$$

where each of the $n-1$ terms represents the probability that a $\theta_{j,t}$ is so small that j rejects if x_i is reduced by dx_i, i.e., $\Pr\left(\theta_{j,t} \leq \widehat{\theta}_j\right)$ for $\widehat{\theta}_j \equiv \frac{\partial U_j\left(\mathbf{x}^i\right)/\partial x_i}{\rho_j \Delta U_j(\mathbf{x}^*)} |dx_i|$. Naturally, the probability that more than one party has such a small $\theta_{j,t}$ vanishes when $|dx_i| \to 0$ since f is assumed to have no mass point.

If we combine (15), (16), and (17), we find party i's expected payoff when proposing x_i^i. This payoff, written on the left-hand side in the following inequality, is smaller than i's payoff if i sticks to the SSPE by proposing x_i^* if and only if:

$$\left(1+\sum_{j\neq i}\max\left\{0,\frac{\partial U_j\left(\mathbf{x}^*\right)/\partial x_i}{\rho_j\Delta U_j\left(\mathbf{x}^*\right)}\right\}f_j\left(0\right)dx_i\right)\left(U_i\left(\mathbf{x}^*\right)+dx_i\frac{\partial U_i\left(\mathbf{x}^*\right)}{\partial x_i}\right) \tag{18}$$
$$-\sum_{j\neq i}\max\left\{0,\frac{\partial U_j\left(\mathbf{x}^*\right)/\partial x_i}{\rho_j\Delta U_j\left(\mathbf{x}^*\right)}\right\}dx_i f_j\left(0\right)\left(1-\mathrm{E}\left(\theta_{i,t}\mid\theta_{j,t}\leq\widehat{\theta}_{j,t}\right)\rho_i\Delta\right)U_i\left(\mathbf{x}^*\right)\leq U_i\left(\mathbf{x}^*\right),$$

where $\mathrm{E}\left(\theta_{i,t}\mid\theta_{j,t}\leq\widehat{\theta}_j\right)$ follows from Bayes' rule:

$$\mathrm{E}\left(\theta_{i,t}\mid\theta_{j,t}\leq\widehat{\theta}_j\right)\equiv\frac{\int_0^{\widehat{\theta}_j}\int_{\Theta_{-j}}\theta_{i,t}f\left(\theta_t\right)d\theta_j d\Theta_{-j}}{\int_0^{\widehat{\theta}_j}\int_{\Theta_{-j}}f\left(\theta_t\right)d\theta_j d\Theta_{-j}},\ \mathrm{E}\left(\theta_{i,t}\mid\theta_{j,t}=0\right)\equiv\lim_{dx_i\uparrow 0}\frac{\int_0^{\widehat{\theta}_j}\int_{\Theta_{-j}}\theta_{i,t}f\left(\theta_t\right)d\theta_j d\Theta_{-j}}{\int_0^{\widehat{\theta}_j}\int_{\Theta_{-j}}f\left(\theta_t\right)d\theta_j d\Theta_{-j}},$$

and, as already defined, $\Theta_{-j}\equiv\prod_{k\neq j}\left[0,\overline{\theta}_k\right]$ and $\widehat{\theta}_j\equiv\frac{\partial U_j(\mathbf{x}^*)/\partial x_i}{\rho_j\Delta U_j(\mathbf{x}^*)}\left|dx_i\right|$.

When both sides of (18) are divided by $|dx_i|$ and $dx_i\uparrow 0$, (18) can be rewritten as (12).

The proof of part (b) is analogous and thus omitted.

Remark on the proof of Theorem 1 (Folk theorem): I will now construct strategies that can support as an SPE any $\mathbf{U}^*\in\mho_{\mathbf{U}}$, where $\mho_{\mathbf{U}}$ is an open set. In this case, if $\mathbf{U}^*$ can be supported as an SPE, then so can also $\mathbf{U}^i$, where $U_j^*=k_jU_j^i$ for $k_j=1$ when $j\neq i$, and $k_i\in(0,1)$. The idea is that if i deviates, then the parties punish i by switching from $\mathbf{U}^*$ to $\mathbf{U}^i$. In this situation, i loses from the deviation if and only if (i.e., eq. (18) becomes):

$$\left(1+\sum_{j\neq i}\left[\max\left\{0,\frac{\partial U_j\left(\mathbf{x}^*\right)/\partial x_i}{\rho_j\Delta U_j\left(\mathbf{x}^*\right)}\right\}f_j\left(0\right)dx_i\right]\right)\left(U_i\left(\mathbf{x}^*\right)+dx_i\frac{\partial U_i\left(\mathbf{x}^*\right)}{\partial x_i}\right)$$
$$-\sum_{j\neq i}\left[\max\left\{0,\frac{\partial U_j\left(\mathbf{x}^*\right)/\partial x_i}{\rho_j\Delta U_j\left(\mathbf{x}^*\right)}\right\}dx_i f_j\left(0\right)\left(1-E\left(\theta_{i,t}\mid\theta_{j,t}\leq\widehat{\theta}_{j,t}\right)\right)\right]\rho_i\Delta k_i U_i\left(\mathbf{x}^*\right)\leq U_i\left(\mathbf{x}^*\right).$$

If $\Delta\downarrow 0$, while $dx_i\nrightarrow 0$, then $P\left(\mathbf{x}^i\right)$ grows and hits 1, so such a deviation cannot be beneficial to i. Suppose, thus, that $dx_i \uparrow 0$. The inequality can then be written as:

$$\begin{aligned}
&(-dx_i)\left(-\frac{\partial U_i(\mathbf{x}^*)}{\partial x_i}\right) \leq \\
&(-dx_i)\sum_{j\neq i}\max\left\{0, \frac{\partial U_j(\mathbf{x}^*)/\partial x_i}{\rho_j \Delta U_j(\mathbf{x}^*)}\right\} f_j(0)\left[1-\left(1-E\left(\theta_{i,t}\mid\theta_{j,t}\leq\widehat{\theta}_{j,t}\right)\rho_i\Delta\right)k_i\right]U_i(\mathbf{x}^*) \Rightarrow \\
&-\frac{\partial U_i(\mathbf{x}^*)}{\partial x_i} \leq \sum_{j\neq i}\max\left\{0, \frac{\partial U_j(\mathbf{x}^*)/\partial x_i}{\rho_j U_j(\mathbf{x}^*)}\right\} f_j(0)\left[\frac{1-k_i}{\Delta}+E\left(\theta_{i,t}\mid\theta_{j,t}\leq\widehat{\theta}_{j,t}\right)\rho_i k_i\right]U_i(\mathbf{x}^*).
\end{aligned}$$

When $k_i = 1$, as in the proof of Theorem 2, the role of Δ vanished. With $k_i \in (0,1)$, the right-hand side increases without bounds when $\Delta \downarrow 0$, ensuring that there exists $\underline{\Delta} \in (0,\infty)$ such that the inequality holds for every $\Delta \in (0,\underline{\Delta}]$ and, in this case, i does not benefit from the deviation. Note that the conclusion is the same if ρ_j, instead of Δ, is small.

Since $k_i > 0$, $U_j^i > 0 \forall (i,j) \in N^2$, so deviations from the punishment can be deterred in the same way.‖

Proof of Theorem 3

A continuum of $\mathbf{x}^*$'s can satisfy the equilibrium condition in Theorem A-2. To provide an illustration of this, note that if (12) binds then (12) continues to be satisfied when x_i^* is reduced. The idea of local perfection is to introduce trembles such that equilibrium offers can be rejected (i.e., $P(\mathbf{x}^*) > 0$) and thus we must check that i cannot benefit from marginally increasing *or* decreasing x_i^i from x_i^* to reduce $P(\mathbf{x}^i)$.[23] It will now be proven that, with trembles, i will strictly benefit from $dx_i > 0$ when (4) is strict, and thus it must hold with equality at $\mathbf{x}^*$.

The vector s_t is i.i.d. over time according to some cdf, $p(\cdot)$, with is assumed to have a bounded support and $\partial p(\mathbf{0})/\partial s_{i,t} > 0$ on a neighborhood of $\mathbf{0}$. When j considers whether to accept $U_j(\mathbf{x}^i + \epsilon \mathbf{s}_t)$, after i has deviated and the vector of pledges is $\mathbf{x}^i$, then j faces the continuation value $V_j(\mathbf{x}^*)$ by rejecting, where $V_j(\mathbf{x}^*)$ takes into account that $\mathbf{x}^*$ can be rejected in the future (if the future s_t's are sufficiently small):

To write the equation for $V_j(\mathbf{x}^*)$, note that it is the *combination* of the $s_{i,t}$'s and the

[23] Above, $P(\cdot)$ was the probability that any $j \in N \backslash i$ rejects. For simplicity, I here refer to $P(\cdot)$ as the probability that anyone (including i) rejects. This simplification is inconsequential when the trembles vanish, as I assume.

$\theta_{j,t}$'s that determines whether j rejects $\mathbf{x}^*$: let $\Phi_A(\mathbf{x}^*)$ be the set of $s_{i,t}$'s and $\theta_{j,t}$'s such that every j accepts $\mathbf{x}^*$, while $\Phi_R(\mathbf{x}^*)$ is the complementary set.[24] We then have $P(\mathbf{x}^*) = \Pr\{(\mathbf{s}_t, \theta_t) \in \Phi_R(\mathbf{x}^*)\}$, where $\theta_t = (\theta_{1,t}, ..., \theta_{n,t})$, and:

$$V_j(\mathbf{x}^*) = (1 - P(\mathbf{x}^*)) \mathrm{E}_{\mathbf{s}_t:(\mathbf{s}_t,\theta_t)\in\Phi_A(\mathbf{x}^*)} U_j(\mathbf{x}^* + \epsilon \mathbf{s}_t) \tag{19}$$
$$+ P(\mathbf{x}^*) V_j(\mathbf{x}^*) \mathrm{E}_{\theta_{j,t}:(\mathbf{s}_t,\theta_t)\in\Phi_R(\mathbf{x}^*)} (1 - \theta_{j,t}\rho_j\Delta).$$

The shocks, combined with the option to reject, imply that $V_j(\mathbf{x}^*) > 0$ even if $U_j(\mathbf{x}^*) = 0$, so there is no longer any need to assume $U_j(\mathbf{x}^*) > 0 \forall j$.

With this, party $j \neq i$ rejects $\mathbf{x}^i$ if and only if:

$$\left(1 - P_{-j}\left(\mathbf{x}^i\right)\right) U_j\left(\mathbf{x}^i + \epsilon \mathbf{s}_t\right) + P_{-j}\left(\mathbf{x}^i\right)(1 - \rho_{j,t}\Delta) V_j(\mathbf{x}^*) < (1 - \rho_{j,t}\Delta) V_j(\mathbf{x}^*) \Leftrightarrow$$

$$1 - \theta_{j,t}\rho_j\Delta > \frac{U_j(\mathbf{x}^i + \epsilon \mathbf{s}_t)}{V_j(\mathbf{x}^*)} \Longleftrightarrow \theta_{j,t} < \widetilde{\theta}_j\left(\mathbf{x}^i\right) \equiv \frac{V_j(\mathbf{x}^*) - U_j(\mathbf{x}^i + \epsilon \mathbf{s}_t)}{\rho_j \Delta V_j(\mathbf{x}^*)}. \tag{20}$$

Here, $\widetilde{\theta}_j(\mathbf{x}^i)$ is a function of s_t. To simplify the notation, I assume $\widetilde{\theta}_j(\mathbf{x}^i) \in \left(0, \overline{\theta}_i\right)$ for every s_t when $dx_i < 0$ is small (when dx_i is small, then $\widetilde{\theta}_j(\mathbf{x}^i)$ is proportional to dx_i). The probability that every $j \neq i$ accepts can then be written as:

$$1 - P\left(\mathbf{x}^i\right) = \int_{\mathbf{s}_t} G\left(\widetilde{\theta}_1\left(\mathbf{x}^i\right), ..., \widetilde{\theta}_{i-1}\left(\mathbf{x}^i\right), \widetilde{\theta}_{i+1}\left(\mathbf{x}^i\right), ...\widetilde{\theta}_n\left(\mathbf{x}^i\right)\right) dp(\mathbf{s}_t)$$
$$\equiv \int_{\mathbf{s}_t} \int_0^{\overline{\theta}_i} \left[\int_{\widetilde{\theta}_1(\mathbf{x}^i)}^{\overline{\theta}_1} ... \int_{\widetilde{\theta}_{i-1}(\mathbf{x}^i)}^{\overline{\theta}_{i-1}} \int_{\widetilde{\theta}_{i+1}(\mathbf{x}^i)}^{\overline{\theta}_{i+1}} ... \int_{\widetilde{\theta}_n(\mathbf{x}^i)}^{\overline{\theta}_n} f(\theta_t)\, d\theta_{-i,t}\right] d\theta_i dp(\mathbf{s}_t) \Rightarrow$$

$$-\frac{\partial P\left(\mathbf{x}^i\right)}{\partial x_i} = \mathrm{E}_{\mathbf{s}_t} \sum_{j\neq i} -\frac{\partial U_j(\mathbf{x}^i + \epsilon \mathbf{s}_t)/\partial x_i}{\rho_j \Delta V_j(\mathbf{x}^*)} G'_j\left(\widetilde{\theta}_1\left(\mathbf{x}^i\right), ..., \widetilde{\theta}_{i-1}\left(\mathbf{x}^i\right), \widetilde{\theta}_{i+1}\left(\mathbf{x}^i\right), ...\widetilde{\theta}_n\left(\mathbf{x}^i\right)\right).$$

The condition under which i does not benefit from a marginal change $dx_i < 0$ is given

[24] By referring to (20), below, $\Phi_A(\mathbf{x}^*)$ and $\Phi_R(\mathbf{x}^*)$ are defined as:

$$\Phi_A(\mathbf{x}^*) = \left\{(\mathbf{s}_t, \theta_t) : \theta_{j,t} \geq \frac{V_j(\mathbf{x}^*) - U_j(\mathbf{x}^* + \epsilon \mathbf{s}_t)}{\rho_j \Delta V_j(\mathbf{x}^*)} \forall j\right\},$$
$$\Phi_R(\mathbf{x}^*) = \left\{(\mathbf{s}_t, \theta_t) : \theta_{j,t} < \frac{V_j(\mathbf{x}^*) - U_j(\mathbf{x}^* + \epsilon \mathbf{s}_t)}{\rho_j \Delta V_j(\mathbf{x}^*)} \text{ for at least one } j\right\}.$$

by an equation that is analogous to (18), although we now have to take into account the trembles:

$$\begin{aligned}&\mathrm{E}_{\mathbf{s}_t:(\mathbf{s}_t,\theta_t)\in\Phi_A(\mathbf{x}^i)}\left(1-P\left(\mathbf{x}^*\right)-\frac{\partial P\left(\mathbf{x}^*\right)}{\partial x_i}dx_i\right)\left(U_i\left(\mathbf{x}^*+\epsilon\mathbf{s}_t\right)+\frac{\partial U_i\left(\mathbf{x}^*+\epsilon\mathbf{s}_t\right)}{\partial x_i}dx_i\right)+ \\ &\mathrm{E}_{(\mathbf{s}_t,\theta_t):(\mathbf{s}_t,\theta_t)\in\Phi_R(\mathbf{x}^i)}\left[P\left(\mathbf{x}^*\right)+\sum_{j\neq i}\left[\frac{\partial U_j\left(\mathbf{x}^*+\epsilon\mathbf{s}_t\right)/\partial x_i}{\rho_j\Delta V_j\left(\mathbf{x}^*\right)}dx_iG_j'\left(\frac{V_1\left(\mathbf{x}^*\right)-U_1\left(\mathbf{x}^*+\epsilon\mathbf{s}_t\right)}{\rho_1\Delta V_1\left(\mathbf{x}^*\right)},...\right)\right]\right] \\ &\cdot\left(1-\theta_{i,t}\rho_i\Delta\right)V_i\left(\mathbf{x}^*\right)\leq V_i\left(\mathbf{x}^*\right).\end{aligned} \tag{21}$$

Since the trembles imply that $P\left(\mathbf{x}^*\right)>0$, i might benefit from reducing this risk and consider a marginal increase $dx_i>0$. Party i will not benefit from $dx_i>0$ if (21) holds with the reverse inequality sign ($\geq$); the proof for this is analogous to the proof above for when party i would not benefit from $dx_i<0$. Consequently, (21) must hold with equality for no marginal deviation to be beneficial to i. (Note that (21) must hold with equality regardless of whether $U_i\left(\cdot\right)$ would increase when $dx_i>0$ or when $dx_i<0$, so, we do not need to impose the assumptions $\partial U_i\left(\cdot\right)/\partial x_j>0$ for $j\neq i$ and <0 for $j=i$.)

When we let $\epsilon\to 0$, so that the trembles vanish, then we can see from (19) and (20) that $P\left(\mathbf{x}^*\right)\to 0$ and $V_j\left(\mathbf{x}^*\right)\to U_j\left(\mathbf{x}^*\right)$. When these limits are substituted into (21), holding with equality, and we divide both sides by dx_i before we let $dx_i\to 0$ and $\epsilon s_t\to 0$, then (21) can be rewritten as:

$$\frac{\partial U_i\left(\mathbf{x}^*\right)}{\partial x_i}+\sum_{j\neq i}\frac{\partial U_j\left(\mathbf{x}^*\right)/\partial x_i}{\rho_j\Delta U_j\left(\mathbf{x}^*\right)}f_j\left(0\right)\mathrm{E}\left(\theta_{i,t}\mid\theta_{j,t}=0\right)\rho_i\Delta U_i\left(\mathbf{x}^*\right)=0, \tag{22}$$

which coincides with the first-order condition of

$$\arg\max_{x_i}\prod_{j\in N}\left(U_j\left(x_i,\mathbf{x}_{-i}^*\right)\right)^{w_j^i},$$

when $\frac{w_j^i}{w_i^i}=\frac{\rho_i}{\rho_j}f_j\left(0\right)\mathrm{E}(\theta_{i,t}\mid\theta_{j,t}=0)$, $\forall j\neq i$. $\|$

Proof of Theorem 4

With (8), a binding (4) implies:

$$\begin{aligned}
x_i^* &= \arg\max_{x_i} \prod_{j\in N} \left(d_j\left(x_j\right)p\left(\mathbf{x}\right)\right)^{w_j^i} = \arg\max_{x_i} d_i\left(x_i\right)p\left(\mathbf{x}\right)^{\sum_j w_j^i/w_i^i} \\
&= \arg\max_{x_i} d_i\left(x_i\right)^{w_i^i/\sum_j w_j^i} p\left(\mathbf{x}\right) = \arg\max_{x_i} \prod_{j\in N} d_j\left(x_j\right)^{w_j^j/\sum_k w_k^j} p\left(\mathbf{x}\right), \text{ so} \\
\mathbf{x}^* &= \arg\max_{\mathbf{x}} \prod_{j\in N} d_j\left(x_j\right)^{w_j^j/\sum_k w_k^j} p\left(\mathbf{x}\right),
\end{aligned}$$

which can be written as (9), given the definitions ϱ_i and ω. Given $\mathbf{x}^*$, (9) can be rewritten as (10). ‖

Proofs of Theorems 5 and 6

The proof of Theorem 5(i) is very similar to the proof of Theorem 2, while the proofs of Theorem 5(ii) and Theorem 6 are both analogous to the proof of Theorem 3. These proofs are thus omitted but they are available upon request.

FOR ONLINE PUBLICATION ONLY

APPENDIX B: P&R BARGAINING AND UNCERTAINTY

This online appendix builds on the proof of Theorem A-2 to investigate conditions under which contributions can be positive with P&R bargaining and how the outcome can be characterized under alternative assumptions. In short, I show that contributions can be positive even if $f_j(0) = 0$ if instead either $\Delta \to 0$ or if there is a boundary for how small the reduction in x_i might be. For simplicity, I make the additional assumptions that increasing $x_i > 0$ is costly for i but beneficial for everyone else.

No uncertainty: I start with the basic situation in which there is no uncertainty on the discount rates. Consider the restriction that $x_i = \Delta_i^x \varsigma$, where ς can be any positive integer. That is, if i reduces x_i from x_i^*, i must reduce x_i by at least the amount Δ_i^x. For example, if x_i must be described by a real number with at most ϑ_i decimals, then $\Delta_i^x = 1/10^{\vartheta_i}$. I am especially interested in the limit $\Delta_i^x \to 0$, so that x_i can approximate any real number. If both $\Delta_i^x \to 0$ and $\Delta \to 0$, $\chi_i \equiv \Delta_i^x/\Delta$ might be a finite and strictly positive number.

If i deviates by offering $x_i^i = x_i^* - \Delta_i^x$, then j rejects if and only if:

$$U_j\left(\mathbf{x}^i\right) < (1 - \rho_j \Delta) U_j(\mathbf{x}^*).$$

When Δ_i^x is small, this inequality is approximated as:

$$\begin{aligned} U_j\left(\mathbf{x}^i\right) = U_j(\mathbf{x}^*) - \frac{\partial U_j(\mathbf{x}^*)}{\partial x_i}\Delta_i^x < (1 - \rho_j \Delta) U_j(\mathbf{x}^*) \Leftrightarrow \\ \frac{\partial U_j(\mathbf{x}^*)/\partial x_i}{U_j(\mathbf{x}^*)} > \frac{\rho_j}{\chi_i}. \end{aligned} \tag{23}$$

Thus, for $\mathbf{x}^*$ to be an equilibrium, $\frac{\partial U_j(\mathbf{x}^*)/\partial x_i}{U_j(\mathbf{x}^*)}$ cannot be very small for every j, since then every j would have accepted a small reduction in x_i instead of waiting for $\mathbf{x}^*$. However, the condition does not rule out that x_i^* can be above i's preferred level: if j anticipates $\mathbf{x}^* \gg 0$, then j will reject a smaller x_i whenever x_i^* is so small that $\frac{\partial U_j(\mathbf{x}^*)/\partial x_i}{U_j(\mathbf{x}^*)}$ is larger than ρ_j/χ_i.

Theorem B-1. *Consider a situation with no uncertainty. If $U_i(\mathbf{x}^*) > 0 \forall i \in N$, $\mathbf{x}^*$ can be a part of a nontrivial SSPE if and only if for every $i \in N$, there exists some $j \neq i$ such* *that:*

$$1 < \frac{\partial U_j(\mathbf{x}^*)/\partial x_i}{\rho_j U_j(\mathbf{x}^*)}\chi_i. \tag{24}$$

Intuitively, with P&R bargaining, i is willing to contribute beyond i's bliss point if j is willing to reject a reduction of Δ_i^x. The number Δ_i^x can be arbitrarily close to zero if also Δ is close to zero. For any $\chi_i \equiv \Delta_i^x/\Delta \in (0,\infty)$, the right-hand side of (24) will grow when the contributions fall since then $\partial U_j(\mathbf{x}^*)/\partial x_i$ grows while $U_j(\mathbf{x}^*)$ approaches zero. For sufficiently small (but positive) contributions, (24) holds.

Uncertainty under alternative assumptions.—Section 3 assumed $\rho_{j,t} = \theta_{j,t}^{\rho}\rho_j$ (although the shock $\theta_{j,t}^{\rho}$ was then referred to as $\theta_{j,t}$). We might also consider the possibility that j's expectation over the lag before the next acceptance stage is $\Delta_{j,t} = \theta_{j,t}^{D}\Delta$, where Δ is the common mean for this expectation, while $\theta_{j,t}^{D}$ is a shock with mean 1. This shock might capture a situation in which the delay or lag before the next proposal stage is unknown and different parties obtain different subjective beliefs regarding what the lag will be. Similarly, with a stochastic $\frac{\partial U_j^{\theta}(\mathbf{x}^*)/\partial x_i}{U_j^{\theta}(\mathbf{x}^*)}$, suppose we can write $\frac{\partial U_j^{\theta}(\mathbf{x}^*)/\partial x_i}{U_j^{\theta}(\mathbf{x}^*)} = \frac{\partial U_j(\mathbf{x}^*)/\partial x_i}{U_j(\mathbf{x}^*)}/\theta_{j,t}^{U}$, where $\mathrm{E}\left(1/\theta_{j,t}^{U}\right) = 1$. Here, $\theta_{j,t}^{U}$ can be interpreted as a shock that influences j marginal utility of x_i, j's absolute level of utility, or both. All shocks are realized and observed after offers but before acceptance decisions are made, and all shocks are i.i.d. over time.[25] As will be shown in the proof below, the rejection condition becomes uncertain in the presence of *any* of these three shocks (or with two or all three of them): of importance is the product of the three shocks:

$$\theta_{j,t} \equiv \theta_{j,t}^{\rho}\theta_{j,t}^{D}\theta_{j,t}^{U}.$$

The $\theta_{j,t}$'s are assumed to be jointly distributed according to F, as before. Clearly, the support of $\theta_{j,t}$ will include zero as long as zero is included in the support of at least one of the three shocks. I will say that there is no uncertainty if every $\theta_{j,t}^{\rho}$, $\theta_{j,t}^{D}$, and $\theta_{j,t}^{U}$ is deterministic.

[25] Admittedly, the sources of the various shocks are here simply black boxes. A more serious future investigation should provide a careful micro-foundation for the shocks and relate them to the primitives of the model as well as to real-world evidence.

The condition under which j rejects, (23), can now be written as:

$$\theta_{j,t} \equiv \theta_{j,t}^{\rho}\theta_{j,t}^{D}\theta_{j,t}^{U} < \widetilde{\theta}_j\left(\mathbf{x}^i\right) \equiv \frac{\partial U_j\left(\mathbf{x}^*\right)/\partial x_i}{\rho_j U_j\left(\mathbf{x}^*\right)}\chi_i, \tag{25}$$

replacing (13). With this definition of $\widetilde{\theta}_j\left(\mathbf{x}^i\right)$, we can define the cdf G just as in (14). This G, which is the probability that (25) fails for every j (i.e., everyone accepts the deviation $\mathbf{x}^i$), is clearly a function of Δ_i^x. Write this function as $G_{i,\mathbf{x}^*}\left(\Delta_i^x\right)$.

As in the proof of Theorem A-1, i seeks to maximize (15). For $\mathbf{x}^*$ to be part of an SSPE, i cannot benefit from proposing marginally less. Party i does not benefit from offering the marginal amount Δ_i^x less if and only if:

$$\begin{aligned}
&\mathrm{E}\left[U_i\left(\mathbf{x}^*\right) - \left(\partial U_i\left(\mathbf{x}^*\right)/\partial x_i\right)\Delta_i^x\right] G_{i,\mathbf{x}^*}\left(\Delta_i^x\right) + \left(1 - G_{i,\mathbf{x}^*}\left(\Delta_i^x\right)\right) U_i\left(\mathbf{x}^*\right)\left(1 - \rho_{i,t}\Delta_{i,t}\right) \\
&\qquad < \mathrm{E}U_i\left(\mathbf{x}^*\right) \Leftrightarrow \\
&\left(-\frac{\partial U_i\left(\mathbf{x}^*\right)/\partial x_i}{U_i\left(\mathbf{x}^*\right)\rho_i}\right)\chi_i \leq \frac{1 - G_{i,\mathbf{x}^*}\left(\Delta_i^x\right)}{G_{i,\mathbf{x}^*}\left(\Delta_i^x\right)}.
\end{aligned} \tag{26}$$

The right-hand side of (26) is a positive number when $\chi_i > 0$ as long as it is possible that $\theta_{j,t}$ is small enough to satisfy (25) for some $j \neq i$. When all contributions fall, the right-hand side of (25) increases and approaches infinity when $U_j\left(\mathbf{x}^*\right) \to 0$, so naturally (25) will be satisfied before all contributions are zero.

Theorem B-2. *Consider a situation with uncertainty and a nontrivial SSPE in which* $U_i\left(\mathbf{x}^*\right) > 0 \forall i$. *For every* $i \in N$, *(26) holds.*

As in Section 4.4 and Theorem 3, we can impose trembling-hand perfection to show that the inequality in (26) must bind in a locally perfect SSPE.

Theorem B-2 limits how large the contributions can be. However, strictly positive contributions can be supported in equilibrium for the same reason as in Section 4: Any deviation by i may be rejected by one of the opponents with a sufficiently large probability. As above, $\Delta_{i,t}$ can be arbitrarily small if also Δ is small. Intuitively, if the contributions and payoffs are small, it doesn't take much for a party to reject an offer if the party, in return, can expect a marginally better offer quite soon. Thus, the threshold $\widetilde{\theta}_j$ is strictly positive and it does not approach zero even if $\Delta_i^x \to 0$, if just $\chi_i \equiv \Delta_i^x/\Delta > 0$. On the contrary, if $\chi_i > 0$, $\widetilde{\theta}_j$ grows without bounds when contributions and payoffs become small.

These results prove that the qualitative result of Section 4—that P&R bargaining can lead to positive contributions—does not hinge on the assumption that the discount rate can be arbitrarily close to zero. However, the assumptions in Section 3 are helpful because the outcome simplifies and it can be related to the asymmetric NBS in a way that is not possible under the alternative assumptions considered here.

To see this, the proof below shows that a second-order Taylor expansion of the right-hand side of (26) implies:

$$\begin{aligned}
-\frac{\partial U_i(\mathbf{x}^*)/\partial x_i}{U_i(\mathbf{x}^*)\rho_i} \leq & \sum_{j\neq i} f_j(\mathbf{0}) \frac{\partial U_j(\mathbf{x}^*)/\partial x_i}{\rho_j U_j(\mathbf{x}^*)} \\
& + \frac{\chi_i}{2} \sum_{j\neq i}\sum_{k\neq i} \frac{\partial f_j(\mathbf{0})}{\partial \widetilde{\theta}_k} \left(\frac{\partial U_j(\mathbf{x}^*)/\partial x_i}{\rho_j U_j(\mathbf{x}^*)}\right)\left(\frac{\partial U_k(\mathbf{x}^*)/\partial x_i}{\rho_k U_k(\mathbf{x}^*)}\right) \\
& + \chi_i \left(\sum_{j\neq i} \frac{\partial f_j(\mathbf{0})}{\partial \widetilde{\theta}_j} \frac{\partial U_j(\mathbf{x}^*)/\partial x_i/}{\rho_j U_j(\mathbf{x}^*)}\right)^2.
\end{aligned}$$

If $\chi_i \to 0$, the last two terms are zero and we are left with the same condition as in Theorem A-1(a). If instead $f_j(\mathbf{0}) \to 0$, the first term on the right-hand side is zero. The second term is zero if shocks are uncorrelated, and, in that case, we are left with the final term. The inequality can then be written as:

$$-\frac{\partial U_i(\mathbf{x}^*)/\partial x_i}{U_i(\mathbf{x}^*)\rho_i} \leq \chi_i \left(\sum_{j\neq i} \frac{\partial f_j(\mathbf{0})}{\partial \widetilde{\theta}_j} \frac{\partial U_j(\mathbf{x}^*)/\partial x_i/}{\rho_j U_j(\mathbf{x}^*)}\right)^2.$$

Here, the right-hand side is positive (and positive contributions can be supported) even if $f_j(\mathbf{0}) = 0$ if just $\partial f_j(\mathbf{0})/\partial\widetilde{\theta}_j > 0$. The fact that the term on the right-hand side is quadratic implies that the outcome cannot easily be related to the asymmetric NBS.

Corollary B-1. *Consider Example E with parameters, $U_i(\mathbf{x}) = \alpha \sum_{j\neq i} x_j - \beta x_i^2/2$, and symmetric $\chi_i = \chi$.*

(i) Suppose there is no uncertainty. Symmetric positive contributions x_i^* can be a part of a

nontrivial SSPE if and only if:

$$x_i^* \in \left(0, \frac{\alpha(n-1)}{\beta} - \frac{\alpha}{\beta}\sqrt{(n-1)^2 - 2\beta\chi/\alpha\rho}\right).$$

(ii) Suppose there is uncertainty and $\chi \to 0$*. If* $x_i^* > 0$ *is a part of a symmetric nontrivial SSPE in which* $U_i(\mathbf{x}^*) > 0 \forall i$*, then:*

$$x_i^* \leq (n-1)\frac{\alpha}{\beta} f^{\theta}(\mathbf{0}).$$

(iii) Suppose there is uncertainty, shocks are uncorrelated, and $f_j(0) \to 0 \forall j$*. If* $x_i^* > 0$ *is part of a symmetric nontrivial SSPE in which* $U_i(\mathbf{x}^*) > 0 \forall i$*, then the second-order Taylor approximation of (26) implies:*

$$x_i^* \leq (n-1)\frac{\partial f_j(0)}{\partial \theta_j}\frac{\chi\alpha^2}{2\beta\rho}$$

The comparative static w.r.t. the mean discount rates, for example, is the same as in Section 4. The above inequalities also give a new comparative static: If χ is larger (so that the time lag Δ goes to zero very fast relative to how finely one can set x_i), then the upper boundary for the thresholds is larger.

Part (ii) suggests that Theorem 2 may continue to hold if $\chi \to 0$ (as in the main text) when $f_j(0) > 0$, even if the shock $\theta_{j,t}$ can be derived from the alternative sources, as defined in (25). The proof below confirms this to be the case.

Corollary B-2. *Suppose* $\chi \to 0$ *and* $f_j(0) > 0 \forall j \in N$*. Theorems 2, 3, and 4 continue to hold with* $\theta_{j,t}$ *defined by (25).*

Proofs of Theorem B-2, Corollary B-1, and Corollary B-2

From (14) we can define:

$$G'_{i,\mathbf{x}^*} \equiv \frac{dG_{i,\mathbf{x}^*}(0)}{d\Delta_i^x} = \sum_{j\neq i} \frac{\partial G(\mathbf{0})}{\partial \widetilde{\theta}_j} \frac{\partial U_j(\mathbf{x}^*)/\partial x_i}{\rho_j \Delta U_j(\mathbf{x}^*)}, \text{ and} \tag{27}$$

$$G''_{i,\mathbf{x}^*} \equiv \frac{d^2 G_{i,\mathbf{x}^*}(0)}{(d\Delta_i^x)^2} = \sum_{j\neq i} \frac{\partial G(\mathbf{0})}{\partial \widetilde{\theta}_j} \frac{\partial^2 U_j(\mathbf{x}^*)/(\partial x_i)^2}{\rho_j \Delta U_j(\mathbf{x}^*)} + \sum_{j\neq i}\sum_{k\neq i} \frac{\partial^2 G(\mathbf{0})}{\partial \widetilde{\theta}_j \partial \widetilde{\theta}_k} \left(\frac{\partial U_j(\mathbf{x}^*)/\partial x_i}{\rho_j \Delta U_j(\mathbf{x}^*)}\right)\left(\frac{\partial U_k(\mathbf{x}^*)/\partial x_i}{\rho_k \Delta U_k(\mathbf{x}^*)}\right).$$

Consider a second-order Taylor expansion of the right-hand side of (26), $\frac{1-G}{G}$. To derive this, note that:

$$\frac{d}{d\Delta_i^x}\left(\frac{1-G}{G}\right) = \frac{-G'G-(1-G)G'}{G^2} = \frac{-G'}{G^2}, \text{ and}$$

$$\frac{d^2}{(d\Delta_i^x)^2}\left(\frac{1-G}{G}\right) = \frac{-G''G^2+2G'G'G}{G^4} = \frac{-G''+2G'G'/G}{G^2}.$$

Therefore, the second-order Taylor expansion of the right-hand side of (26) is given by:

$$\frac{1-G_{i,\mathbf{x}^*}(\Delta_i^x)}{G_{i,\mathbf{x}^*}(\Delta_i^x)} \approx \frac{1-G_{i,\mathbf{x}^*}(0)}{G_{i,\mathbf{x}^*}(0)} + \frac{-G'_{i,\mathbf{x}^*}}{(G_{i,\mathbf{x}^*}(0))^2}\Delta_i^x + \frac{(\Delta_i^x)^2}{2}\left(\frac{-G''_{i,\mathbf{x}^*}+2\left(G'_{i,\mathbf{x}^*}\right)^2/G_{i,\mathbf{x}^*}(0)}{(G_{i,\mathbf{x}^*}(0))^2}\right).$$

The first term is zero since $G_{i,\mathbf{x}^*}(0) = 1$. If we substitute in for $G'_{i,\mathbf{x}^*}$ and $G''_{i,\mathbf{x}^*}$ using (27), we get:

$$\frac{1-G_{i,\mathbf{x}^*}(\Delta_i^x)}{G_{i,\mathbf{x}^*}(\Delta_i^x)} \approx -\Delta_i^x \sum_{j\neq i} \frac{\partial G(\mathbf{0})}{\partial \widetilde{\theta}_j} \frac{\mathrm{E}(\partial U_j(\mathbf{x}^*)/\partial x_i)/U_j(\mathbf{x}^*)}{\rho_j \Delta} \tag{28}$$

$$- \frac{(\Delta_i^x)^2}{2} \sum_{j\neq i} \frac{\partial G(\mathbf{0})}{\partial \widetilde{\theta}_j} \frac{\mathrm{E}\left(\partial^2 U_j(\mathbf{x}^*)/(\partial x_i)^2\right)/U_j(\mathbf{x}^*)}{\rho_j \Delta}$$

$$- \frac{(\Delta_i^x)^2}{2} \sum_{j\neq i}\sum_{k\neq i} \frac{\partial^2 G(\mathbf{0})}{\partial \widetilde{\theta}_j \partial \widetilde{\theta}_k} \left(\frac{\mathrm{E}(\partial U_j(\mathbf{x}^*)/\partial x_i)/U_j(\mathbf{x}^*)}{\rho_j \Delta}\right)\left(\frac{\mathrm{E}(\partial U_k(\mathbf{x}^*)/\partial x_i)/U_k(\mathbf{x}^*)}{\rho_k \Delta}\right)$$

$$+ (\Delta_i^x)^2 \left(\sum_{j\neq i} \frac{\partial G(\mathbf{0})}{\partial \widetilde{\theta}_j} \frac{\mathrm{E}(\partial U_j(\mathbf{x}^*)/\partial x_i)/U_j(\mathbf{x}^*)}{\rho_j \Delta}\right)^2.$$

Note that the second term is zero when $\Delta_i^x \to 0$, even if $\Delta_i^x/\Delta \to \chi_i > 0$.

If $\Delta_i^x/\Delta \to 0$, the third and fourth terms in (28) also become zero, so we are left with only the first term. When this term is substituted into (26), we arrive at

$$-\frac{\partial U_i(\mathbf{x}^*)/\partial x_i}{U_i(\mathbf{x}^*)\rho_i} \leq \sum_{j\neq i}\left(-\frac{\partial G(\mathbf{0})}{\partial\widetilde{\theta}_j}\right)\frac{\mathrm{E}\left(\partial U_j(\mathbf{x}^*)/\partial x_i\right)/U_j(\mathbf{x}^*)}{\rho_j},$$

which is the same condition as in Theorem A-2(a) since $-\frac{\partial G(\mathbf{0})}{\partial\widetilde{\theta}_j} = f_j(0)$. This implies that Theorem 2 continues to hold in this case, as claimed by Corollary B-2. The fact that Theorems 3 and 4 hold, as well, follows because the proofs of Theorems 3 and 4 are unchanged even though the definition of $\theta_{j,t}$ is changed. The proof of Corollary B-1, part (ii), follows straightforwardly.

If instead $-\frac{\partial G(\mathbf{0})}{\partial\widetilde{\theta}_j} = f_j(0) \approx 0$, so that the density of the shocks on $\widetilde{\theta}_j$ is zero when $\widetilde{\theta}_j \to 0$, then the first and fourth terms in (28) become zero, and we are left with only the third term. When we substitute this term into (26), and divide both sides by $\frac{\Delta_i^x}{\Delta}$, (26) becomes:

$$-\frac{\partial U_i(\mathbf{x}^*)/\partial x_i}{U_i(\mathbf{x}^*)\rho_i} \leq \frac{\chi_i}{2}\sum_{j\neq i}\sum_{k\neq i}\left(-\frac{\partial^2 G(\mathbf{0})}{\partial\widetilde{\theta}_j\partial\widetilde{\theta}_k}\right)\left(\frac{\partial U_j(\mathbf{x}^*)/\partial x_i}{\rho_j U_j(\mathbf{x}^*)}\right)\left(\frac{\partial U_k(\mathbf{x}^*)/\partial x_i}{\rho_k U_k(\mathbf{x}^*)}\right),$$

where $-\frac{\partial^2 G(\mathbf{0})}{\partial\widetilde{\theta}_j\partial\widetilde{\theta}_k} = \frac{\partial f_j(0)}{\partial\theta_k}$. If the shocks are not correlated, $\frac{\partial f_j(0)}{\partial\theta_k} = 0$ when $k \neq j$, and this inequality simplifies to:

$$-\frac{\partial U_i(\mathbf{x}^*)/\partial x_i}{U_i(\mathbf{x}^*)\rho_i} \leq \sum_{j\neq i} f_j'(0)\left(\frac{\mathrm{E}\left(\partial U_j(\mathbf{x}^*)/\partial x_i\right)/U_j(\mathbf{x}^*)}{\rho_j}\right)^2\frac{\chi_i}{2},$$

where the right-hand side is positive when some $f_j(\theta_j)$ is strictly convex at $\theta_j = 0$. When this inequality is combined with $U_i(\mathbf{x}) = \alpha\sum_{j\neq i} x_j - \beta x_i^2/2$, it can be rewritten to Corollary B-1. $\|$