Last-Iterate Convergence of No-Regret Learning for Equilibria in Bargaining Games

Serafina Kamp, Reese Liebman, Ben Fish

Abstract

Bargaining games, where agents attempt to agree on how to split a total utility, are an important class of games used to study economic behavior, which motivates a study of online learning algorithms in these games. In this work, we tackle when no-regret learning algorithms converge to Nash equilibria in bargaining games. While recent results have shown that online algorithms related to Follow the Regularized Leader (FTRL) converge to Nash equilibria (NE) in a wide variety of games, including zero-sum games, this does not include bargaining games. Because of the possibility of take-it-or-leave-it offers, bargaining games are not zero-sum or the like. This includes the ultimatum game, which features a single take-it-or-leave-it offer. Nonetheless, we establish that FTRL (without the modifications necessary for zero-sum games) achieves last-iterate convergence to an approximate NE in the ultimatum game. Further, we provide experimental results to demonstrate that asymmetric initial conditions can cause convergence to asymmetric NE, both in the ultimatum game and in bargaining games with multiple rounds. In doing so, this work demonstrates how complex economic behavior (e.g. learning to use threats and the existence of many possible outcomes) can result from using a simple learning algorithm, and that FTRL can converge to equilibria in a more diverse set of games than previously known.

1 Introduction

Bargaining games are an important class of games that have implications across a range of economic behavior, including price setting, wage setting, and firm interactions in a market (Korenok and Munro, 2021; Feri and Gantner, 2011; Prasad et al., 2019). Further, the Ultimatum game is a kind of bargaining game whose variations have been studied extensively to understand fairness norms (Debove et al., 2016; Falk and Fischbacher, 2006; Nowak et al., 2000; Rand et al., 2013; Thaler, 1988). So, it is important to understand how agents learn to play strategies in these game, and we consider the setting where agents learn bargaining strategies online. However, bargaining games inherently involve non-convex utility functions due to discontinuities of deal breakdowns and this makes learning, let alone online learning, difficult in general.

The goal for a successful online learning algorithm in this setting is to achieve no-regret for a single agent learning and to achieve last-iterate convergence to Nash equilibrium when multiple agents are learning. The no-regret guarantee provides a motivation for why an agent would use such a procedure to choose a strategy in the first place while the last-iterate convergence guarantee gives a realistic guarantee for how agents would use a strategy they learn online. The online algorithm we use is the popular no-regret algorithm Follow-the-Regularized-Leader (FTRL) (Shalev-Shwartz et al., 2012). In particular, FTRL and its variants have been previously been used to establish last-iterate convergence results in a variety of games, including monotone games, non-negative regret games, and strictly variationally stable games (Anagnostides et al., 2022; Giannou et al., 2021; Hsieh et al., 2021; Vlatakis-Gkaragkounis et al., 2020).

Brgaining games do not have the properties previously necessary to prove last-iterate convergence in any online learning setting: They do not have concave utility functions, are not zero-sum, not equivalent to a potential game, and need not have strict Nash equilibrium solution concepts which implies they are sometimes degenerate and not strictly variationally stable. Further, bargaining games have infinitely many Nash equilibria [1] and even sometimes infinitely many subgame perfect equilibria in some extensive form multi-round bargaining games (Ponsati and Sákovics, 1998). Thus, there are two interesting questions to be answered for online learning in bargaining games.

- Does FTRL converge in the last-iterate to any kind of Nash equilibrium in bargaining games?
- Are there multiple different Nash equilibrium outcomes FTRL converges to?

In this work, we consider two kinds of bargaining games: the Ultimatum game and a 2 round alternating bargaining game. Previous work shows the possibility of last-iterate convergence to Nash equilibrium for the Ultimatum game under FTRL with ℓ_1 norm and particular learning rates (Kamp and Fish, 2024). However, this version of FTRL is not no-regret, so we provide the following stronger guarantees which answer both of our questions in the affirmative:

- FTRL with the Euclidean regularizer, any learning rate, and any initial strategy choice achieves last-iterate convergence to an ϵ-Nash equilibrium in the normal form Ultimatum game.
- Experiments reveal FTRL converges to a variety of Nash equilibria in both the normal form Ultimatum game and the 2-round alternating bargaining game in the extensive form.

The first implication of these results is that FTRL may enjoy stronger guarantees of last-iterate convergence given the difference in properties of bargaining games and previous classes of games where convergence results hold. Further,

[1] See (Osborne, 1990) for an overview of Bargaining games.

since FTRL simultaneously achieves no-regret for a single learner and last-iterate convergence to a Nash equilibrium in at least one kind of bargaining game, this opens the possibility for strong learning guarantees in other variations of bargaining games.

Next, we observe that FTRL converges to many different Nash equilibrium outcomes, depending on the initial conditions of the algorithm. In some settings, such as wage or price setting, there are important fairness concerns for which Nash equilibrium agents play at. In particular, any asymmetric Nash equilibrium outcome where equal-merit agents are getting different payoffs can potentially be considered an unfair, yet stable outcome (Fish and Stark, 2022; Kamp and Fish, 2024). Given this concern along with the aforementioned interest in understanding the multiplicity of outcomes in real world experiments of the Ultimatum game, there is value in our results since we make progress in describing *how* agents choose bargaining strategies. To highlight the importance of these concerns, we choose a model which simulates a wage negotiation process between a single firm f and a single worker w. The agents bargain over the split of the surplus generated by the worker's employment normalized to 1 and we assume both agents are equally entitled to the surplus. Additionally, we interpret our experimental results through this lens to demonstrate how FTRL allows for threat-like behavior to develop while learning in order for one agent to improve their payoff compared to other equilibrium outcomes.

In Section 2, we review related work and highlight how our setting is different from previous last-iterate convergence guarantees. In Section 3 we introduce our model and the bargaining games we consider and in Section 4 we introduce the learning setting for these games. In Section 5, we present our theoretical contribution and in Section 6 we discuss experimental results. Finally, we conclude with a discussion and future work in Section 7.

2 Related Work

There is extensive literature on online learning in games, and we provide comparisons to only a select few papers to highlight the relevant differences between bargaining games and previous classes of games that have last-iterate Nash equilibrium convergence guarantees. To start, we prove convergence to a *mixed* Nash equilibrium in our theoretical results, but since our game is degenerate the impossibility result of Vlatakis-Gkaragkounis et al. (2020) does preclude our results.

Next, the bargaining game we consider is not strictly variationally stable (Azizian et al., 2021; Hsieh et al., 2021; Mertikopoulos and Zhou, 2019), does not have strict Nash equilibria (Giannou et al., 2021; Vlatakis-Gkaragkounis et al., 2020), is not zero-sum (Cai et al., 2024; Gilpin et al., 2012), is not monotone (Cai et al., 2022), is not a non-negative regret game or equivalent to a potential game (Anagnostides et al., 2022), and is not an auction game (Deng et al., 2022).

Finally, previous work shows convergence of weakly acyclic games (which includes our bargaining game) to Nash equilibrium (Marden et al., 2007) and convergence of the normal form Ultimatum game to Nash equilibrium under FTRL with an ℓ_1 regularizer and particular learning rates(Kamp and Fish, 2024). However, both works use algorithms that are not no-regret, so our result that FTRL can simultaneously get no-regret and convergence in the last-iterate to a Nash equilibrium is quite stronger.

3 Bargaining Games

The setting we consider is a bargaining game between a single firm f and a single worker w. We assume the agents are bargaining over the split of a surplus normalized to 1. We assume throughout that the firm is always the first to propose a surplus split and the worker is always the first to respond. The action set of the proposing agent is to make an offer to the responding agent from the set $\mathcal{A} = [0, 1]$. The action set of the responding agent is to specify whether they would accept or reject each possible offer. The payoff to the agents is given as a tuple (u_f, u_w) where u_f is the payoff to the firm and u_w is the payoff to the worker.

We consider two versions of the bargaining game: The Ultimatum game in the normal form and the 2-round alternating bargaining game in the extensive form. In this section, we introduce both games with continuous action sets, and in Section 4 we describe the convex version of each game used for learning.

3.1 Normal Form Ultimatum Game

In the Ultimatum game [2], the firm makes an offer $a \in \mathcal{A}$ and the worker can either accept a or reject a. If the worker accepts, the payoff to the agents is $(1 - a, a)$, and if the worker rejects, the payoff to the agents is $(0, 0)$. In the normal form version of the game, the agents specify their actions simultaneously. In this version, we assume the firm still chooses an offer $a_f \in \mathcal{A}$, but we now assume the worker chooses an acceptance threshold $a_w \in \mathcal{A}$ which specifies the lower bound on offers they are willing to accept. We will refer to the strategy profile of the agents as a tuple specifying each agent's action: (a_f, a_w). Finally, the utility functions of the agents are

$$\begin{aligned} u_f(a_f, a_w) &= (1 - a_f) \cdot \mathbf{1}\{a_w \leq a_f\}, \\ u_w(a_f, a_w) &= a_f \cdot \mathbf{1}\{a_w \leq a_f\}. \end{aligned}$$

There are infinitely many Nash equilibria for this game. For each $a \in \mathcal{A}$, the strategy profile (a, a) is in Nash Equilibrium. Given the worker's acceptance threshold, the firm gets the most utility by making the lowest possible offer that will get accepted, and, given the firm's offer, the worker gets equal utility from any acceptance threshold at or below this offer. As a result, the Nash

[2] See Tadelis (2013) for an overview of variations of the Ultimatum game.

equilibria are not *strict*, i.e., for an offer $a_f > 0$,

$$u_w(a_f, a_w) = a_f, \forall a_w \leq a_f, a_w \in \mathcal{A}.$$

There are also mixed Nash equilibria in this game which follows the structure of the firm making a pure offer $a_f \in \mathcal{A}$ and the worker mixing over acceptance thresholds $a_w \in \mathcal{A}$ where the largest acceptance threshold the worker plays with non-zero probability is a_f. Notably, in order to be in Nash equilibrium, the worker must be playing the acceptance threshold a_f with sufficiently high probability to prevent the firm from preferring a lower offer. We will define this mixed Nash equilibrium in detail when we introduce the convex version of this game in Section 4.

Finally, it is of note that in the sequential version of the game there is a unique subgame perfect equilibrium where the worker would accept any offer greater than 0, so the firm proposes the lowest possible non-zero offer. However, we are interested in the conditions that lead to convergence to different equilibria, especially given the divergence from the subgame perfect outcome in real-world experiments of the Ultimatum game (Debove et al., 2016), so our results focus on the normal form version of the game.

3.2 2-Round Alternating Bargaining Game in the Extensive Form

For our experimental results, we also consider a 2-round alternating bargaining game in the extensive form with complete information and perfect recall. Actions are now performed sequentially instead of simultaneously and the agents take turns making offers and responding to offers. Here, the firm makes the first offer $a_f \in \mathcal{A}$. Then, the worker either accepts or rejects the offer. If the offer is accepted, the agents receive the payoff $(1 - a_f, a_f)$. Otherwise, the agents switch roles and the worker now makes a counter-offer $a_w \in \mathcal{A}$ to which the firm can either accept or reject. A time discount factor $0 < \delta < 1$ is applied to payoffs in the second round. So, if the firm accepts the counter offer the agents receive the payoff $\delta(a_w, 1 - a_w)$, and if the firm rejects the counter offer the agents receive $\delta(0, 0)$. The extensive form game tree is provided in Figure 1. Note for space we condense the nodes of the worker responding to the offer a_f and counter offering some a_w if they reject.

4 Learning Bargaining Strategies Online

In this paper, we are interested in how agents learn to play strategies in the kinds of bargaining games described in the previous section. Online learning is a useful framework for this problem because, here, agents update their strategies based on the utility feedback they see from their previous actions and the actions of their opponent. Further, there exist *no-regret* algorithms where the strategy an agent learn online gets as much utility, on average, as the best-in-hindsight strategy. Formally, let $a_i^{(t)} \in \mathcal{A}$ be the action agent i took at time t and let

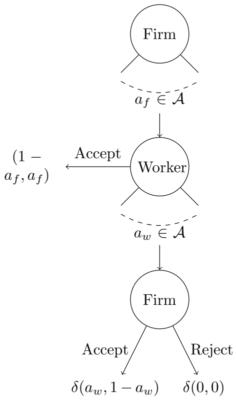


Figure 1: The extensive form game tree for the 2-round alternating bargaining game.

$u_i^{(t)}(a)$ be the utility agent i receives at time step t when playing action a. Then, the regret of an algorithm after T time steps is

$$\texttt{Regret}_T = \arg\max_{a \in \mathcal{A}} \sum_{t=1}^{T} u_i^{(t)}(a) - \sum_{t=1}^{T} u_i^{(t)}(a_i^{(t)}).$$

An algorithm is said to be no-regret if $\texttt{Regret}_T$ is sublinear in T for any arbitrary sequence of utility feedback functions $u_i^{(1)}, \ldots, u_i^{(T)}$ drawn from a class of utility feedback functions.

Online learning algorithms are particularly well-suited for convex optimization problems [3]. However, the utility functions in each of our games are non-convex, so we discretize our action set to use the convex expected utility function instead. Notably, previous work has shown that no-regret strategies for the discretized normal form Ultimatum game get no-regret with respect to the best action in hindsight from the original continuous space (Kamp and Fish, 2024).

4.1 Convex Game Representation

4.1.1 Normal Form Ultimatum Game with Mixed Strategies

We now consider the action space $[0, 1]$ discretized by an integer $D > 1$, i.e., $\mathcal{A} = \{0, \frac{1}{D}, \ldots, \frac{D-1}{D}, 1\}$. Let $\mathcal{G}^{(1)}$ be the normal form Ultimatum game with such a discreteized action set. The agents then learn mixed strategies over $\Delta(\mathcal{A})$. Let $x_i^{(t)} \in \Delta(\mathcal{A})$ be the mixed strategy of agent i at time t for $i \in \{f, w\}$ and let $x_{i,a}^{(t)}$ be the probability mass agent i puts on action a at time t. Further, for all $a \in \mathcal{A}$, let $\mathbf{1}_a$ be a pure strategy of action a. Next, let $U_{i,a}^{(t)}$ be the cumulative

[3] See Hazan et al. (2016) for an overview of online convex optimization.

payoff to agent i through time t when they play a pure action a, given the mixed strategy history of agent $-i$, the other agent, i.e.,

$$U_{i,a}^{(t)}(\{x_{-i}^{(\tau)}\}_{\tau=1}^{t}) = \sum_{\tau=1}^{t} u_i(a, x_{-i}^{(\tau)}).$$

where

$$\begin{aligned} u_f(a, x_w) &= \mathop{\mathbb{E}}_{a_r \sim x_w}[(1-a)\mathbf{1}\{a_r \leq a\}], \\ u_w(x_f, a) &= \mathop{\mathbb{E}}_{a_p \sim x_f}[a_p \mathbf{1}\{a_p \geq a\}]. \end{aligned}$$

Unless otherwise specified, we will omit the input history $\{x_{-i}^{(\tau)}\}_{\tau=1}^{t}$ from the notation and refer to the cumulative payoff for a specific action through time t as $U_{i,a}^{(t)}$. Finally, we will take $U_i^{(t)}$ to be the cumulative payoff vector of all actions $a \in \mathcal{A}$.

The pure Nash equilibria in this representation are still of the form $(\mathbf{1}_a, \mathbf{1}_a)$ for all $a \in \mathcal{A}$. Additionally, $(\mathbf{1}_0, \mathbf{1}_1)$ is a pure Nash equilibrium where both agents get 0. Finally, there is only one kind of mixed Nash equilibrium possible in this game. For $x_f, x_w \in \Delta(\mathcal{A})$, the strategy profile (x_f, x_w) is in mixed Nash equilibrium if $x_f = \mathbf{1}_{a_f}$ for $a_f \in \mathcal{A}$, $\max\{a_w | x_{w,a_w} > 0\} = a_f$, and

$$(1 - a_f) \geq (1 - a) \cdot \sum_{a_w \leq a} x_{w,a_w}, \forall a < a_f. \tag{1}$$

Here, since $\max\{a_w | x_{w,a_w} > 0\} = a_f$, then the worker accepts an offer of a_f with probability 1, so the expected utility to the firm for an offer of a_f is $(1 - a_f)$. Any higher offer would also be accepted with probability 1, so the firm would get strictly worse utility from making an offer higher than a_f. Further, the condition 1 ensures the firm does not get more expected utility by lowering their offer. When, $x_f = \mathbf{1}_{a_f}$ the expected utility of the worker is a_f for any distribution over acceptance thresholds $a \leq a_f$ is a_f and all other mixed strategies get strictly less than a_f. Therefore, the agents are in mixed Nash equilibrium by definition. In Section 5, we prove last-iterate convergence to an approximate mixed Nash Equilibrium of this kind.

4.1.2 2-Round Alternating Bargaining Game in Sequence Form Representation

First, we use the same offer space as above, $\mathcal{A} = \{0, \frac{1}{D}, \ldots, \frac{D-1}{D}, 1\}$. Let $\mathcal{G}^{(2)}$ be the 2-round alternating bargaining game with this action set. Let I_i for $i \in \{f, w\}$ be the information set for each agent. Since our game is complete information, note that there is exactly one node for each $I \in I_i$. Let $h_{i,p,\sigma}$ be the node where agent i is making a proposal after their opponent's previous action σ and $h_{i,r,\sigma}$ be the node where agent i is responding after their opponent's previous action σ. Then, an agent's behavioral strategy is

$$\beta_i : I \times \mathcal{A} \cup \{\text{Accept}, \text{Reject}\} \to [0, 1].$$

Then, the convex version of an extensive form game can be derived from its sequence form representation [4]. Let r_i be the realization plan of agent i mapping action sequences of player i to probability masses. Let $\mathcal{Q}_i$ be the set of valid realization plans of agent i. We abuse notation slightly and suppose $r \in \mathcal{Q}_i$ is represented as a vector of probability masses on sequences leading to payoff nodes. Then, the expected utility of a realization plan, given a cumulative expected utility vector $U_i^{(t)}$, can be denoted as $\langle U_i^{(t)}, r\rangle$. Finally, note that every realization plan has a one-to-one correspondence with a behavioral strategy.

4.2 Follow-the-Regularized-Leader

The online algorithm we consider is Follow-the-Regularized-Leader (FTRL) (Shalev-Shwartz et al., 2012). We use the standard Euclidean regularizer throughout and let $\eta > 0$ be the learning rate. First, the update step of FTRL for game $\mathcal{G}^{(1)}$ for each agent $i \in \{f, w\}$ at time t:

$$\arg\max_{x \in \Delta(\mathcal{A})} \eta\langle U_i^{(t)}, x\rangle - \frac{1}{2}\|x - \alpha_i\|_2^2 \tag{1}$$

Next, For game $\mathcal{G}^{(2)}$, the update step of FTRL for game $\mathcal{G}^{(1)}$ for each agent $i \in \{f, w\}$ at time t:

$$\arg\max_{r \in \mathcal{Q}_i} \eta\langle U_i^{(t)}, r\rangle - \frac{1}{2}\|r - \alpha_i\|_2^2 \tag{2}$$

The term α_i is the *reference point* of the regularizer. We will assume a reference point of $\alpha_i = \mathbf{0}$ throughout Section 5, but in Section 4 we experiment with a variety of reference points to demonstrate their influence on which Nash equilibrium the agents converge to.

5 Last-Iterate Convergence to ϵ-Nash Equilibrium

We are now ready to state the main result of our work. In Theorem 11, we show that, regardless of the initial conditions, agents learning bargaining strategies for $\mathcal{G}^{(1)}$ via Algorithm 1 will converge to an approximate mixed Nash equilibrium in finite time.

Theorem 11. *Suppose agents learn strategies for $\mathcal{G}^{(1)}$ using Algorithm 1 with $\alpha_i = \mathbf{0}$, any $\eta > 0, D > 2$, and arbitrary initial conditions $x_w^{(1)}, x_f^{(1)} \in \Delta(\mathcal{A})$.*

[4] See the Appendix for details of the sequence form of the 2 Round Alternating Bargaining game and see Shoham and Leyton-Brown (2008) for more details on the sequence form representation of extensive form games in general.

Then, for any $\epsilon > 0$, there exists a finite time t_ϵ where $(x_f^{(\tau)}, x_w^{(\tau)})$ is in ϵ-Nash Equilibrium for all $\tau \geq t_\epsilon$.

This strong convergence result demonstrates the promise of online learning algorithms in the area of learning bargaining strategies. Additionally, this result extends the equilibrium convergence guarantees of FTRL to a game that is degenerate, not variationally stable, and not zero-sum.

In the proof, there is one important point in each agent's strategy to track at each time t: The largest acceptance threshold that the worker plays with non-zero probability and the smallest offer the firm makes with non-zero probability, notated as follows.

$$w_{\max}^{(t)} = \max\{a | x_{w,a}^{(t)} > 0\},$$
$$f_{\min}^{(t)} = \min\{a | x_{f,a}^{(t)} > 0\}.$$

At a high-level, in Algorithm 1, the firm strictly prefers a lower offer if it will be accepted by $x_w^{(t)}$ with probability 1, i.e., the firm prefers to offer $w_{\max}^{(t)}$ than any greater offer. Further, given condition 1 of mixed NE of $\mathcal{G}^{(1)}$, the firm will also prefer to lower their offer if $x_{w,w_{\max}^{(t)}}^{(t)}$ is not sufficiently large. At the same time, any acceptance threshold the worker uses less than or equal to $f_{\min}^{(t)}$ gets equal expected utility and strictly more expected utility than any greater acceptance threshold. As a result, the cumulative utility of smaller acceptance thresholds grows comparatively more than larger acceptance thresholds, so there is an incentive for the worker to lower their acceptance threshold over time to match $f_{\min}^{(t)}$ and remain fixed if their acceptance threshold is less than $f_{\min}^{(t)}$. This structure of the worker's utility function is also sufficient to cause $w_{\max}^{(t)}$ to be non-increasing over time. So, $f_{\min}^{(t)} < w_{\max}^{(t)}$ causes $x_{w,w_{\max}^{(t)}}^{(t)}$ to decrease while $w_{\max}^{(t)} < f_{\min}^{(t)}$ causes $f_{\min}^{(t)}$ to decrease while $w_{\max}^{(t)}$ remains fixed.

The proof of Theorem 11 uses this relation between $w_{\max}^{(t)}$ and $f_{\min}^{(t)}$ to show that it takes finite time for $w_{\max}^{(t)}$ and $f_{\min}^{(t)}$ to decrease until $x_{w,w_{\max}^{(t)}}^{(t)}$ is large enough to pass condition 1. We show there always exists a time where $x_{w,w_{\max}^{(t)}}^{(t)}$ is large enough to pass condition 1 for all future time steps, and finally, that this suffices for the firm to approach the pure strategy $\mathbf{1}_{w_{\max}^{(t)}}$ in the limit.

Additionally, the last-iterate strategy profile of Algorithm 1 is also an ϵ-Nash equilibrium with respect to the normal form Ultimatum game with the action set $\mathcal{A} = [0, 1]$.

Corollary 0.1. *Suppose the last iterate strategy profile of Algorithm 1, $(x_f^{(T)}, x_w^{(T)})$, is an ϵ-Nash Equilibrium for some $\epsilon > 0$ with respect to mixed strategies over the action set $\{0, \ldots, 1\}$. Then, $(x_f^{(T)}, x_w^{(T)})$ is an ϵ-Nash Equilibrium with respect to pure strategies from the action set $[0, 1]$.*

Proof. Let $\epsilon > 0$. Suppose after running Algorithm 1 for T time steps, $(x_f^{(T)}, x_w^{(T)})$ is in ϵ-Nash Equilibrium. Then, by Theorem 11, there exists $k \in \{1, \ldots, D\}$ such that $w^{(T)}_{\max} = \frac{k}{D}$ where $x^{(T)}_{w,w^{(T)}_{\max}} \geq \frac{1}{D-k+1}$ and $x^{(T)}_{f,w^{(T)}_{\max}} \geq 1-\epsilon$. We will now show that $(x^{(T)}_f, x^{(T)}_w)$ is an ϵ-best response for the firm and worker, respectively, in the continuous game. That is, there is no action in the continuous set of actions that gets at least ϵ more utility than $x^{(T)}_f$ and $x^{(T)}_w$, respectively.

First, for the worker, by definition of their utility function, the most utility they can get at time T is from an acceptance threshold at $f^{(T)}_{\min}$, but $x^{(T)}_{f,w^{(T)}_{\max}} \geq 1-\epsilon$ implies

$$u_w(x^{(T)}_f, x^{(T)}_w) \geq u_w(x^{(T)}_f, \mathbf{1}_{f^{(T)}_{\min}}) - \epsilon,$$

Therefore, $x^{(T)}_w$ is an ϵ-best response for the worker in the continuous game as well.

Next, for the firm, for all $\ell \in \{1, \ldots, k\}$, consider an offer $a \in [0,1]$ from the continuous game where $w^{(T)}_{\max} - \frac{\ell}{D} \leq a < w^{(T)}_{\max} - \frac{\ell-1}{D}$. Then,

$$\begin{aligned} u_f(a, x^{(T)}_r) &= (1 - \sum_{i=0}^{\ell-1} x^{(T)}_{w,w^{(T)}_{\max}-\frac{i}{D}}) \cdot (1-a) \\ &\leq (1 - \sum_{i=0}^{\ell-1} x^{(T)}_{w,w^{(T)}_{\max}-\frac{i}{D}}) \cdot (1 - (w^{(T)}_{\max} - \frac{\ell}{D})) \\ &= u_f(w^{(T)}_{\max} - \frac{\ell}{D}, x^{(T)}_r) \end{aligned}$$

So, if $w^{(T)}_{\max}$ is a best response offer with respect to the offer set $\{\frac{1}{D}, \ldots, 1\}$, then, it must also be a best response with respect to the offer set $[0,1]$. Therefore, if $x^{(T)}_{f,w^{(T)}_{\max}} \geq 1-\epsilon$, then $x^{(T)}_f$ is an ϵ-best response for the firm in the continuous game as well. □

This result shows that FTRL is a powerful learning algorithm in terms of being both no-regret and converging last-iterate to approximate Nash equilibria in bargaining games. We now turn to our experimental results to expand on the implications of our theoretical result.

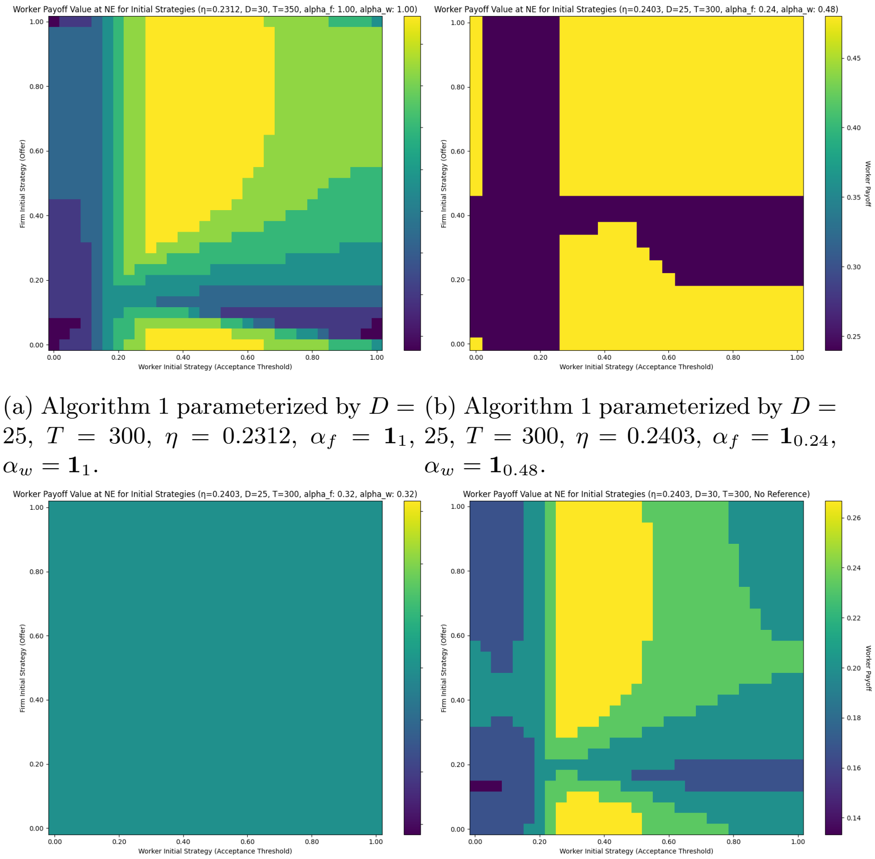
6 Experimental Results

We implemented Algorithm 1 to simulate the agents learning strategies for $\mathcal{G}^{(1)}$ and Algorithm 2 for $\mathcal{G}^{(2)}$ using CVXPY (Diamond and Boyd, 2016). The goal of our experiments is to 1) validate our theoretical findings and 2) demonstrate the link between initial conditions and the Nash equilibrium outcome.

Not only do the experiments demonstrate convergence to Nash equilibria in a variety of settings, but they also show that the algorithm converges to different Nash equilibria, depending on the initial conditions. This highlights the importance of our results: FTRL has the ability to learn no-regret and Nash equilibrium strategies, so it could also offer some explanation for how agents end up in one Nash equilibria over another. For example, we will show how credible and incredible threats may arise while learning strategies for $\mathcal{G}^{(2)}$ via Algorithm 2.

6.1 Normal Form Ultimatum Game Experiments

In these experiments, we run Algorithm 1 over a range of initial conditions and provide graph outputs displaying the payoff to the worker after the agents converge to a Nash equilibrium. For a given D value, we sweep over all possible initial pure strategy profiles $(\mathbf{1}_{a,f}, \mathbf{1}_{a_w})$ for $a_f, a_w \in \mathcal{A}$ and we use a variety of α_i settings. The results show a range of output patterns which indicates the kind of Nash equilibrium the agents converge to depends largely on the initial conditions. The results for these experiments are in Figure 2.



(a) Algorithm 1 parameterized by $D = 25$, $T = 300$, $\eta = 0.2312$, $\alpha_f = \mathbf{1}_1$, $\alpha_w = \mathbf{1}_1$.

(b) Algorithm 1 parameterized by $D = 25$, $T = 300$, $\eta = 0.2403$, $\alpha_f = \mathbf{1}_{0.24}$, $\alpha_w = \mathbf{1}_{0.48}$.

(c) Algorithm 1 parameterized by $D = 25$, $T = 300$, $\eta = 0.2403$, $\alpha_f = \mathbf{1}_{0.32}$, $\alpha_w = \mathbf{1}_{0.32}$.

(d) Algorithm 1 parameterized by $D = 30$, $T = 300$, $\eta = 0.2403$, $\alpha_f = \mathbf{0}$, $\alpha_w = \mathbf{0}$.

Figure 2: Nash equilibrium payoff outcomes for the worker when agents are learning strategies for $\mathcal{G}^{(1)}$ using Algorithm 1.

In all of our results, once we observe convergence in the strategy profiles, the firm is playing an approximately pure strategy $\mathbf{1}_a$ for some $a \in \mathcal{A}$ and the worker satisfies $w_{\max}^{(T)} = a$ with either a mixture over smaller acceptance thresholds if $a \neq \alpha_w$ or they are playing the acceptance threshold a approximately purely. The graphs show the value of a across all combinations of initial pure strategy profiles for different values of D and reference points.

In Graphs 2a and 2d, the reference points are either both set at 1 or both set at 0. Here, the Nash equilibrium outcomes depend more on the initial strategies, and the worker's payoff ranges from 0.1 to 0.3. Recall the payoff value represents the percentage split of a surplus between the firm and worker, so each 0.01 difference represents a 1% change to the worker's wage.

In Graph 2b, the worker has a larger reference point than the firm, and the payoff is one of the two values. The worker's reference point is more likely when the worker has a larger initial acceptance threshold while the firm's reference point becomes more likely when the worker has a smaller initial acceptance threshold.

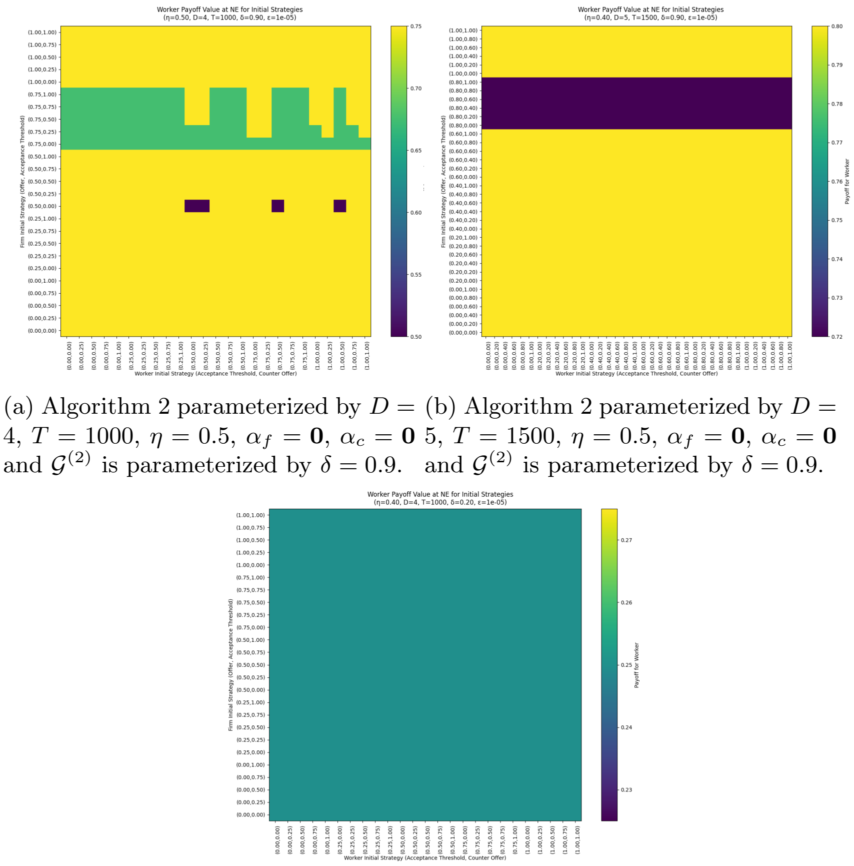
Finally, in Graph 2c, the agents both have a references point at 0.32 and they converge to this Nash equilibrium regardless of their initial strategies. Intuitively, $(\mathbf{1}_{0.32}, \mathbf{1}_{0.32})$ is a Nash equilibrium in $\mathcal{G}^{(1)}$, so when agents agree on this outcome via their reference point, then both agents maximize their objective value simultaneously when the strategy $\mathbf{1}_{0.32}$ has the most mass. Interestingly, the strategy profile $(\mathbf{1}_1, \mathbf{1}_1)$ is also a Nash equilibrium, but as demonstrated by Graph 2a, this does not imply convergence to this Nash equilibrium when $\alpha_f = \alpha_w = \mathbf{1}_1$.

6.2 2-Round Alternating Bargaining Game Experiments

In these experiments, we run Algorithm 2 over a range of initial conditions and provide graph outputs displaying the payoff to the worker after the agents converge to a Nash equilibrium. For a given D value, we sweep over a large subset of all possible initial pure strategy profiles. In particular, we consider behavioral strategy profiles characterized by the offer the agent makes as a proposer and the acceptance threshold it sets as a responder. That is, each agent i chooses $a_{i,p}, a_{i,r} \in \mathcal{A}$ and sets β_i as follows:

$$\beta_i(h, a) = \begin{cases} 1 & h = h_{i,r,\sigma}, a \geq a_{i,r}, \\ 1 & h = h_{i,p,\sigma}, a = a_{i,p}, \\ 0 & \text{otherwise} \end{cases}$$

For these experiments, we use $\alpha_f = \alpha_w = \mathbf{0}$. Further, we use $D = 4$ and $D = 5$ due to longer times until convergence. The results of our experiments are in Figure 3.



(a) Algorithm 2 parameterized by $D = 4$, $T = 1000$, $\eta = 0.5$, $\alpha_f = \mathbf{0}$, $\alpha_c = \mathbf{0}$ and $\mathcal{G}^{(2)}$ is parameterized by $\delta = 0.9$.

(b) Algorithm 2 parameterized by $D = 5$, $T = 1500$, $\eta = 0.5$, $\alpha_f = \mathbf{0}$, $\alpha_c = \mathbf{0}$ and $\mathcal{G}^{(2)}$ is parameterized by $\delta = 0.9$.

(c) Algorithm 2 parameterized by $D = 4$, $T = 1000$, $\eta = 0.4$, $\alpha_f = \mathbf{0}$, $\alpha_c = \mathbf{0}$ and $\mathcal{G}^{(2)}$ is parameterized by $\delta = 0.2$.

Figure 3: Nash equilibrium payoff outcomes for the worker when agents are learning strategies for $\mathcal{G}^{(2)}$ using Algorithm 2.

A realization plan has a one-to-one correspondence with a behavioral strategy, so we describe the outcomes based on the corresponding $\beta_i^{(T)}$ of $r_i^{(T)}$ for each agent i. Once we observe convergence of $(\beta_f^{(T)}, \beta_w^{(T)})$, we observe two kinds of Nash equilibria. in The first is an approximate pure Nash equilibrium where the firm proposes $a_f \in \mathcal{A}$ approximately purely and the worker accepts the offer approximately purely. The second is an approximate mixed Nash equilibrium is where firm mixes over a set of first offers where the worker rejects the first offer and counter offers $a_w \in \mathcal{A}$ which the firm approximately purely accepts.

First, in all cases, $\beta_f^{(T)}(h_{f,r,0}, \text{Accept}) = \frac{1}{2}$, i.e., the firm always accepts a counter offer of 0 with probability $\frac{1}{2}$. This is because the firm gets 0 utility from either accepting or rejecting an offer of 0, so the Lagrangian of the update step of Algorithm 2 requires $\beta_f^{(t)}(h_{f,r,0}, \text{Accept}) = \beta_f^{(t)}(h_{f,r,0}, \text{Reject})$ for all time steps t. Therefore, for $D > 2$, the most utility that the worker can get in round 2 is $\delta \cdot \frac{D-1}{D}$ from a counter offer of $a_w = \frac{1}{D}$. The most common outcome in all the graphs (the most frequent color in all the graphs) is when the firm approximately purely makes the smallest offer $a_f \in \mathcal{A}$ such that $\delta \cdot \frac{D-1}{D} < a_f$ to which the worker approximately purely accepts. Further, in response to all first offers $a < a_f$, the worker has indeed converged to approximately purely rejecting and counter offering $a_w = \frac{1}{D}$. Since this is off the equilibrium path, this represents a *credible threat.*

This case follows the logic of backwards induction, and the only reason we do not observe convergence to approximately pure subgame perfect equilibrium is because, by the Lagrangian of the update step of Algorithm 2, the worker only updates the probability mass values of responses to first offers that are played with non-zero probability by $r_f^{(t)}$ and otherwise they remain fixed. The same is true of the firm's update step: The firm does not update the value of their counter offer response probabilities when $r_w^{(t)}$ has 0 probability mass on making such a counter offer. As a result, this allows for the possibility of *incredible threats* in the last-iterate strategy. We highlight two cases where incredible threats cause the agents to end in a strategy profile that is 1) worse for both agents and 2) worse for the worker, but better for the firm.

In Graph 3a, case 1) occurs at outcomes where the worker is getting a payoff of 0.675 (the second most frequent color), the agents have converged to a strategy profile where the firm is mixing over the first offers $a_f \in \{0, 0.25, 0.5\}$ only and the worker is responding by rejecting and approximately purely proposing the counter offer 0.25. Note that due to the discount factor δ, this is strictly worse for both agents than the outcome where the worker accepts 0.75 in the first round. The reason this case occurs is because the initial conditions must have been set such that the firm puts 0 probability mass on the offer 0.75 such that the worker gets stuck at a strategy where they are not accepting 0.75 with high probability. Thus, the incredible threat of rejecting 0.75 leads to an outcome that is worse for both agents. The same case qualitatively happens in Graph 3b at the outcomes where the worker gets 0.72 (the least frequent color).

Finally, case 2) occurs in Graph 3a at outcomes where the worker is getting a payoff of 0.5 (the least frequent color). Here, the initial conditions equate to the firm making an incredible threat early on that they would accept a counter offer of 0.25 with low probability if the firm rejects an offer of 0.5, so the worker converges to accept the lower offer of 0.5 approximately purely.

7 Conclusion and Future Work

In Section 5, Theorem 11 establishes that FTRL has stronger equilibrium convergence guarantees than previously established since $\mathcal{G}^{(1)}$ is degenerate, not a zero-sum game, not strictly variationally stable, and yet FTRL is still guaranteed to have last-iterate convergence to an approximate mixed Nash Equilibria for any initial conditions. Additionally, our results demonstrate the applicability of no-regret algorithms to bargaining games in general. This opens a direction for future work in using these algorithms to learn strategies in more complicated bargaining games including n-round and infinite round bargaining as well as bargaining with outside options.

Further, the property of an algorithm being no-regret provides strong justification for why any individual agent would use such a procedure to learn a strategy, so this line of work could contribute to a more realistic understanding of how bargaining strategies are chosen, especially given the inconsistency of theoretical results and empirical observations of the Ultimatum game. For example, one interpretation of Graph 2c with equal reference points in each agent's regularizer could be the agreement of a relevant social norm that agents coordinate on implicitly (Roth et al., 1995). Additionally, the patterns of Graphs 2a and 2d suggest, for some reference point settings, there is a consistent relationship between the initial strategies and the Nash equilibrium that is converged to. Explicating this correlation could have implications for interpreting how opening offers influence the trajectory of a bargaining game and we leave this for future work.

Finally, this line of work also has implications for algorithmic fairness concerns. Given the variety of possible Nash equilibrium outcomes demonstrated by our empirical results, especially with asymmetric payoff outcomes, it is all the more important to study algorithms that could implicitly lead to optimal, yet discriminatory outcomes. Our work makes progress in this area by describing the dynamics of $w_{\max}$ and $f_{\min}$ that drive agents to convergence in the proof of Theorem 11. We hope to inspire future work on how algorithm design and game structures influence the kind of equilibrium an algorithm converges to.

References

Ioannis Anagnostides, Ioannis Panageas, Gabriele Farina, and Tuomas Sandholm. 2022. On last-iterate convergence beyond zero-sum games. In *International Conference on Machine Learning*. PMLR, 536–581.

Waïss Azizian, Franck Iutzeler, Jérôme Malick, and Panayotis Mertikopoulos. 2021. The last-iterate convergence rate of optimistic mirror descent in stochastic variational inequalities. In *Conference on Learning Theory*. PMLR, 326–358.

Yang Cai, Haipeng Luo, Chen-Yu Wei, and Weiqiang Zheng. 2024. Uncoupled and convergent learning in two-player zero-sum Markov games with bandit feedback. *Advances in Neural Information Processing Systems* 36 (2024).

Yang Cai, Argyris Oikonomou, and Weiqiang Zheng. 2022. Finite-time last-iterate convergence for learning in multi-player games. *Advances in Neural Information Processing Systems* 35 (2022), 33904–33919.

Stéphane Debove, Nicolas Baumard, and Jean-Baptiste André. 2016. Models of the evolution of fairness in the ultimatum game: a review and classification. *Evolution and Human Behavior* 37, 3 (2016), 245–254.

Xiaotie Deng, Xinyan Hu, Tao Lin, and Weiqiang Zheng. 2022. Nash convergence of mean-based learning algorithms in first price auctions. In *Proceedings of the ACM Web Conference 2022*. 141–150.

Steven Diamond and Stephen Boyd. 2016. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research* 17, 83 (2016), 1–5.

Armin Falk and Urs Fischbacher. 2006. A theory of reciprocity. *Games and economic behavior* 54, 2 (2006), 293–315.

Francesco Feri and Anita Gantner. 2011. Bargaining or searching for a better price?–an experimental study. *Games and Economic Behavior* 72, 2 (2011), 376–399.

Benjamin Fish and Luke Stark. 2022. It's not fairness, and it's not fair: the failure of distributional equality and the promise of relational equality in complete-information hiring games. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–15.

Angeliki Giannou, Emmanouil Vasileios Vlatakis-Gkaragkounis, and Panayotis Mertikopoulos. 2021. Survival of the strictest: Stable and unstable equilibria under regularized learning with partial information. In *Conference on Learning Theory*. PMLR, 2147–2148.

Andrew Gilpin, Javier Pena, and Tuomas Sandholm. 2012. First-order algorithm with convergence for-equilibrium in two-person zero-sum games. *Mathematical programming* 133, 1 (2012), 279–298.

Elad Hazan et al. 2016. Introduction to online convex optimization. *Foundations and Trends® in Optimization* 2, 3-4 (2016), 157–325.

Yu-Guan Hsieh, Kimon Antonakopoulos, and Panayotis Mertikopoulos. 2021. Adaptive learning in continuous games: Optimal regret bounds and convergence to nash equilibrium. In *Conference on Learning Theory*. PMLR, 2388–2422.

Serafina Kamp and Benjamin Fish. 2024. Equal Merit Does Not Imply Equality: Discrimination at Equilibrium in a Hiring Market with Symmetric Agents. *arXiv preprint arXiv:2412.15162* (2024).

Oleg Korenok and David Munro. 2021. Wage bargaining in a matching market: Experimental evidence. *Labour Economics* 73 (2021), 102078.

Jason R Marden, Gürdal Arslan, and Jeff S Shamma. 2007. Regret based dynamics: convergence in weakly acyclic games. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*. 1–8.

Panayotis Mertikopoulos and Zhengyuan Zhou. 2019. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming* 173 (2019), 465–507.

Martin A Nowak, Karen M Page, and Karl Sigmund. 2000. Fairness versus reason in the ultimatum game. *Science* 289, 5485 (2000), 1773–1775.

Martin Osborne. 1990. Bargaining and Markets.

Clara Ponsati and József Sákovics. 1998. Rubinstein bargaining with two-sided outside options. *Economic Theory* 11 (1998), 667–672.

Sanjay Prasad, Ravi Shankar, and Sreejit Roy. 2019. Impact of bargaining power on supply chain profit allocation: a game-theoretic study. *Journal of Advances in Management Research* 16, 3 (2019), 398–416.

David G Rand, Corina E Tarnita, Hisashi Ohtsuki, and Martin A Nowak. 2013. Evolution of fairness in the one-shot anonymous ultimatum game. *Proceedings of the National Academy of Sciences* 110, 7 (2013), 2581–2586.

Alvin E Roth et al. 1995. Bargaining experiments. (1995).

Shai Shalev-Shwartz et al. 2012. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning* 4, 2 (2012), 107–194.

Yoav Shoham and Kevin Leyton-Brown. 2008. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations.* Cambridge University Press.

Steven Tadelis. 2013. Game Theory; an Introduction. (2013).

Richard H Thaler. 1988. Anomalies: The ultimatum game. *Journal of economic perspectives* 2, 4 (1988), 195–206.

Emmanouil-Vasileios Vlatakis-Gkaragkounis, Lampros Flokas, Thanasis Lianeas, Panayotis Mertikopoulos, and Georgios Piliouras. 2020. No-regret learning and mixed nash equilibria: They do not mix. *Advances in Neural Information Processing Systems* 33 (2020), 1380–1391.

A $\mathcal{G}^{(1)}$ Under FTRL-NFG Always Converges to ϵ-Mixed NE

A.1 Set-up and the Lagrangian

The Lagrangian of this quadratic program for Algorithm 1 is

$$\mathcal{L}_i(x_i, \lambda_i, \mu_i) = \frac{1}{2}\|x_i\|_2^2 - \eta\langle U_i^t, x_i\rangle + \lambda_i\left(\sum_{a\in\mathcal{A}} x_{i,a} - 1\right) - \sum_{a\in\mathcal{A}} \mu_{i,a}x_{i,a}$$

The dual of this Lagrangian is

$$\max_{\lambda_i,\mu_i} \min_{x\in\mathbb{R}^{|\mathcal{A}|}} \mathcal{L}_i(x, \lambda_i, \mu_i)$$
$$\text{subject to}$$
$$\mu_i \geq 0$$

The quadratic program has strong duality by Slater's condition since the objective function is convex, the inequality constraint is convex, the equality constraint is affine, and there exists a point $x \in \Delta(\mathcal{A})$ where the equality constraint is satisfied and the inequality is strictly satisfied.

Then, by the KKT theorem, any problem that satisfies strong duality also satisfies the following KKT conditions:

- **Stationarity:** $0 \in \nabla\mathcal{L}(x, \lambda_i, \mu_i)|_{x=x^*}$ for the primal optimal x^*.
- **Primal Feasibility:** The primal constraints are satisfied for the primal optimal x^*.
- **Dual Feasibility:** $\mu_a \geq 0, \forall a \in \mathcal{A}$ for the dual optimal variables.
- **Complementary Slackness:** $\mu_a x_a^* = 0$.

Notably, by stationarity, for each $i \in \{f, w\}$ and for each $a \in \mathcal{A}$,

$$x_{i,a}^{(t+1)} = \eta U_{i,a}^{(t)} - \lambda_i + \mu_{i,a}.$$

Claim 1. *If $x_{i,a}^{(t+1)} > 0$ and $x_{i,a'}^{(t+1)} > 0$, then*

$$x_{i,a}^{(t+1)} - x_{i,a'}^{(t+1)} = \eta U_{i,a}^{(t)} - \eta U_{i,a'}^{(t)} \tag{1}$$

Proof. If $x_{i,a}^{(t+1)} > 0, x_{i,a'}^{(t+1)} > 0$, then by complementary slackness, we have $\mu_{i,a} = \mu_{i,a'} = 0$.

By stationarity, this implies that

$$x_{i,a}^{(t+1)} = \eta U_{i,a}^{(t)} - \lambda_i,$$

and

$$x_{i,a'}^{(t+1)} = \eta U_{i,a'}^{(t)} - \lambda_i.$$

The claim immediately follows. □

Claim 2. *Consider agent i and two possible strategies of i: $a, a' \in \mathcal{A}$. If at least one of $x_{i,a}^{(t+1)}, x_{i,a'}^{(t+1)}$ has non-zero probability mass, then $x_{i,a}^{(t+1)} \geq x_{i,a'}^{(t+1)}$ if and only if $U_{i,a}^{(t)} \geq U_{i,a'}^{(t)}$ with equality if and only if $x_{i,a}^{(t+1)} = x_{i,a'}^{(t+1)}$.*

Proof. To begin, the KKT conditions imply

$$x_{i,a}^{(t+1)} - \eta U_{i,a}^{(t)} + \lambda_i \geq 0,$$

and

$$x_{i,a}^{(t+1)}(x_{i,a}^{(t+1)} - \eta U_{i,a}^{(t)} + \lambda_i) = 0.$$

This implies

$$x_{i,a}^{(t)} = \begin{cases} \eta U_{i,a}^{(t)} - \lambda_i & \lambda_i < \eta U_{i,a}^{(t)} \\ 0 & \lambda_i \geq \eta U_{i,a}^{(t)} \end{cases}$$

First, suppose $x_{i,a}^{(t+1)} \geq x_{i,a'}^{(t+1)}$. If $x_{i,a}^{(t+1)} > 0$ and $x_{i,a'}^{(t+1)} = 0$, then from above we have $\eta U_{i,a'}^{(t)} \leq \lambda_i < \eta U_{i,a}^{(t)}$ and immediately we have $U_{i,a'}^{(t)} < U_{i,a}^{(t)}$. If both $x_{i,a}^{(t+1)} > 0$ and $x_{i,a'}^{(t+1)} > 0$, then from Claim 1, we have

$$x_{i,a}^{(t+1)} - x_{i,a'}^{(t+1)} = \eta \left(U_{i,a}^{(t)} - U_{i,a'}^{(t)} \right) \geq 0,$$

so we have $U_{i,a'}^{(t)} \leq U_{i,a}^{(t)}$ with equality if and only if $x_{i,a}^{(t+1)} = x_{i,a'}^{(t+1)}$.
Next, suppose $U_{i,a'}^{(t)} \leq U_{i,a}^{(t)}$. If both $x_{i,a}^{(t+1)} > 0$ and $x_{i,a'}^{(t+1)} > 0$, then, $x_{i,a}^{(t+1)} \geq x_{i,a'}^{(t+1)}$ with equality if and only if $U_{i,a'}^{(t)} = U_{i,a}^{(t)}$ immediately follows from Claim 1. Next, if $x_{i,a}^{(t+1)} > 0$ and $x_{i,a'}^{(t+1)} = 0$, then we immediately have $x_{i,a}^{(t+1)} > x_{i,a'}^{(t+1)}$. Further, it cannot be the case that $U_{i,a'}^{(t)} = U_{i,a}^{(t)}$ because this case implies $\eta U_{i,a'}^{(t)} \leq \lambda_i < \eta U_{i,a}^{(t)}$. Finally, if $x_{i,a}^{(t+1)} = 0$ and $x_{i,a'}^{(t+1)} > 0$, then we must have $\eta U_{i,a}^{(t)} \leq \lambda_i < \eta U_{i,a'}^{(t)}$ which contradicts our original assumption $U_{i,a'}^{(t)} \leq U_{i,a}^{(t)}$ and we can conclude that such a probability assignment in $x_i^{(t+1)}$ is not possible.

□

A.2 Convergence to ϵ-Mixed NE

First we state the main theorem to prove in this section. We will then prove several lemmas that will be necessary to prove this theorem. We will close by proving the main theorem. Throughout this section, we assume a firm and worker agent are learning strategies for $\mathcal{G}^{(1)}$ using Algorithm 1 parameterized by any $\eta > 0, D > 2$.

Theorem 11. *Suppose agents learn strategies for $\mathcal{G}^{(1)}$ using Algorithm 1 with $\alpha_i = \mathbf{0}$, any $\eta > 0, D > 2$, and arbitrary initial conditions $x_w^{(1)}, x_f^{(1)} \in \Delta(\mathcal{A})$. Then, for any $\epsilon > 0$, there exists a finite time t_ϵ where $(x_f^{(\tau)}, x_w^{(\tau)})$ is in ϵ-Nash Equilibrium for all $\tau \geq t_\epsilon$.*

Lemma 1. *The sequences $x_{w,0}^{(t)}, \ldots, x_{w,1}^{(t)}$ is non-increasing at all time steps $t > 1$ for any arbitrary sequence of firm mixed strategies $x_f^{(1)}, \ldots, x_f^{(t-1)}$. Further, $u_w(x_f^{(t)}, 0), \ldots, u_w(x_f^{(t)}, 1)$ is non-increasing at all time steps $t > 1$ for any arbitrary firm mixed strategy $x_f^{(t)}$.*

Proof. For any arbitrary sequence of firm mixed strategies $x_f^{(1)}, \ldots, x_f^{(t-1)}$, the cumulative utility the worker gets through time $t-1$ of an acceptance threshold $a \in \mathcal{A}$ is

$$U_{w,a}^{(t-1)} = \sum_{\tau=1}^{t-1} \sum_{a_p \geq a} x_{f,a_p}^{(\tau)} \cdot a_p.$$

This implies the following cumulative utility relation between subsequent strategies $a_k < a_{k+1}$:

$$U_{w,a_k}^{(t-1)} = U_{w,a_{k+1}}^{(t-1)} + \sum_{\tau=1}^{t-1} x_{f,a_k}^{(\tau)} \cdot a_k.$$

Since x_f is a probability distribution and each a_k is non-negative, we can conclude

$$U_{w,a_0}^{(t-1)} \geq \ldots \geq U_{w,a_D}^{(t-1)}.$$

By Claim 2, $x_{w,a_k}^{(t)} \geq x_{w,a_{k+1}}^{(t)}$ if and only if $U_{w,a_k}^{(t-1)} \geq U_{w,a_{k+1}}^{(t-1)}$ with equality if and only if $x_{w,a_k}^{(t)} = x_{w,a_{k+1}}^{(t)}$. Therefore, the sequence $x_{w,0}^{(t)}, \ldots, x_{w,1}^{(t)}$ is non-increasing. The result above holds for the expected utility to the worker at any time step t as well:

$$u_w(x_f^{(t)}, a_k) = u_w(x_f^{(t)}, a_{k+1}) + x_{f,a_k}^{(t)} \cdot a_k,$$

so we may conclude

$$u_w(x_f^{(t)}, a_0) \geq \ldots \geq u_w(x_f^{(t)}, a_D).$$

□

Lemma 2. *The sequence $x_{f,0}^{(t)}, \dots, x_{f,1}^{(t)}$ is unimodal at all time steps $t > 1$ for any arbitrary sequence of worker mixed strategies $x_w^{(1)}, \dots, x_w^{(t-1)}$ that satisfy Lemma 1. Further, $u_f(0, x_w^{(t)}), \dots, u_f(1, x_w^{(t)})$ is unimodal at all time steps $t > 1$ for any worker mixed strategy $x_w^{(t)}$ that satisfies Lemma 1.*

Proof. For any arbitrary sequence of worker mixed strategies $x_w^{(1)}, \dots, x_w^{(t-1)}$, the cumulative utility the firm gets through time $t-1$ for an offer of $a \in \mathcal{A}$ is

$$U_{f,a}^{(t-1)} = \sum_{\tau=1}^{t-1} \sum_{a_r \leq a} x_{w,a_r}^{(\tau)} \cdot (1-a).$$

We begin by showing the sequence $U_{f,0}^{(t-1)}, \dots, U_{f,1}^{(t-1)}$ is unimodal when the sequence $x_w^{(1)}, \dots, x_w^{(t-1)}$ satisfies Lemma 1.

Consider subsequent strategies $a_\ell = \frac{\ell}{D}, a_{\ell+1} = \frac{\ell+1}{D}$, then we have

$$U_{f,a_{\ell+1}}^{(t-1)} - U_{f,a_\ell}^{(t-1)} = \sum_{\tau=1}^{t-1} \left[x_{w,a_{\ell+1}}^{(\tau)} \left(\frac{D-\ell-1}{D} \right) - \sum_{a \leq a_\ell} x_{w,a}^{(\tau)} \frac{1}{D} \right]. \tag{1}$$

From expression 1, note that

$$U_{f,a_\ell}^{(t-1)} \leq U_{f,a_{\ell+1}}^{(t-1)} \iff \frac{\sum_{\tau=1}^{t-1} x_{w,a_{\ell+1}}^{(\tau)}}{\sum_{\tau=1}^{t-1} \sum_{a \leq a_\ell} x_{w,a}^{(\tau)}} \geq \frac{1}{D-\ell-1}. \tag{2}$$

Suppose $a_k = \frac{k}{D}$ is a cumulative utility maximizer, i.e.,

$$U_{f,a_k}^{(t-1)} - U_{f,a}^{(t-1)} \geq 0, \forall a \neq a_k \in \mathcal{A}. \tag{3}$$

So, to establish that $U_{f,a_1}^{(t-1)}, \dots, U_{f,a_{D-1}}^{(t-1)}$ is unimodal, it suffices to show

$$\begin{aligned} U_{f,a_{\ell-1}}^{(t-1)} - U_{f,a_\ell}^{(t-1)}, &\leq 0 \qquad && \forall \ell \in \{1, \dots, k\} \\ U_{f,a_\ell}^{(t-1)} - U_{f,a_{\ell+1}}^{(t-1)} &\geq 0. && \forall \ell \in \{k, \dots, D-1\} \end{aligned}$$

By Lemma 1, $x_w^{(\tau)}$ is non-increasing as the acceptance thresholds $a \to 1$ at every time step τ. Further, each $x_{w,a}^{(\tau)} \geq 0$ at every time step τ, so whenever $i \geq j$,

$$\sum_{a \leq a_i} x_{w,a}^{(\tau)} \geq \sum_{a \leq a_j} x_{w,a}^{(\tau)}.$$

Therefore, $\forall \ell \in \{1, \dots, k\}$,

$$\frac{\sum_{\tau=1}^{t-1} x_{w,a_\ell}^{(\tau)}}{\sum_{\tau=1}^{t-1} \sum_{a \leq a_{\ell-1}} x_{w,a}^{(\tau)}} \geq \frac{\sum_{\tau=1}^{t-1} x_{w,a_k}^{(\tau)}}{\sum_{\tau=1}^{t-1} \sum_{a \leq a_{k-1}} x_{w,a}^{(\tau)}} \geq \frac{1}{D-k},$$

and $\forall \ell \in \{k, \ldots, D-1\}$,

$$\frac{\sum_{\tau=1}^{t-1} x_{w,a_{\ell+1}}^{(\tau)}}{\sum_{\tau=1}^{t-1} \sum_{a \leq a_\ell} x_{w,a}^{(\tau)}} \leq \frac{\sum_{\tau=1}^{t-1} x_{w,a_{k+1}}^{(\tau)}}{\sum_{\tau=1}^{t-1} \sum_{a \leq a_k} x_{w,a}^{(\tau)}} \leq \frac{1}{D-k-1},$$

where the last inequality in each expression follows from combining expressions 2 and 3. Therefore,

$$\frac{\sum_{\tau=1}^{t-1} x_{w,a_\ell}^{(\tau)}}{\sum_{\tau=1}^{t-1} \sum_{a \leq a_{\ell-1}} x_{w,a}^{(\tau)}} \geq \frac{1}{D-\ell} \qquad \forall \ell \in \{1, \ldots, k\},$$

$$\frac{\sum_{\tau=1}^{t-1} x_{w,a_{\ell+1}}^{(\tau)}}{\sum_{\tau=1}^{t-1} \sum_{a \leq a_\ell} x_{w,a}^{(\tau)}} \leq \frac{1}{D-\ell-1} \qquad \forall \ell \in \{k, \ldots, D-1\}.$$

Finally, Claim 2 implies that if the sequence $U_{f,0}^{(t-1)}, \ldots, U_{f,1}^{(t-1)}$ is unimodal, then the sequence $x_{f,0}^{(t)}, \ldots, x_{f,1}^{(t)}$ is unimodal as well.
Further, the above logic holds for any time step t where $x_w^{(t)}$ satisfies Lemma 1, so we can conclude $u_f(0, x_w^{(t)}), \ldots, u_f(1, x_w^{(t)})$ is unimodal as well.

□

Recall the following notation which will be used throughout the subsequent lemmas.

$$w_{\max}^{(t)} = \max\{a | x_{w,a}^{(t)} > 0\},$$
$$f_{\min}^{(t)} = \min\{a | x_{f,a}^{(t)} > 0\}.$$

Lemma 3. *If agents play strategies at time t such that $w_{\max}^{(t)} \leq f_{\min}^{(t)}$, then $x_w^{(t+1)} = x_w^{(t)}$.*

Proof. Notice that, for any acceptance threshold $a' \leq f_{\min}^{(t)}$,

$$u_w(x_f^{(t)}, a') = \sum_{a \geq f_{\min}^{(t)}} x_{f,a}^{(t)} \cdot a$$

because $x_{f,a}^{(t)} = 0$ for all $a < f_{\min}^{(t)}$ by definition. This implies, for any $a, a' \leq f_{\min}^{(t)}$,

$$U_{w,a}^{(t)} - U_{w,a'}^{(t)} = U_{w,a}^{(t-1)} - U_{w,a'}^{(t-1)}. \tag{1}$$

First, we show that any acceptance threshold that gets some mass at time t and time $t+1$ must have the same probability mass difference with other such acceptance thresholds. Since $w_{\max}^{(t)} \leq f_{\min}^{(t)}$, then for any $a, a' \in \mathcal{A}$ where $x_{w,a}^{(t)} > 0, x_{w,a'}^{(t)} > 0$, by Claim 1 and equation 1,

$$x_{w,a}^{(t)} - x_{w,a'}^{(t)} = \eta(U_{w,a}^{(t-1)} - U_{w,a'}^{(t-1)}) = \eta(U_{w,a}^{(t)} - U_{w,a'}^{(t)}).$$

This implies that, if $x_{w,a}^{(t+1)} > 0, x_{w,a'}^{(t+1)} > 0$, then

$$x_{w,a}^{(t+1)} - x_{w,a'}^{(t+1)} = x_{w,a}^{(t)} - x_{w,a'}^{(t)} \tag{2}$$

Next, suppose for some $a, a' \leq w_{\max}^{(t)}$, we have $x_{w,a}^{(t)} > 0, x_{w,a'}^{(t)} > 0$, but $x_{w,a}^{(t+1)} > 0, x_{w,a'}^{(t+1)} = 0$. By Claim 2, the only way for this case to occur is for at least one acceptance threshold $a' \leq w_{\max}^{(t)}$ to get 0 mass at time $t+1$ and no acceptance threshold greater than $w_{\max}^{(t)}$, which gets 0 mass at time t by definition, to have non-zero mass at time $t+1$. Therefore, this is the only case to consider for an acceptance threshold getting 0 mass at time $t+1$ after having non-zero mass at time t. Here, the number of acceptance thresholds that get mass must be strictly less than those that do at time t, so by the primal constraints and equation 2 we must have

$$x_{w,a}^{(t+1)} > x_{w,a}^{(t)}.$$

Then, by the KKT conditions,

$$\lambda_r^{(t+1)} = \eta(U_{w,a}^{(t-1)} + u_w(x_f^{(t)}, a)) - x_{w,a}^{(t+1)} \geq \eta(U_{w,a'}^{(t-1)} + u_w(x_f^{(t)}, a')).$$

Since $w_{\max}^{(t)} \leq f_{\min}^{(t)}$, then $u_w(x_f^{(t)}, a) = u_w(x_f^{(t)}, a')$, so

$$\eta(U_{w,a}^{(t-1)} - U_{w,a'}^{(t-1)}) \geq x_{w,a}^{(t+1)}.$$

By Claim 1, this implies

$$x_{w,a}^{(t)} - x_{w,a'}^{(t)} \geq x_{w,a}^{(t+1)},$$

however, since $x_{w,a'}^{(t)} > 0$, this implies the contradiction

$$x_{w,a}^{(t)} > x_{w,a}^{(t+1)}.$$

Therefore, it is impossible for some $a, a' \leq w_{\max}^{(t)}$ to satisfy $x_{w,a}^{(t)} > 0, x_{w,a'}^{(t)} > 0$, but $x_{w,a}^{(t+1)} > 0, x_{w,a'}^{(t+1)} = 0$.

Finally, suppose for some $a \leq w_{\max}^{(t)} < a'$, we have $x_{w,a}^{(t)} > 0, x_{w,a'}^{(t)} = 0$, but $x_{w,a}^{(t+1)} > 0, x_{w,a'}^{(t+1)} > 0$. Note by Claim 2, it is impossible for $x_{w,a}^{(t)} = 0, x_{w,a'}^{(t)} > 0$ since the cumulative utility functions are non-increasing as a increases, so this is the only case to consider for an acceptance threshold gaining mass at time $t+1$ after having 0 mass at time t. Further, by the primal constraints and equation 2, this implies

$$x_{w,a}^{(t+1)} < x_{w,a}^{(t)}.$$

First, by the KKT conditions,

$$\eta U_{w,a'}^{(t-1)} \leq \lambda_r^{(t)} = \eta U_{w,a}^{(t-1)} - x_{w,a}^{(t)},$$

but if $x_{w,a'}^{(t+1)} > 0$

$$\lambda_r^{(t+1)} < \eta U_{w,a'}^{(t)}.$$

Since $x_{w,a}^{(t+1)} < x_{w,a}^{(t)}$,

$$\lambda_r^{(t+1)} = \eta(U_{w,a}^{(t-1)} + u_w(x_f^{(t)}, a)) - x_{w,a}^{(t+1)} \geq \lambda_r^{(t)} + \eta u_w(x_f^{(t)}, a).$$

However, since $u_w(x_f^{(t)}, a) \geq u_w(x_f^{(t)}, a')$ by Lemma 1,

$$\lambda_r^{(t)} + \eta u_w(x_f^{(t)}, a) \geq \eta(U_{w,a'}^{(t-1)} + u_w(x_f^{(t)}, a')),$$

which implies the contradiction

$$\lambda_r^{(t+1)} \geq \eta U_{w,a'}^{(t)}.$$

Therefore, the same acceptance thresholds get non-zero probability mass at time t and $t+1$ and their probability mass differences must remain the same, thus, $x_w^{(t+1)} = x_w^{(t)}$. □

Lemma 4. *Suppose at time t that $f_{\min}^{(t)} < w_{\max}^{(t)}$. Then, $x_{w,w_{\max}^{(t)}}^{(t+1)} < x_{w,w_{\max}^{(t)}}^{(t)}$.*

Proof. To begin, if $x_{w,w_{\max}^{(t)}}^{(t+1)} = 0$, then immediately by definition of $w_{\max}^{(t)}$,

$$x_{w,w_{\max}^{(t)}}^{(t+1)} < x_{w,w_{\max}^{(t)}}^{(t)}.$$

Next suppose $x_{w,w_{\max}^{(t)}}^{(t+1)} > 0$. First, by Lemma 1 and the definition of $w_{\max}^{(t)}$, for all $a < w_{\max}^{(t)}$,

$$x_{w,a}^{(t)} \geq x_{w,w_{\max}^{(t)}}^{(t)} > 0,$$

and Claim 2 implies

$$U_{w,a}^{(t-1)} \geq U_{w,w_{\max}^{(t)}}^{(t-1)}.$$

Next, since $f_{\min}^{(t)} < w_{\max}^{(t)}$ it must be the case that, for all $a < w_{\max}^{(t)}$,

$$u_w(x_f^{(t)}, a) - u_w(x_f^{(t)}, w_{\max}^{(t)}) > 0.$$

Therefore,

$$U_{w,a}^{(t)} - U_{w,w_{\max}^{(t)}}^{(t)} > U_{w,a}^{(t-1)} - U_{w,w_{\max}^{(t)}}^{(t-1)} \tag{1}$$

Since $x_{w,w_{\max}^{(t)}}^{(t+1)} > 0$, then by Lemma 1, $x_{w,a}^{(t+1)} > 0$ for all $a < w_{\max}^{(t)}$. Then, Claim 1 and inequality 1 implies for all $a < w_{\max}^{(t)}$,

$$x_{w,a}^{(t+1)} - x_{w,w_{\max}^{(t)}}^{(t+1)} > x_{w,a}^{(t)} - x_{w,w_{\max}^{(t)}}^{(t)}.$$

If $x^{(t+1)}_{w,w^{(t)}_{\max}} \geq x^{(t)}_{w,w^{(t)}_{\max}}$, then this implies for all $a < w^{(t)}_{\max}$,

$$x^{(t+1)}_{w,a} > x^{(t)}_{w,a} \tag{2}$$

However, by the primal constraint

$$\sum_{a \leq w^{(t)}_{\max}} x^{(t)}_{w,a} = 1,$$

so the assumption $x^{(t+1)}_{w,w^{(t)}_{\max}} \geq x^{(t)}_{w,w^{(t)}_{\max}}$ along with inequality 2 implies that

$$\sum_{a \leq w^{(t)}_{\max}} x^{(t+1)}_{w,a} > 1,$$

which violates the primal constraint at time $t+1$. Therefore,

$$x^{(t+1)}_{w,w^{(t)}_{\max}} < x^{(t)}_{w,w^{(t)}_{\max}}.$$

□

Lemma 5. *At any time step t, $w^{(\tau)}_{\max} \leq w^{(t)}_{\max}$ for all $\tau \geq t$.*

Proof. First, if $f^{(t)}_{\min} \geq w^{(t)}_{\max}$, then by Lemma 3,

$$x^{(t+1)}_{w} = x^{(t)}_{w} \implies x^{(t+1)}_{w,w^{(t)}_{\max}} = x^{(t)}_{w,w^{(t)}_{\max}}.$$

Otherwise if $f^{(t)}_{\min} < w^{(t)}_{\max}$, then we will show it's impossible to have $a > w^{(t)}_{\max}$ with $x^{(t)}_{w,a} = 0$ but $x^{(t+1)}_{w,a} > 0$. First, by Lemma 4,

$$x^{(t+1)}_{w,w^{(t)}_{\max}} < x^{(t)}_{w,w^{(t)}_{\max}}.$$

If $x^{(t+1)}_{w,w^{(t)}_{\max}} = 0$, then $x^{(t+1)}_{w,a} = 0$ necessarily by Lemma 1.
Otherwise suppose $x^{(t+1)}_{w,w^{(t)}_{\max}} > 0$. Note that the KKT conditions at time t imply

$$\eta U^{(t-1)}_{w,w^{(t)}_{\max}} - \eta U^{(t-1)}_{w,a} \geq x^{(t)}_{w,w^{(t)}_{\max}}.$$

Then, if $x^{(t+1)}_{w,a} > 0$ and $x^{(t+1)}_{w,w^{(t+1)}_{\max}} > 0$, by Claim 1, it must be true that

$$x^{(t+1)}_{w,w^{(t)}_{\max}} - x^{(t+1)}_{w,a} = \eta U^{(t)}_{w,w^{(t)}_{\max}} - \eta U^{(t)}_{w,a},$$

which implies

$$x^{(t+1)}_{w,w^{(t)}_{\max}} > \eta U^{(t)}_{w,w^{(t)}_{\max}} - \eta U^{(t)}_{w,a}.$$

However, since $x^{(t+1)}_{w,w^{(t)}_{\max}} < x^{(t)}_{w,w^{(t)}_{\max}}$, then this implies

$$U^{(t-1)}_{w,w^{(t)}_{\max}} - \eta U^{(t-1)}_{w,a} > \eta U^{(t)}_{w,w^{(t)}_{\max}} - \eta U^{(t)}_{w,a},$$

or

$$u_w(x^{(t)}_f, a) > u_w(x^{(t)}_f, w^{(t)}_{\max})$$

which is impossible by Lemma 1. Therefore, $w^{(\tau)}_{\max} \le w^{(t)}_{\max}$ for all $\tau \ge t$. □

Lemma 6. *Suppose at time t that $w^{(t)}_{\max} < f^{(t)}_{\min}$. Then, there is always a time $t' > t$ where $f^{(t')}_{\min} \le w^{(t')}_{\max}$.*

Proof. Suppose instead it is the case that for all $\tau \ge t$, $w^{(\tau)}_{\max} < f^{(\tau)}_{\min}$. First, notice by Lemma 3 that for all $\tau \ge t$ where $w^{(\tau)}_{\max} < f^{(\tau)}_{\min}$,

$$x^{(\tau+1)}_w = x^{(\tau)}_w,$$

thus, $w^{(\tau)}_{\max} = w^{(t)}_{\max}$ for all $\tau \ge t$.
Next, by definition of the firm's utility function,

$$u_f(w^{(\tau)}_{\max}, x^{(\tau)}_w) \ge u_f(a, x^{(\tau)}_w) + \frac{1}{D}, \forall a > w^{(\tau)}_{\max}$$

Therefore, since $w^{(\tau)}_{\max}$ is fixed for all $\tau \ge t$, there exists a time $t' > t$ where

$$U^{(t')}_{f,w^{(t')}_{\max}} \ge U^{(t')}_{f,a}, \forall a > w^{(t')}_{\max}$$

So, if there exists $a > w^{(t')}_{\max}$ where $x^{(t')}_{f,a} > 0$, then by Claim 2, it must be the case that $x^{(t')}_{f,w^{(t')}_{\max}} \ge x^{(t')}_{f,a} > 0$ which implies

$$f^{(t')}_{\min} \le w^{(t')}_{\max}.$$

Otherwise, if no such $a > w^{(t')}_{\max}$ where $x^{(t')}_{f,a} > 0$ exists, then by definition of $f^{(t')}_{\min}$ and the primal constraints we again have

$$f^{(t')}_{\min} \le w^{(t')}_{\max}.$$

Therefore, by contradiction, there always exists a time $t' > t$ where $f^{(t')}_{\min} \le w^{(t')}_{\max}$. □

Lemma 7. *Suppose at time t, $f^{(t)}_{\min} \le w^{(t)}_{\max}$. Then, there exists a finite time $t' \ge t$ where $f^{(\tau)}_{\min} \le w^{(t)}_{\max}$ for all $\tau \ge t'$.*

Proof. Suppose at time t, $f^{(t)}_{\min} \leq w^{(t)}_{\max}$. First, by Lemma 5, $w^{(\tau)}_{\max} \leq w^{(t)}_{\max}$ for all $\tau \geq t$. As a result, for all $\tau \geq t$,

$$u_f(w^{(t)}_{\max}, x^{(\tau)}_w) \geq u_f(a, x^{(\tau)}_w) + \frac{1}{D}, \forall a > w^{(t)}_{\max}.$$

which implies there exists a time $t' \geq t$ where for all $a > w^{(t)}_{\max}$ and all $\tau \geq t'$,

$$U^{(\tau)}_{f,w^{(t)}_{\max}} \geq U^{(\tau)}_{f,a}.$$

Therefore, by Claim 2, for all $\tau \geq t'$, it is impossible for at least one $a > w^{(t)}_{\max}$ to have $x^{(\tau)}_{f,a} > 0$, but $x^{(\tau)}_{f,w^{(t)}_{\max}} = 0$. Thus, $f^{(\tau)}_{\min} \leq w^{(t)}_{\max}$ for all $\tau \geq t'$. □

Lemma 8. *Suppose at time t, $w^{(t)}_{\max} = \frac{k}{D}$ for some $k \in \{2, \ldots, D\}$, $x^{(t)}_{w,w^{(t)}_{\max}} < \frac{1}{D-k+1}$, and $f^{(t)}_{\min} = w^{(t)}_{\max}$. Then, there is a finite time $t' > t$ where $f^{(\tau)}_{\min} < w^{(t')}_{\max}$ for all $\tau \geq t'$.*

Proof. Suppose at time t, $f^{(t)}_{\min} = w^{(t)}_{\max}$. First, by Lemma 7, there exists a finite time $t' \geq t$ where $f^{(\tau)}_{\min} \leq w^{(t)}_{\max}$ for all $\tau \geq t'$, so suppose $f^{(\tau)}_{\min} = w^{(t)}_{\max}$ for all $\tau \geq t'$. Then, by Lemma 3, $x^{(\tau)}_w = x^{(t)}_w$ for all $\tau \geq t'$. Note that since $k \geq 2$, there exists a smaller action than $w^{(t)}_{\max}$: $w^{(t)}_{\max} - \frac{1}{D} \in \mathcal{A}$. Then $x^{(\tau)}_{w^{(t)}_{\max}} < \frac{1}{D-k+1}$ implies for all $\tau \geq t'$,

$$u_f(w^{(t)}_{\max} - \frac{1}{D}, x^{(\tau)}_w) > \left(1 - \frac{1}{D-k+1}\right) \cdot \left(1 - w^{(t)}_{\max} + \frac{1}{D}\right) = u_f(w^{(t)}_{\max}, x^{(\tau)}_w)$$

which implies

$$u_f(w^{(t)}_{\max} - \frac{1}{D}, x^{(\tau)}_w) - u_f(w^{(t)}_{\max}, x^{(\tau)}_w)$$

is a constant, positive value for all $\tau \geq t'$. Therefore, there exists another time t^* where

$$U^{(t^*)}_{f,w^{(t)}_{\max}-\frac{1}{D}} \geq U^{(t^*)}_{f,w^{(t)}_{\max}}$$

Since $f^{(\tau)}_{\min} = w^{(t)}_{\max}$ for all $\tau \geq t'$, then we must have $x^{(t^*)}_{f,w^{(t)}_{\max}} > 0$, but by Claim 2,

$$x^{(t^*)}_{f,w^{(t)}_{\max}-\frac{1}{D}} \geq x^{(t^*}_{f,w^{(t)}_{\max}} > 0,$$

which immediately implies $f^{(t^*)}_{\min} < w^{(t)}_{\max}$.

Next, suppose at time t, $f^{(t)}_{\min} < w^{(t)}_{\max}$ and $x^{(t)}_{w,w^{(t)}_{\max}} < \frac{1}{D-k+1}$. Then, we will show it is impossible at time step $t+1$ to have $x^{(t+1)}_{f,a} = 0$ for all $a < w^{(t)}_{\max}$.

To begin, since $x^{(t)}_{w,w^{(t)}_{\max}} < \frac{1}{D-k+1}$ and $u_f(w^{(t)}_{\max}, x^{(t)}_w) \geq u_f(a, x^{(t)}_w) + \frac{1}{D}$ for all $a > w^{(t)}_{\max}$,

$$u_f(w^{(t)}_{\max} - \frac{1}{D}, x^{(t)}_w) > u_f(a, x^{(t)}_w), \forall a \geq w^{(t)}_{\max} \tag{1}$$

First, suppose $x^{(t)}_{f,w^{(t)}_{\max}-\frac{1}{D}} > 0$. By Claim 1, this implies for all $a \geq w^{(t)}_{\max}$ where $x^{(t)}_{f,a} > 0$,

$$x^{(t)}_{f,w^{(t)}_{\max}-\frac{1}{D}} - x^{(t)}_{f,a} = \eta U^{(t-1)}_{f,w^{(t)}_{\max}-\frac{1}{D}} - \eta U^{(t-1)}_{f,a}, \tag{2}$$

and for all $a \geq w^{(t)}_{\max}$ where $x^{(t)}_{f,a} = 0$,

$$U^{(t-1)}_{f,w^{(t)}_{\max}-\frac{1}{D}} > U^{(t-1)}_{f,a}. \tag{3}$$

Then, suppose $x^{(t+1)}_{f,w^{(t)}_{\max}-\frac{1}{D}} = 0$ and there exists $a \geq w^{(t)}_{\max}$ where $x^{(t+1)}_{f,a} > 0$. This implies

$$U^{(t-1)}_{f,a} > U^{(t-1)}_{f,w^{(t)}_{\max}-\frac{1}{D}},$$

so by inequalities 1 and 3, it is impossible for such an a to have $x^{(t)}_{f,a} = 0$. So, it must be the case that $x^{(t)}_{f,a} > 0$. Then by the KKT conditions,

$$x^{(t+1)}_{f,a} \leq \eta U^{(t)}_{f,a} - \eta U^{(t)}_{f,w^{(t)}_{\max}-\frac{1}{D}}$$

Subtracting both sides by equation 2 and applying inequality 1 implies

$$x^{(t+1)}_{f,a} < x^{(t)}_{f,a}.$$

Since this is true for any $a \geq w^{(t)}_{\max}$, then

$$\sum_{a \geq w^{(t)}_{\max}} x^{(t+1)}_{f,a} < \sum_{a \geq w^{(t)}_{\max}} x^{(t)}_{f,a} \leq 1,$$

which implies by the primal constraints that it is impossible for all $a < w^{(t)}_{\max}$ to have $x^{(t+1)}_{f,a} = 0$ and this contradicts $x^{(t+1)}_{f,w^{(t)}_{\max}-\frac{1}{D}} = 0$.

Next, suppose $x^{(t)}_{f,w^{(t)}_{\max}-\frac{1}{D}} = 0$. Then by Lemma 2 there exists $f^{(t)}_{\min} \leq a^* < w^{(t)}_{\max} - \frac{1}{D}$ such that

$$U^{(t-1)}_{f,a^*} \geq U^{(t-1)}_{f,a}, \forall a \neq a^* \tag{4}$$

This implies, by Lemma 2,

$$U^{(t-1)}_{f,w^{(t)}_{\max}-\frac{1}{D}} \geq U^{(t-1)}_{f,a}, \forall a \geq w^{(t)}_{\max} \tag{5}$$

Then, if there exists $a \geq w^{(t)}_{\max}$ where $x^{(t+1)}_{f,a} > 0$, but all $a' < w^{(t)}_{\max}$ have $x^{(t+1)}_{f,a'} = 0$, then by Claim 2,

$$U^{(t)}_{f,a} > U^{(t)}_{f,a^*},$$

which along with inequality 4 implies

$$u_f(a, x_w^{(t)}) > u_f(a^*, x_w^{(t)}).$$

However, by inequalities 1 and 5,

$$U^{(t)}_{f,w^{(t)}_{\max}-\frac{1}{D}} \geq U^{(t)}_{f,a},$$

so by Claim 2, if $x^{(t+1)}_{f,a} > 0$, then $x^{(t)}_{f,w^{(t)}_{\max}-\frac{1}{D}} > 0$ which contradicts all $a' < w^{(t)}_{\max}$ have $x^{(t+1)}_{f,a'} = 0$.

Therefore, in all possible cases, $f^{(t+1)}_{\min} < w^{(t)}_{\max}$. Further, by Lemma 4, it is also the case that

$$x^{(t+1)}_{w,w^{(t)}_{\max}} < x^{(t)}_{w,w^{(t)}_{\max}} < \frac{1}{D-k+1},$$

so we can conclude $f^{(\tau)}_{\min} < w^{(t)}_{\max}$ for all $\tau \geq t$.

□

Lemma 9. *Suppose at time t, $w^{(t)}_{\max} = \frac{k}{D}$ for some $k \in \{2, \ldots, D\}$, $x^{(t)}_{w,w^{(t)}_{\max}} < \frac{1}{D-k+1}$, and $f^{(\tau)}_{\min} < w^{(t)}_{\max}$ for all $\tau \geq t$. Then, there is a finite time $t' > t$ where $w^{(t')}_{\max} < w^{(t)}_{\max}$.*

Proof. Suppose $f^{(\tau)}_{\min} < w^{(t)}_{\max}$ for all $\tau \geq t$. Then, we will show there must be another finite time $t' > t$ where $x^{(t')}_{w,w^{(t)}_{\max}} = 0$. Then by Lemma 1 this implies it is also the case that $x^{(t')}_{w,a} = 0$ for all $a \geq w^{(t)}_{\max}$, and we can conclude $w^{(t')}_{\max} < w^{(t)}_{\max}$.

First, note that since $k \geq 2$, there exists a smaller action than $w^{(t)}_{\max}$: $w^{(t)}_{\max} - \frac{1}{D} \in \mathcal{A}$. Then, since $x^{(t)}_{w,w^{(t)}_{\max}} < \frac{1}{D-k+1}$,

$$u_f(w^{(t)}_{\max} - \frac{1}{D}, x^{(t)}_w) > u_f(w^{(t)}_{\max}, x^{(t)}_w).$$

Further, since $f^{(\tau)}_{\min} < w^{(t)}_{\max}$ for all $\tau \geq t$, by Lemma 4 either $x^{(\tau)}_{w,w^{(t)}_{\max}} = 0$ or $x^{(\tau+1)}_{w,w^{(t)}_{\max}} < x^{(\tau)}_{w,w^{(t)}_{\max}}$ for all $\tau \geq t$. Then, this implies by the definition of the firm's utility function that for all $\tau \geq t$,

$$u_f(w^{(t)}_{\max} - \frac{1}{D}, x^{(\tau)}_w) - u_f(w^{(t)}_{\max}, x^{(\tau)}_w) > u_f(w^{(t)}_{\max} - \frac{1}{D}, x^{(t)}_w) - u_f(w^{(t)}_{\max}, x^{(t)}_w) > 0.$$

This implies there exists a time t^* where for all $\tau \geq t^*$,

$$U^{(\tau)}_{f,w^{(t)}_{\max}-\frac{1}{D}} - U^{(\tau)}_{f,w^{(t)}_{\max}} \geq \frac{1}{\eta},$$

and by Claim 1 and the primal constraints, it cannot be the case that both

$$x^{(\tau)}_{f,w^{(t)}_{\max}-\frac{1}{D}} > 0 \text{ and } x^{(\tau)}_{f,w^{(t)}_{\max}} > 0,$$

and since

$$U^{(\tau)}_{f,w^{(t)}_{\max}-\frac{1}{D}} > U^{(\tau)}_{f,w^{(t)}_{\max}},$$

by Claim 2, we can conclude

$$x^{(\tau)}_{f,w^{(t)}_{\max}} = 0, \forall \tau \geq t^*.$$

So, there must always be a time $t^* > t$ where either $x^{(t^*)}_{w,w^{(t)}_{\max}} = 0$ or $x^{(t^*)}_{w,w^{(t)}_{\max}} > 0$ and $x^{(\tau)}_{f,w^{(t)}_{\max}} = 0$ for all $\tau \geq t^*$. The latter case implies that for all $\tau \geq t^*$,

$$U^{(\tau)}_{w,w^{(t)}_{\max}} = U^{(t^*-1)}_{w,w^{(t)}_{\max}}.$$

However, since $k \geq 2$, $\frac{1}{D}$ is a lower acceptance threshold for the worker than $w^{(t)}_{\max}$. Then, by the worker's utility function

$$u_w(x^{(t)}_f, \frac{1}{D}) = \sum_{a \geq \frac{1}{D}} x^{(t)}_{f,a} \cdot a \geq (1 - x^{(t)}_{f,0}) \cdot \frac{1}{D}. \tag{1}$$

Note that by the fact that an acceptance threshold of 0 for the worker cannot get more utility than an acceptance threshold of $\frac{1}{D}$, then by Lemma 1

$$U^{(t)}_{w,0} = U^{(t)}_{w,\frac{1}{D}} \geq U^{(t)}_{w,a}, \forall a > \frac{1}{D}$$

for all time steps t, so by Claim 1 it is always the case that $x^{(t)}_{w,0} = x^{(t)}_{w,\frac{1}{D}} > 0$. This implies that when $D > 2$,

$$u_f(\frac{1}{D}, x^{(t)}_w) > u_f(0, x^{(t)}_w), \forall t.$$

Therefore,

$$U^{(t)}_{f,\frac{1}{D}} > U^{(t)}_{f,0}, \forall t$$

So, by Claim 2

$$x^{(t)}_{f,\frac{1}{D}} > x^{(t)}_{f,0}, \forall t.$$

By the primal constraints, this implies $x^{(t)}_{f,0} < \frac{1}{2}$ for all t, so combining this fact with the lower bound 1, then for any time step t,

$$U^{(t)}_{w,\frac{1}{D}} \geq U^{(t-1)}_{w,\frac{1}{D}} + \frac{1}{2D}.$$

Therefore, since the cumulative utility of the offer $w^{(t)}_{\max}$ stops growing after time t^*, there must be a finite time $t' \geq t^*$ where

$$U^{(t')}_{w,\frac{1}{D}} - U^{(t')}_{w,w^{(t)}_{\max}} \geq \frac{1}{\eta},$$

and by Claim 1 and the primal constraints, it must be the case that $x^{(t')}_{w,w^{(t)}_{\max}} = 0$ and we can conclude $w^{(t')}_{\max} < w^{(t)}_{\max}$.

□

Lemma 10. *Suppose at time t, $w^{(t)}_{\max} = \frac{k}{D}$ for some $k \in \{1, \ldots, D-1\}$, $f^{(\tau)}_{\min} \leq w^{(t)}_{\max}$, and $x^{(\tau)}_{w,w^{(t)}_{\max}} \geq \frac{1}{D-k+1}$ for all $\tau \geq t$. Then, for any $\epsilon > 0$, there exists a time $t_\epsilon \geq t$ where $(x^{(t_\epsilon)}_f, x^{(t_\epsilon)}_w)$ is in an ϵ-mixed Nash Equilibrium.*

Proof. Suppose at time t, $f^{(\tau)}_{\min} \leq w^{(t)}_{\max}$ and $x^{(\tau)}_{w,w^{(t)}_{\max}} \geq \frac{1}{D-k+1}$ for all $\tau \geq t$. First, by definition of $w^{(t)}_{\max}$ and the firm's expected utility function,

$$u_f(w^{(t)}_{\max}, x^{(\tau)}_w) \geq u_f(a, x^{(\tau)}_w) + \frac{1}{D}, \forall a > w^{(t)}_{\max} \tag{1}$$

This implies there exists a time $t' \geq t$ where for all $\tau \geq t'$

$$U^{(\tau)}_{f,w^{(t)}_{\max}} - U^{(\tau)}_{f,a} \geq \frac{1}{\eta}, \forall a > w^{(t)}_{\max}.$$

Then, by Claim 1, for each $a > w^{(t)}_{\max}$ and all time steps $\tau \geq t'$, it is impossible for

$$x^{(\tau)}_{f,w^{(t)}_{\max}} > 0, x^{(\tau)}_{f,a} > 0.$$

By expression 1, it must be the case that for all $\tau \geq t'$,

$$x^{(\tau)}_{f,a} = 0, \forall a > w^{(t)}_{\max} \tag{2}$$

Next, since

$$x^{(\tau)}_{w,w^{(t)}_{\max}} \geq \frac{1}{D-k+1}, \forall \tau \geq t,$$

then for all $\tau \geq t$,

$$\begin{aligned} u_f(w^{(t)}_{\max} - \frac{1}{D}, x^{(\tau)}_w) &\leq (1 - \frac{1}{D-k+1})\frac{D-k+1}{D} \\ &= \frac{D-k}{D} \\ &= u_f(w^{(t)}_{\max}, x^{(\tau)}_w) \end{aligned}$$

By Lemma 2, if $u_f(w^{(t)}_{\max} - \frac{1}{D}, x^{(\tau)}_w) \leq u_f(w^{(t)}_{\max}, x^{(\tau)}_w)$, then it must also be the case that for all $a < w^{(t)}_{\max}$,

$$u_f(a, x^{(\tau)}_w) \leq u_f(w^{(t)}_{\max}, x^{(\tau)}_w), \forall \tau \geq t.$$

Combining this with expression 1, we can conclude for all $\tau \geq t$,

$$u_f(w^{(t)}_{\max}, x^{(\tau)}_w) \geq u_f(a, x^{(\tau)}_w), \forall a \neq w^{(t)}_{\max} \tag{3}$$

We now break into the individual cases of $f_{\min}^{(\tau)} = w_{\max}^{(t)}$ for all $\tau \geq t$ and $f_{\min}^{(\tau)} < w_{\max}^{(t)}$ for all $\tau \geq t$ to finish the proof. It is sufficient to consider these two cases because after time t' where expression 2 becomes true, then if $f_{\min}^{(t')} < w_{\max}^{(t)}$, but there exists a time $t^* > t'$ where $f_{\min}^{(t^*)} = w_{\max}^{(t)}$, then we immediately have $x_{f,w_{\max}^{(t)}}^{(t^*)} = 1$ and the first case below shows this implies convergence.

In the first case, suppose $f_{\min}^{(\tau)} = w_{\max}^{(t)}$ for all $\tau \geq t$ which implies for all $\tau \geq t$,

$$x_{f,a}^{(\tau)} = 0, \forall a < w_{\max}^{(t)}.$$

Combining this with expression 2, we can conclude that it must be the case that

$$x_{f,w_{\max}^{(t)}}^{(\tau)} = 1, \forall \tau \geq t'.$$

Therefore, there exists a finite time where the firm purely offers $w_{\max}^{(t)}$ for all future time steps and by expression 3, this offer will always be a best response to the worker's strategy. Further, $w_{\max}^{(t)}$ is the largest acceptance threshold with non-zero probability by definition, it is impossible for the worker switch acceptance thresholds to get more utility than $w_{\max}^{(t)}$. So, any mixture over acceptance thresholds $a \leq w_{\max}^{(t)}$ is a best response to $x_{f,w_{\max}^{(t)}}^{(\tau)} = 1$. Therefore, we can conclude the agents have converged to the strategy profile $(x_w^{(t')}, x_f^{(t')})$ and that the strategy profile is a mixed Nash Equilibrium.

In the second case, suppose $f_{\min}^{(\tau)} < w_{\max}^{(t)}$, but $x_{w,w_{\max}^{(t)}}^{(\tau)} \geq \frac{1}{D-k+1}$ for all $\tau \geq t$. Let $t' \geq t$ be the time where expression 2 guarantees offers greater than $w_{\max}^{(t)}$ get 0 probability mass in all future time steps. Further, the following two properties must hold in this case

$$x_{w,w_{\max}^{(t)}}^{(\tau)} > \frac{1}{D-k+1}, \forall \tau \geq t, \tag{4}$$

and there exists a $t^* \geq t'$ where

$$x_{f,w_{\max}^{(t)}}^{(t^*)} \geq x_{f,a}^{(t^*)}, \forall a \neq w_{\max}^{(t)}. \tag{5}$$

By Lemma 4 $x_{w,w_{\max}^{(t)}}^{(\tau+1)} < x_{w,w_{\max}^{(t)}}^{(\tau)}$ when $f_{\min}^{(\tau)} < w_{\max}^{(t)}$, so $x_{w,w_{\max}^{(t)}}^{(\tau+1)} < \frac{1}{D-k+1}$ if $x_{w,w_{\max}^{(t)}}^{(\tau)} = \frac{1}{D-k+1}$, so property 4 must hold. Next, if for all $\tau \geq t'$ there exists $f_{\min}^{(\tau)} \leq a < w_{\max}^{(t)}$ where

$$x_{f,a}^{(\tau)} > x_{f,w_{\max}^{(t)}}^{(\tau)},$$

then by the primal constraints this implies

$$\sum_{f_{\min}^{(\tau)} \leq a < w_{\max}^{(t)}} x_{f,a}^{(\tau)} > \frac{1}{2},$$

so by the worker's utility function

$$u_w(x_f^{(\tau)}, \frac{1}{D}) \geq u_w(x_f^{(\tau)}, w_{\max}^{(t)}) + \frac{1}{2D},$$

since a lower acceptance threshold, $\frac{1}{D}$ gets at least $\frac{1}{D}$ more utility than the acceptance threshold $w_{\max}^{(t)}$ with probability at least $\frac{1}{2}$. This implies there exists a time $t^* > t'$ where

$$U_{w,\frac{1}{D}}^{(t^*)} - U_{w,w_{\max}^{(t)}}^{(t^*)} \geq \frac{1}{\eta},$$

which implies $x_{w,w_{\max}^{(t)}}^{(t^*)} = 0$. Therefore, property 5 must be true as well.

Now, by property 4, then by the definition of the firm's utility function, for all $\tau \geq t$

$$u_f(w_{\max}^{(t)}, x_w^{(\tau)}) > u_f(w_{\max}^{(t)} - \frac{1}{D}, x_w^{(\tau)}),$$

so by Lemma 2,

$$u_f(w_{\max}^{(t)}, x_w^{(\tau)}) > u_f(a, x_w^{(\tau)}), \forall a < w_{\max}^{(t)},$$

and combining this with expression 1,

$$u_f(w_{\max}^{(t)}, x_w^{(\tau)}) > u_f(a, x_w^{(\tau)}), \forall a \neq w_{\max}^{(t)} \tag{4}$$

Next, by property 5, Claim 2, and expression 4, then for all $\tau \geq t^*$

$$U_{f,w_{\max}^{(t)}}^{(\tau+1)} - U_{f,a}^{(\tau+1)} > U_{f,w_{\max}^{(t)}}^{(\tau)} - U_{f,a}^{(\tau)} \geq 0, \forall a \neq w_{\max}^{(t)}.$$

Then, by Claim 1, this implies for all $\tau \geq t^*$ and for all $a \neq w_{\max}^{(t)}$ where $x_{f,a}^{(\tau)} > 0$ and $x_{f,a}^{(\tau+1)} > 0$,

$$x_{f,w_{\max}^{(t)}}^{(\tau+1)} - x_{f,a}^{(\tau+1)} > x_{f,w_{\max}^{(t)}}^{(\tau)} - x_{f,a}^{(\tau)}.$$

Further, it cannot be the case that $x_{f,w_{\max}^{(t)}}^{(\tau+1)} = x_{f,w_{\max}^{(t)}}^{(\tau)}$ while $x_{f,a}^{(\tau+1)} < x_{f,a}^{(\tau)}$ for all such $a \neq w_{\max}^{(t)}$ because this implies

$$\sum_{a \in \mathcal{A}} x_{f,a}^{(\tau+1)} < \sum_{a \in \mathcal{A}} x_{f,a}^{(\tau)} = 1,$$

which would violate the primal constraint at time $t+1$.

So, we can conclude $x_{f,w_{\max}^{(t)}}^{(\tau+1)} > x_{f,w_{\max}^{(t)}}^{(\tau)}$ for all $\tau \geq t^*$. Since for all $a > w_{\max}^{(t)}$, $x_{f,a}^{(\tau)} = 0$ for all $\tau \geq t'$ and $t^* \geq t'$, then this immediately implies

$$\sum_{f_{\min}^{(\tau+1)} \leq a < w_{\max}^{(t)}} x_{f,a}^{(\tau+1)} < \sum_{f_{\min}^{(\tau)} \leq a < w_{\max}^{(t)}} x_{f,a}^{(\tau)}, \forall \tau \geq t^*$$

So, by the primal constraints, we can conclude

$$\lim_{\tau\to\infty} \sum_{f_{\min}^{(\tau)} \le a < w_{\max}^{(t)}} x_{f,a}^{(\tau)} = 0,$$

and

$$\lim_{\tau\to\infty} x_{f,w_{\max}^{(t)}}^{(\tau)} = 1.$$

Therefore, for any $\epsilon > 0$ there exists a time t_ϵ where

$$x_{f,w_{\max}^{(t)}}^{(\tau)} > 1 - \epsilon, \forall \tau \ge t_\epsilon.$$

By expression 4, the offer $w_{\max}^{(t)}$ is a best-response for the firm for all $\tau \ge t$, so $x_{f,w_{\max}^{(t)}}^{(\tau)} > 1 - \epsilon$ implies

$$u_f(x_f^{(\tau)}, x_w^{(\tau)}) \ge u_f(x'_f, x_w^{(\tau)}) - \epsilon, \forall x'_f \in \Delta(\mathcal{A}).$$

Further, $x_{f,w_{\max}^{(t)}}^{(\tau)} > 1 - \epsilon$ implies the worker gets at most ϵ more utility by lowering their acceptance threshold from $w_{\max}^{(t)}$, so we also have

$$u_w(x_f^{(\tau)}, x_w^{(\tau)}) \ge u_w(x_f^{(\tau)}, x'_w) - \epsilon, \forall x'_w \in \Delta(\mathcal{A}).$$

Therefore, the strategy profile $(x_f^{(t_\epsilon)}, x_w^{(t_\epsilon)})$ is an ϵ-mixed NE. □

Finally, we prove the main theorem of this section.

Theorem 11. *Suppose agents learn strategies for $\mathcal{G}^{(1)}$ using Algorithm 1 with $\alpha_i = \mathbf{0}$, any $\eta > 0, D > 2$, and arbitrary initial conditions $x_w^{(1)}, x_f^{(1)} \in \Delta(\mathcal{A})$. Then, for any $\epsilon > 0$, there exists a finite time t_ϵ where $(x_f^{(\tau)}, x_w^{(\tau)})$ is in ϵ-Nash Equilibrium for all $\tau \geq t_\epsilon$.*

Proof. To prove the theorem, we will show that, regardless of the initial conditions, the agents must always reach or approach a mixed Nash Equilibrium (NE) asymptotically, such that we can conclude the agents end in an ϵ-NE at the last iterate.
We begin by describing all the possible conditions the agents' strategy profile, $(x_f^{(t)}, x_w^{(t)})$, could satisfy at any time t. Then, we use induction to show there is always a finite time where the agents are in one of two conditions for all future time steps. We conclude by showing that this implies the agents have converged to an ϵ-NE for any $\epsilon > 0$.
To begin, at any time step t,

$$w_{\max}^{(t)} = \frac{k}{D},$$

for some $k \in \{1, \ldots, D\}$. Then, exactly one of the following conditions is satisfied by the agents' strategy profile at time t.

1. $w_{\max}^{(t)} < f_{\min}^{(t)}$
2. $w_{\max}^{(t)} = f_{\min}^{(t)}$ and $x_{w, w_{\max}^{(t)}}^{(t)} < \frac{1}{D-k+1}$
3. $f_{\min}^{(t)} < w_{\max}^{(t)}$ and $x_{w, w_{\max}^{(t)}}^{(t)} < \frac{1}{D-k+1}$
4. $w_{\max}^{(t)} = f_{\min}^{(t)}$ and $x_{w, w_{\max}^{(t)}}^{(t)} \geq \frac{1}{D-k+1}$
5. $f_{\min}^{(t)} < w_{\max}^{(t)}$ and $x_{w, w_{\max}^{(t)}}^{(t)} \geq \frac{1}{D-k+1}$

Now, we consider each condition separately, and show the possible conditions that can be satisfied in time step $t+1$, given the condition satisfied at time step t. We say the agents move to condition i if $(x_f^{(t+1)}, x_w^{(t+1)})$ satisfies condition i for $i \in \{1, 2, 3, 4, 5\}$.

First, suppose $(x_f^{(t)}, x_w^{(t)})$ is in **condition 1**. Then, in the next time step, either $f_{\min}^{(t+1)} > w_{\max}^{(t+1)}$ and the agents remain in condition 1 or $f_{\min}^{(t+1)} \leq w_{\max}^{(t+1)}$ and the agents move to condition 2, 3, 4, or 5.

Next, suppose $(x_f^{(t)}, x_w^{(t)})$ is in **condition 2**. First, by Lemma 3,

$$x_w^{(t+1)} = x_w^{(t)},$$

so it cannot be the case that the agents move to condition 4 or 5. If, $f_{\min}^{(t+1)} > w_{\max}^{(t+1)}$, then the agents move to condition 1. Next, if $f_{\min}^{(t+1)} = f_{\min}^{(t)}$, then the agents remain in condition 2. Finally, if $f_{\min}^{(t+1)} < f_{\min}^{(t)}$, then this implies $f_{\min}^{(t+1)} < w_{\max}^{(t+1)}$ since $f_{\min}^{(t)} = w_{\max}^{(t)} = w_{\max}^{(t+1)}$ and the agents move to condition 3.

Next, suppose $(x_f^{(t)}, x_w^{(t)})$ is in **condition 3**. First, by Lemma 4,

$$x_{w,w_{\max}^{(t)}}^{(t+1)} < x_{w,w_{\max}^{(t)}}^{(t)}.$$

If $x_{w,w_{\max}^{(t)}}^{(t+1)} = 0$, then by definition $w_{\max}^{(t+1)} \neq w_{\max}^{(t)}$, so by Lemma 5, $w_{\max}^{(t+1)} < w_{\max}^{(t)}$. Now, the agents can move to condition 1, 2, 3, 4, or 5, with the new $w_{\max}$ value. Otherwise, if $x_{w,w_{\max}^{(t)}}^{(t+1)} > 0$, then by Lemma 5, $w_{\max}^{(t+1)} = w_{\max}^{(t)}$. Since $x_{w,w_{\max}^{(t)}}^{(t+1)} < x_{w,w_{\max}^{(t)}}^{(t)}$, then the agents cannot move to condition 4 or 5. Further, since $w_{\max}^{(t+1)} = w_{\max}^{(t)}$, then by Lemma 8, $f_{\min}^{(t+1)} < w_{\max}^{(t+1)}$, and the agents cannot move back to condition 1 or 2. So, the agents remain in condition 3 in this case.

Next, suppose $(x_f^{(t)}, x_w^{(t)})$ is in **condition 4**. First, by Lemma 3

$$x_w^{(t+1)} = x_w^{(t)},$$

so the agents cannot move to condition 2 or 3. Next, if $f_{\min}^{(t+1)} > w_{\max}^{(t+1)}$, then the agents move to condition 1. If $f_{\min}^{(t+1)} = f_{\min}^{(t)}$, then the agents remain in condition 4. Otherwise, if $f_{\min}^{(t+1)} < f_{\min}^{(t)}$, then since $f_{\min}^{(t)} = w_{\max}^{(t)} = w_{\max}^{(t+1)}$, this implies $f_{\min}^{(t+1)} < w_{\max}^{(t+1)}$ and the agents move to condition 5.

Next, suppose $(x_f^{(t)}, x_w^{(t)})$ is in **condition 5**. First, by Lemma 4,

$$x_{w,w_{\max}^{(t)}}^{(t+1)} < x_{w,w_{\max}^{(t)}}^{(t)}.$$

If $f_{\min}^{(t+1)} > w_{\max}^{(t+1)}$, then the agents move to condition 1. Next, if $x_{w,w_{\max}^{(t)}}^{(t+1)} \geq \frac{1}{D-k+1}$ and $f_{\min}^{(t+1)} < w_{\max}^{(t+1)}$, then the agents remain in condition 5. Next, if $x_{w,w_{\max}^{(t)}}^{(t+1)} \geq \frac{1}{D-k+1}$ but $f_{\min}^{(t+1)} = w_{\max}^{(t+1)}$, then the agents move to condition 4. Next, if $0 < x_{w,w_{\max}^{(t)}}^{(t+1)} < \frac{1}{D-k+1}$, then the agents move to condition 2 or 3. Finally, if $x_{w,w_{\max}^{(t)}}^{(t+1)} = 0$ then by Lemma 5,

$$w_{\max}^{(t+1)} < w_{\max}^{(t)},$$

and the agents move to condition 1, 2, 3, 4, or 5.

Next, we show there exists a finite time where the agents remain in condition 4 or condition 5 for all future time steps. First, Lemma 5 shows that the value of

$w_{\max}$ is non-increasing in all time steps, and further the number of $w_{\max}$ values is $|\mathcal{A}|$. So, it suffices to show for each unique $w_{\max}$ value that either the agents must remain in condition 4 or condition 5 for all future time steps or the value of $w_{\max}$ must decrease in finite time.

Suppose $(x_f^{(t)}, x_w^{(t)})$ is in condition 1 at time t. From the cases above, the agents either remain in condition 1 or move to one of the other conditions, and Lemma 6 shows that it takes finite time for the agents to move to condition 2, 3, 4, or 5. Further, Lemma 7 shows that once agents leave condition 1 for the first time per unique value of $w_{\max}$, it takes finite time to ensure the agents never enter condition 1 again for that $w_{\max}$ value. So, we may assume that the agents never enter condition 1 for the remainder of the cases. Next, if the agents are in condition 4 or 5, but don't stay there for all future time steps and $w_{\max}$ does not decrease, then there must be a finite time where the agents move to condition 2 or 3. Next, if the agents are in condition 2, then from the cases above, they either remain there or move to condition 3, and Lemma 8 shows that it takes finite time for the agents to move to condition 3. Then, from the cases above, agents must stay in condition 3 until $w_{\max}$ decreases in value, and Lemma 9 shows it takes finite time for $w_{\max}$ to decrease. Therefore, for each unique $w_{\max}$ value, either it takes a finite amount of time for its value to decrease, or the agents never leave condition 4 or 5.

Finally, in the base case, suppose at time t, $w_{\max}^{(t)} = \frac{1}{D}$. Note that since

$$U_{w,0}^{(t)} = U_{w,\frac{1}{D}}^{(t)}, \forall t$$

then by Claim 2, this is the smallest value in $\mathcal{A}$ that $w_{\max}^{(t)}$ can be. By definition of $w_{\max}$, this implies $x_{w,\frac{1}{D}}^{(t)} \geq \frac{1}{2}$ which satisfies the probability mass lower bound of condition 4 and 5 for $D > 2$. Then, Lemma 6, along with the fact that the lower bound of $f_{\min}^{(t)}$ is also $\frac{1}{D}$, shows that it takes finite time for $f_{\min}^{(t)} = w_{\max}^{(t)}$. Therefore, the conditions for the agents being in condition 4 are satisfied at the lowest value of $w_{\max}$. So, we can conclude there is always a finite time where the agents are in condition 4 or condition 5 for all future time steps.

To finish the proof, Lemma 10 shows that if agents are either in condition 4 or condition 5 for all future time steps, they must converge to an ϵ-mixed NE for any $\epsilon > 0$. □

B Sequence Form Representation of 2-Round Alternating Bargaining Game

A *sequence* σ is a sequential string of actions an agent must take to get to some node in the game tree. For example, if agents are at the payoff node $(1 - a_i, a_i)$, then the sequence the firm took is $\sigma_f = a_i$ and the sequence the worker took is $\sigma_w = A_{a_i}$. The sequences and associated payoffs for the two round bargaining game parameterized by discount factor $0 < \delta < 1$ are given in the table below with the firm's sequences on the rows and the worker's sequences on the columns where $a, b \in \mathcal{A}$ are arbitrary offers in $\mathcal{A}$. The line $-$ indicates that the combination of sequences does not result in a terminal node.

	A_a	$R_a b$
a	$(1-a, a)$	$-$
aA_b	$-$	$\delta(b, 1-b)$
aR_b	$-$	$\delta(0,0)$

Let Σ_i be the set of all terminal sequences of agent i for $i \in \{f, w\}$ and let $\emptyset$ represent the root node of the extensive form game tree. Then,

$$\Sigma_f = \{\emptyset, a, aA_b, aR_b | a, b \in \mathcal{A}\}$$
$$\Sigma_w = \{\emptyset, A_a, R_a b | a, b \in \mathcal{A}\}$$

Next, let I_i be the information set of agent i. Since our game is complete information and perfect recall, for each $I \in I_i$, I is a singleton set with one node h and, further, there is a unique sequence $\sigma_i \in \Sigma_i$ that leads to h. For each $I \in I_i$, let $\texttt{ext}(I)$ be the set of extensions of the unique sequence $\sigma_h \in \Sigma_i$ that leads to the node $h \in I$ by 1 valid action in Σ_i. For example, if $h \in I$ is the node corresponding to the firm responding to a counteroffer of $b \in \mathcal{A}$ from the worker after giving an initial offer of $a \in \mathcal{A}$, then $\sigma_h = a$ is the unique sequence leading to the node $h \in I$ and

$$\texttt{ext}(I) = \{aA_b, aR_b | b \in \mathcal{A}\}.$$

Next, a *realization plan* represents the probability mass an agent puts on reaching each terminal sequence. Formally, $r_i : \Sigma_i \to [0, 1]$ such that

$$\begin{aligned} r_i(\emptyset) &= 1 & \\ \sum_{\sigma^+ \in \texttt{ext}(I)} r_i(\sigma^+) &= r_i(\sigma) & \forall I \in I_i \\ r_i(\sigma) &\geq 0 & \forall \sigma \in \Sigma_i \end{aligned}$$

From a realization plan r_i, a behavioral strategy of agent i can be recovered. Let $\sigma_i \in \Sigma_i$ where σ_i is the unique sequence leading to $I_{\sigma_i} \in I_i$. Then, let $\sigma_i a_i \in \texttt{ext}(I_{\sigma_i})$, and the behavioral strategy at action a_i is:

$$\beta_i(I_{\sigma_i}, \sigma_i a_i) = \frac{r_i(\sigma_i a)}{r_i(\sigma_i)}.$$

Let $U_{i,\sigma}^{(t)}(\{r_{-i}^{(\tau)}\}_{1 \leq \tau \leq t})$ be the cumulative expected utility agent i gets at terminal sequence $\sigma \in \Sigma_i \setminus \{\emptyset\}$ through time t:

$$U_{i,\sigma}^{(t)}(\{r_{-i}^{(\tau)}\}_{1 \leq \tau \leq t}) = \sum_{\tau=1}^{t} u_i(\sigma, r_{-i}^{(\tau)})$$

where, for all first round offers $a \in \mathcal{A}$ and second round offers $b \in \mathcal{A}$,

$$u_f(\sigma, r_w^{(\tau)}) = \begin{cases} (1-a) \cdot r_w^{(\tau)}(A_a) & \sigma = a \\ \delta \cdot b \cdot r_w^{(\tau)}(R_a b) & \sigma = aA_b \\ \delta \cdot 0 \cdot r_w^{(\tau)}(R_a b) & \sigma = aR_b \end{cases}$$

and

$$u_w(r_f^{(\tau)}, \sigma) = \begin{cases} a \cdot r_f^{(\tau)}(a) & \sigma = A_a \\ \delta \cdot (1-b) \cdot r_f^{(\tau)}(aA_b) & \sigma = R_a b \end{cases}$$

For ease of notation, we will shorten the cumulative expected utility of agent i at terminal sequence σ to $U_{i,\sigma}^{(t)}$ and we will refer to the realization plan mass that agent i puts on terminal sequence σ at time t as $r_{i,\sigma}^{(t)}$. Then, $U_i^{(t)}$ is the cumulative expected utility vector of agent i at time t and $r_i^{(t)}$ is the realization plan of agent i at time t. Finally, let $\mathcal{Q}_i$ be the set of valid realization plans of agent i. We abuse notation slightly and suppose $r \in \mathcal{Q}_i$ is represented as a vector. Then, the expected utility of a realization plan, given a cumulative expected utility vector $U_i^{(t)}$, can be denoted as

$$\langle U_i^{(t)}, r \rangle.$$