Biased-Belief Equilibrium in Finite Dynamic Games with Stochastic Private States

by

Vladimir Khismatullin

Abstract

This paper addresses a Bayesian persuasion problem in multi-stage games with public actions and private information evolving over time. We first demonstrate that solutions in pure strategies are infeasible for such settings and derive an alternative solution based on the concept of biased beliefs — tendencies observed in real-world players, who seek to derive private information from public actions in a predictable way. Finally, we analyze the conditions under which normal players achieve equilibrium and identify key factors that provide significant advantages in these multi-stage scenarios.

Contents

1 Introduction and Setup

1.1 Introduction

Most long-term games in life involve hidden states, where information is gradually revealed through the observable actions of individuals. These actions depend on a combination of private information, changes in the state of nature, and publicly available information. Such interactions can be thought of as a game with a sequential increase in information. Examples include dating, legal proceedings, renting an apartment, updating a professor on research progress, and poker.

Although this framework appears straightforward, the presence of private information creates a strategic challenge to players. What happens is that the first player seeks to achieve three goals: infer the opponents private information as soon as possible, hide her own information, and maximize the final utility. Clearly all 3 objective are non-independent and an increase in one of them should usually lead to a decrease in others. This, however, is not the biggest complication, as the incentives of the second player, affect the first player, in turn affecting the second player and so on. In other words, the optimal solution is implicitly defined.

Another complication in real-life games is the existence of biases. Every information-based decision-making is subject to personal bias. These personal biases play a significant role in shaping equilibrium-type outcomes in real life (Al-Nowaihi and Dhami 2015). For instance, a rental agency that perceives each customer's credibility pessimistically, may sell a lower total number of contracts, thus under-performing financially, yet still achieve equilibrium (i.e., sustain itself long-term). However, many a time a real-estate owner interested in finding the optimal size of kitchens, number of sinks and width of doors may spend too much time focusing on what is irrelevant, losing all of her potential customers.

The key finding of this paper is that, while under private information, solutions to such sequential games are defined implicitly and are thus impossible to obtain for long games, an equilibrium can be easily derived in a practically relevant setting under the assumption of biased beliefs about the private states. In our model players still play a game in extensive-form pure strategies and derive private information about the other player's state based on his actions, however, they do so based on a strongly biased view. Thus, seemingly complicating the basic game, we make the solution tractable.

1.2 Relevant work

From the mathematics point of view, this model without a biased belief is a generalization of the setup from Billings et al. 2003. As authors point out, such games are typically solved in extensive-form randomized mixed strategies. However, in most problems where action spaces and state spaces are non-trivial, this leads to exponential blow-up in complexity. As we will show, in cases where actions-state space is continuous, this does not look analytically feasible. Using the biased-belied idea (Heller and Winter 2020) we kill two birds with one stone, making the solution both tractable in pure strategies and polynomially complex for continuous action-state space.

It should be further noted that games with private information are well-covered in the existing economics literature. In "Universal Games of Incomplete Information," Reif 1979 studies two-player games where each player has private information unknown to their opponent. He introduces universal games to represent the complexity of incomplete-information scenarios and shows that solving these games is computationally intensive, often requiring doubly-exponential time. Reif also studies "blindfold games," where one player's actions are restricted, proving their universality and exponential space complexity. This work highlights the computational challenges inherent in games with private information, addressed in the second section of this paper.

Cole and Kocherlakota 2001 in "Dynamic Games with Hidden Actions and Hidden States," study dynamic games where players' actions and certain state variables are unobservable to others. They develop an algorithm to identify sequential equilibria where strategies depend solely on privately observed states. This is the paper closest to the problem we address in this paper. The only difference remains in that our model allows the players to makes strategies based on all public information. The simplification here, once again, comes from the fact that players induce biases in their judgment, making far from optimal solutions.

In another paper by Yamamoto 2019 infinite-horizon games are studied. Players observe actions and receive noisy public signals about an unobserved state. He identifies conditions under which the set of feasible and individually rational payoffs remains consistent, regardless of initial state beliefs, as players become more patient. While inspiring, the result over infinite horizon differs significantly from our problem. We use the concept of a public signal representing the private state with some added noise but assume that players interpret the public signal as partially indicative of the private state, rather than guaranteeing that it does reflect it. This assumption seems much more plausible than that used in their paper.

To summarize, all existing models lack flexibility in at least one of the following three areas: accommodating continuous states and continuous actions, handling games that extend beyond a single period, and incorporating public information. The combination of the latter two is crucial for achieving optimality, while disallowing for continuous states significantly complicates derivations, making them polynomially intractable.

In an attempt to simplify the problem, we complicate the framework by introducing biased beliefs, which have shown to significantly enrich the space of possible equilibria. However, none of the existing frameworks (Caragiannis et al. 2014, Haan and Hauck 2023 and Heller and Winter 2020) approach the bias the same way as it is known in sociology. Both Caragiannis et al. 2014, Haan and Hauck 2023 introduce the bias with regards to the optimal solution, while Heller and Winter 2020 defines it as a possibly personal feature, although, once again, goes on to center the idea around the deviation from the optimal solution. We argue that real-life biases do not come as a deviation from the optimal solution, rather they imply a new set of solutions, possibly highly chaotic. In fact, these biases should be modeled more like the asymmetrical information model from Fershtman and Pakes 2012, where actions are biased by ones limited knowledge.

In order to define the appropriate extensive-form game, we use the well-known signaling idea, reflected in Bayesian persuasion (Kamenica and Gentzkow 2011) combined with the decision-analysis framework from Bordon and Fu 2015 (in contrast to Doepke and Townsend 2006) and extend it to a multi-period case when both players have private information that changes over time. To solve such a complicated problem, we make an assumption on the form of utility function that is reflective of multi-stage betting games with private information, where players participate for a given number of rounds. However, that proves insufficient and we implement the idea from the biased-belief equilibrium (Heller and Winter 2020), in a way that does not mutilate the perceived action, rather the perceived intent.

1.3 Explanatory example

This paper does not solve a specific problem, rather provides a theoretical framework for problem solution. That being said, we should mention the example that led to the creation of this model. The game of heads-on Texas hold'em poker (Bowling et al. 2015) is a game in which two players get a pair of cards that is known only to them. One player is randomly selected to be the leading player – the first player to make a move, while the other player always makes a move after him. The game is played in four stages, each beginning with more cards being dealt, updating both players' beliefs about the likelihood of winning. At the beginning of each round, the leading player can make a bet of their liking, after which the other player makes the same decision. The incentive for both players is to make a bet maximizing their winnings. In order to simplify math and make analytical derivations tractable for a general case, we simplify the rules of the game by allowing each player to bet regardless of the other player's move. That is, both players do not have to call the other players bet, leading to each stage consisting of two bets and two bets only. That being said, the case of real games where folding is allowed gets addressed in the extension in the last section of the paper.

1.4 Setup

We study a game, described by the following: We consider a 2-player game, where players take turns one after another. The player to start the game — **action-taker or leader** — observes his hidden state S_0 and makes an action a_0. The second player — **responder or follower** — observes her hidden state Z_0 and makes a response r_0. The game then continues for T more periods, until it reaches a terminal state at time $t = T$, at which point the information is revealed and players obtain their utility based on both hidden states. We model the utility to be a multiple of players' total bets and chances of winning. The information flow in this game is illustrated by graph 4, for a specific case of game with four stages ($T = 3$). The **assumptions** are as follows:

- The initial state (S_0 for the action taker and Z_0 for the responder) is drawn from the same prior distribution. Once again, this information is purely private. At the beginning of each stage, the hidden does not get reset completely, instead, it changes according to a Markov chain rule for hidden states:

$$S_0 \sim \mathbb{P}(S_0); \qquad Z_0 \sim \mathbb{P}(Z_0); \qquad S_{t+1} \sim \mathbb{P}(S_{t+1}|S_t); \qquad Z_{t+1} \sim \mathbb{P}(Z_{t+1}|Z_t)$$

  Both players have perfect knowledge about the prior distribution and the chain rule.
- Players observe each other's actions and derive utility based on the true value of the final hidden state in this zero-sum game. Denote the history of actions made by the sender and receiver respectively as : $A_{:t} = (a_0, a_1, \dots a_t)$ and $R_{:t} = (r_0, r_1 \dots r_t)$, then:

$$U_{\text{action-taker}} = U(A_{:T}, R_{:T}, S_T, Z_T) = -U_{\text{responder}}$$

  We will further specify this utility as a product of players winning chances and the total amount bet with some regularization.
- In order to derive the hidden state of their opponent, a player forms a belief about the underlying state based on their belief about the other person's private state and the action they take. The notation is as follows: at time t player 2 will have belief $\mathcal{B}_t$ about the state of player 1 — S_t,

whereas player 1 will have belief $\mathcal{G}_t$ about Z_t. Note that the first of the two beliefs forms after player 1 takes action, whereas the second forms only once the second player takes action. Due to the Markovian property 4,

$$\mathcal{B}_{t+1} = \mathcal{B}_{t+1}(a_t, \mathcal{B}_t); \qquad \mathcal{B}_0 = \mathcal{B}_0(a_0)$$

$$\mathcal{G}_{t+1} = \mathcal{G}_{t+1}(a_t, \mathcal{G}_t); \qquad \mathcal{G}_0 = \mathcal{G}_0(a_0)$$

Beliefs represent a stochastic information structure of each player, reflective of the information they have obtained so far.

Please note that the hidden state does not represent a specific outcome (a poker hand together with the community cards), rather a relative strength of the hand. For example, if a player is dealt pocket nines (S_0 is very high, > 2), his relative strength will go down ($S_1 < 2$) once he misses the flop (the flop comes off Ace, King, Queen suited).

1.5 Simplifying assumptions

For the rest of this paper, we would like to manifest some assumptions which add specificity to the model, making the derivations tractable. We use a utility function that postulates a bound on the potential bets by adding a risk-aversion regularization term. As we will see, this leads to a Kelly-like criterion, where the amount bet is proportional to the perceived chance of winning.

- States can be represented by an infinite continuous distribution and are homogeneous among players, moreover, we fix the prior to be Normal (0,1):

$$\mathrm{dom}(S) = \mathrm{dom}(Z) = \mathbb{R} \quad \text{Moreover, } \mathbb{P}_S(\cdot) = \mathbb{P}_Z(\cdot) = \mathcal{N}(\cdot|0,1) \tag{1}$$

- Actions can be represented by an unbounded continuum

$$\mathrm{dom}(a) = \mathrm{dom}(Z) = \mathbb{R}$$

- To derive a specific form and payoff, we define the utility function as:

$$U = \left(\sum_{t=0}^{T}(a_t + r_t)\right) \cdot (S_T - Z_T) + \frac{1}{2\gamma_r}\sum_{t=0}^{T} r_t^2 - \frac{1}{2\gamma_a}\sum_{t=0}^{T} a_t^2 \tag{2}$$

Note that in the case of a single-period game, it simplifies to:

$$U = (a_0 + r_0) \cdot (S_0 - Z_0) + \frac{1}{2\gamma_r} r_0^2 - \frac{1}{2\gamma_a} a_0^2 \tag{3}$$

That basically stands for betting proportional to perceived chances of winning.

As we have noted, under this set of assumptions, players win proportionally to how much better they are than their opponent. For example, in dating, one would love to stick to a more beautiful, rich, util-endowed partner. However, they know that over-investing can prove dangerous even under perfect information, as there is still uncertainty as to how long their partner lives. This is another way of saying that the hidden states only represent the relative hand strength, not the outcome. Generally, outcome is a weighting of strengths, however, we do not go into this discussion and simply assume it to be proportional to the difference $S_T - Z_T$.

2 Rational agent games

In this section we show that finding the optimal solution in both pure and mixed strategies, i.e. when both players choose their actions using all available information, is *analytically impossible*. That is, under our set of assumptions, the conditions on optimal actions are too strict.

2.1 Game in pure strategies

First note that the operated move in this case is the pure action a_t. That is, both players are looking for optimal decision rules a_t^*, r_t^*. A rational agent will always try to derive the belief based on all available information. In other words for the receiving player at time t, the belief about the state of the first player is based on the history of all available actions and known states:

$$\mathcal{B}_t = (S_t | A_{:t}, R_{:t-1}, Z_{:t-1})$$

Rational agent will try to find the optimal response under all the available information at time t, which is another way of saying that they will optimize the utility given a belief reflective of all public and private information.

What makes this problem excruciatingly hard to solve is the fact that player 1 is fully aware of the belief he imposes, and thus can mimic an action, sending a false signal to mislead player 2. However, player 2 then understands that player 1 might have made his move in an attempt to mislead, which once again affects the action 2 will take and so one. In mathematical terms, as we will see, that means that the optimal action of the first player is always defined as an implicit function.

As we prove next, under this simple utility model, there does not exist any solution. This does not imply that there is never a solution in this setup, however, it does seem very unlikely.

2.2 Single period game in pure strategies

To illustrate the problem under consideration, we first solve the game in pure strategies that lasts for only one stage — where players participate only during the starting period.

Such decision-making problems are always solved recursively from the end. The first step is thus to find the optimal action of the last player $r_0^*(Z_0, a_0)$ under the utility 3, with additional uncertainty, dictated by the privacy of state S_0

$$\begin{aligned} r_0^* &= \arg\min_{r_0} \mathbb{E}_{S|a_0} \left\{ (a_0 + r_0) \cdot (S_0 - Z_0) + \frac{1}{2\gamma_r} r_0^2 - \frac{1}{2\gamma_a} a_0^2 \right\} \\ &= \arg\min_{r_0} \left\{ (a_0 + r_0) \cdot (\mathbb{E}(S_0|a_0) - Z_0) + \frac{1}{2\gamma_r} r_0^2 \right\} \end{aligned}$$

For which we immediately note, that the second derivative with respect to r_0 is equal to $\frac{1}{\gamma_r}$, and thus, a unique solution exists, defined by the FOC:

$$(\mathbb{E}(S_0|a_0) - Z_0) + \frac{1}{\gamma} r_0 = 0 \implies r_0 = -\gamma(\mathbb{E}(S_0|a_0) - Z_0)$$

That is, the receiver will bet proportional to his chances of winning and scaled by his risk aversion. Remind yourself that $S_0 - Z_0$ stands for the expected frequency of winning. Then this result is similar to Kelly criterion. For now, ignore the definition of $\mathbb{E}(S_0|a_0)$, as it is simply a function of a_0, we proceed to the next step: maximizing the utility by the leading player:

$$
\begin{aligned}
a_0^* &= \arg\max_{a_0} \left\{ \mathbb{E}_z(a_0 - \gamma_r(\mathbb{E}(S_0|a_0) - Z_0)) \cdot (S_0 - Z_0) + \frac{1}{2\gamma_r}\mathbb{E}_z(-\gamma_r(\mathbb{E}(S_0|a_0) - Z_0))^2 - \frac{1}{2\gamma_a}a_0^2 \right\} \\
&= \arg\max_{a_0} \Big\{ a_0 \cdot (S_0 - \mathbb{E}Z_0) - \gamma_r s\mathbb{E}_z(\mathbb{E}(S_0|a_0) - Z_0) + \gamma_r\mathbb{E}_z(\mathbb{E}(S_0|a_0)Z_0 - z_0^2) + \\
&+ \frac{\gamma_r}{2}\mathbb{E}_z(\mathbb{E}(S_0|a_0) - Z_0)^2 - \frac{1}{2\gamma_a}a_0^2 \Big\} \\
&= \text{ Under the normality assumption it simpliefies to } = \\
&= \arg\max_{a_0} \left\{ a_0 S_0 - \gamma_r s\mathbb{E}(S_0|a_0) + \gamma_r 1 + \frac{\gamma_r}{2}((\mathbb{E}(S_0|a_0))^2 + 1)) - \frac{1}{2\gamma_a}a_0^2 \right\} \\
&= \arg\max_{a_0} \left\{ a_0 S_0 - \gamma_r s\mathbb{E}(S_0|a_0) + \frac{\gamma_r}{2}(\mathbb{E}(S_0|a_0))^2 - \frac{1}{2\gamma_a}a_0^2 \right\}
\end{aligned}
$$

Now, the question is, where does the $\mathbb{E}(S_0|a_0)$ come from. The answer to this question lies in the Bayes formula:

$$
\mathbb{P}(S_0|a_0) = \frac{\mathbb{P}(a_0|S_0)\mathbb{P}(S_0)}{\mathbb{P}(a_0)} = \text{use knowledge of optimal strategy} = \frac{[a^*(S_0) = a_0]\mathbb{P}(S_0)}{\int_{s|a^*(s)=a_0} dF(s)}
$$

That, in effect, means that

$$
a_0^* = \arg\max_{a_0} f(a_0, S_0, \gamma_r, \gamma_a, a_0^*(\cdot, \cdot))
$$

Thus, the optimal first action is an implicit function of several factors. Under some functional form constraints and smoothness assumptions, this does have a solution. However, for a multi-period game with dynamic changes in private information S_t, Z_t, the authors of this paper could not find any set of assumptions that would lead to a non-degenerate model where the optimal rule is tractable.

2.3 Mixed strategy extension

Let's assume that the game is no longer in pure, but mixed strategies. That means players choose a distribution of actions: $\pi_a(a), \pi_r(r)$. In the game itself, they will report these distributions maximizing their utility given the observables, i,e,

$$
\pi_a(a_t) = \pi_a(a_t|S_{:t}, R_{:t-1}, A_{:t-1}), \qquad \pi_r(r) = \pi_r(r_t|Z_{:t}, A_{:t}, R_{:t-1})
$$

In this case the problem remains, as the posterior now has the following form:

$$
\mathbb{P}(s|a) = \frac{\mathbb{P}(a|s)\mathbb{P}(s)}{\int \mathbb{P}(a|s)\mathbb{P}(s)ds} = \frac{\pi_a(a|s)\mathbb{P}(s)}{\int \pi_a(a|s)\mathbb{P}(s)ds}
$$

Here, the optimal solution of the second player at time T depends on the posterior

$$
\pi_r(r_T|Z_{:T}, A_{:T}, R_{:T-1}) = \pi_r(r_T|Z_{:T}, A_{:T}, R_{:T-1}; \pi_a(a_T|S_{:T}, R_{:T-1}, A_{:T-1}))
$$

and by the definition of the game, the optimal mixed strategy is

$$\begin{aligned}\pi_a(a_T) &= \underset{\pi_a(\cdot|S_{:T},R_{:T-1},A_{:T-1})}{\arg\max} \quad \mathbb{E}_{Z|S_{:T-1},R_{:T-1},A_{:T-1}} U\{S_T, \hat{Z}_T, R_{:T-1}, A_{:T-1}; \pi_a^T, \pi_r(\cdot|Z_{:T}, A_{:T}, R_{:T-1})\} \\ &= \underset{\pi_a(\cdot|S_{:T},R_{:T-1},A_{:T-1})}{\arg\max} \quad \mathbb{E}_{Z|S_{:T-1},R_{:T-1},A_{:T-1}} U\{S_T, \hat{Z}_T, R_{:T-1}, A_{:T-1}; \pi_a^T, \\ &\qquad\qquad \pi_r(r_T|Z_{:T}, A_{:T}, R_{:T-1}; \pi_a(a_T|S_{:T}, R_{:T-1}, A_{:T-1}))\}\end{aligned}$$

which can not be independent of the choice of optimal mixed strategy. Even if an optimal solution does exist, it is infeasible to estimate in practically relevant settings for real players, as it now represents an implicitly defined functional. In reality, such layers of recursive reasoning—contemplating what others think about your thoughts regarding their thoughts, and so on—resemble an overcomplication more akin to a pathological psychiatric condition than practical rationality

By now, we have noticed that any fraction of information about optimal decision-making used by both players to derive their best response results in an implicit solution. In the next section, we introduce the concept of biased belief to model real-life decision-making, where players take shortcuts and deliberately avoid overly complex reasoning.

3 Biased beliefs games

As we noticed before, rational players have to find implicitly defined functions over continuum spaces to derive their beliefs, which likely does not happen in practice. In this case, they used all the available information known to the moment of action, in order to derive the underlying private information. The optimal policy was implicitly defined due to the dependence of the belief on optimal policy:

$$\mathcal{B}_t = \mathcal{B}_t(Z_{:t-1}, A_{:t}, R_{:t-1}) = \mathcal{B}_t(Z_{:t-1}, A_{:t}, R_{:t-1}, A^*_{:t}) \xrightarrow{\text{defines}} \mathbb{P}(S_t | Z_{:t-1}, A_{:t}, R_{:t-1}, A^*_{:t})$$

In real world, players rarely think that deep, usually reducing their belief about the opponent's strength to the commonly known optimal decision-making rule. So, in order to model their behavior we should simplify the belief structure, **by assuming that each player has their unique way of formulating beliefs, independent of another player's optimal strategy with respect to this individual**. This is equivalent to saying that one formulates their beliefs based on the style of the player, but not his whole thought process.

This does not exclude the possibility of adjusting one's strategy to another player. Formally, first we define each player to be described by a set of characteristics Θ_i, which, for example, include the inverse risk-aversion γ_i from Utility 1 (3). Next, for a player i playing against player j, we define the biased belief as a distribution over perceived underlying states

$$\hat{S}_t \xleftarrow{\text{implied by}} \mathcal{B}_t = f_i(\Theta_j)(Z_{:t-1}, A_{:t}, R_{:t-1}), \text{ where } f_i \text{ is a known functional}$$

In such case, the way of formulating belief is restricted to a functional f_i of opponents characteristics Θ_j. Note that the 'implied by' highlights the fact that one's beliefs might describe more than the posterior distribution of the state given history.

Essentially, the **Biased belief** simplifies the optimization procedure from finding a set of 2 T functions that are $2T$-implicit, to recursively solving a maximization problem over a set of hyper-parameters.

Example 1:

We call player naive if he associates every state with one and one action only

$$S_t \sim f(\Theta_j)(Z_{:t-1}, A_{:t}, R_{:t-1}) = f(a_t), \text{ such that } \exists f^{-1}(a_t)$$

In this case his belief collapses whenever his opponent makes a move. And, as such, history becomes irrelevant. In the first set of assumptions, the ultimately naive player will identify the state S_t equal to the action a_t, i.e. $f(\cdot) = I_{\text{Identity}}$.

In dating, an ultimately naive player would be an equivalent of someone judging their partner's love by the monetary value of their presents. This behavior neither accounts for any of the characteristics of this human being, nor for their potential strategies. Nonetheless, two blindly naive partners will easily reach an equilibrium, by making gifts of constant value, independent of their partners characteristics, or past history.

Example 2:

Assume each player has some factor of unpredictability σ_i, in such case, a σ-naive player (or a normal player) is a player who forms his belief about the opponents state around the action with some some deviation, described by his opponent's level of uncertainty:

$$a_t \sim f(\Theta_j)(Z_{:t-1}, A_{:t}, R_{:t-1}) = f(\sigma_j)(Z_{:t-1}, A_{:t}, R_{:t-1}) = f(\sigma_j)(a_t) = \mathcal{N}(a_t|S_t, \sigma_j^2)$$

That is, the player expects his opponent to make an action distributed normally with mean S_t and variance defined by their unpredictability factor σ. Note that the distribution $\mathbb{P}(S_t|a_t)$ can be derived from the belief using Bayes formula and that this does not lead to oversimplification of the time component, as all history should still be used to derive the belief about the true state:

$$\mathbb{P}(S_t|A_{:t}) = \mathbb{P}(S_t|a_0, a_1, \dots, a_t) \neq \mathbb{P}(S_t|a_t)$$

A typical sigma-naive person will start a relationship by learning their partner's unpredictability and proceed accessing the difference between their and partner's quality. They will be certain that the partner's move is somewhat reflective of the true state, and that averaging them over time will reveal the true quality. It is worth noting that they will only do so under some conditions on the dynamics of the underlying private information. If your potential husband works as a banker, his quality is unlikely to change as much as if he was an undercover agent.

3.1 Simplifying conditions

To solve the game with Utility 1, we introduce two key concepts: the linearity of perceived state and reduced form optimal actions. The first addresses how real players typically form beliefs about their opponents based on past actions. The second outlines how players anticipate and plan their future moves.

3.1.1 The linearity of perceived state condition:

The linearity of perceived state is an important property of biased beliefs. It states that at any point in time t, a player will perceive his opponents state as linear in the opponent's past actions. Formally, $\forall t \geq 0,\ \forall k \geq 0$:

$$\mathbb{E}(S_{t+k}|A_{:t}) = \phi_a(t,k,0)a_0 + \phi_a(t,k,1)a_1 + \cdots + \phi_a(t,k,t-1)a_{t-1} + \phi_a(t,k,t)a_t \tag{4}$$

$$\mathbb{E}(Z_{t+k}|R_{:t-1}) = \phi_r(t,k,0)r_0 + \phi_r(t,k,1)r_1 + \cdots + \phi_r(t,k,t-1)r_{t-1} \tag{5}$$

When thinking about the opponent's state, it is important that the player only considers the opponent's actions. While this assumption is realistic in the set of assumptions 1, it might not be in other sets. For utility 1, players bet proportional to their advantage, implying that the true expected advantage is, in turn, linear in action. We will prove this later.

While this condition does not directly address the feasibility of optimal rules, it proves very helpful in out model. In case when both players only consider the expectation of the state, this assumption excludes the problem we discussed in the last section, where the distribution of the belief was dependent on the optimal action a^*, r^*, making solution tractable in this case.

3.1.2 Reduced form optimal action:

In any extensive-form game, any optimal non-degenerate action depends on the future optimal action. This subsection of the paper, along with the Appendix focuses on the problem of deriving the formulas necessary to simplify the functional form and derivation of the extensive-form decision-making rules. This starts by assuming that periods $T, T-1, \dots k+1, k$ of the game have been solved and the formula for optimal actions at time $k-1$ is yet to be derived.

Suppose that it is currently turn k, with player 1 making his decision. In order to derive the best action, as we will see, he needs to use all future optimal decision-making rules for times $t > k$.

$$r^*_t(Z_t, A_{:t}, R_{:t-1}) \qquad a^*_t(S_t, A_{:t-1}, R_{:t-1})$$

The question arises, as to what this rule looks like when we substitute all r_t, a_t for $t > k$ with their optimal values. Obtaining a rule that is dependent on the history publicly available at time k and future possible values of the observable private information in $Z_k, Z_{k+1}, \dots, Z_T, S_k, S_{k+1}, \dots, S_T$.

Main lemma:

Assume that for all $t > t_0$ the optimal decision-making rule is linear in the underlying state and actions, that is,

$$a^*_t(S_t, A_{:t-1}, R_{:t-1}) = \eta_t S_t + \sum_{j=0}^{t-1}(a_j\alpha^t_j + r_j\beta^t_j) \tag{6}$$

$$r^*_t(Z_t, A_{:t}, R_{:t-1}) = \nu_t Z_t + \sum_{j=0}^{t-1}(a_j\hat{\alpha}^t_j + r_j\hat{\beta}^t_j) + a_t\hat{\alpha}^t_t \tag{7}$$

Than the reduced form of future optimal action is also linear with coefficients $w(i,j,k,b)$, that is, the future optimal rules perceived by players at time k, when player 1 has not made a move yet (and thus a_k is not known) can be written as:

$$a^*(t,k,0) = \sum_{j=k}^{t}\eta_a(t,k,j,0)S_j + \sum_{j=k}^{t-1}\nu_a(t,k,j,0)Z_j + \sum_{j=0}^{k-1}\left(w^a_a(t,k,j,0)a_j + w^r_a(t,k,j,0)r_j\right) \tag{8}$$

$$r^*(t,k,0) = \sum_{j=k}^{t}\eta_r(t,k,j,0)S_j + \sum_{j=k}^{t}\nu_r(t,k,j,0)Z_j + \sum_{j=0}^{k-1}\left(w^a_r(t,k,j,0)a_j + w^r_r(t,k,j,0)r_j\right) \tag{9}$$

for some coefficients w. Once the move has been made by the first player, the information about private state S_k is used in a_k and the functional form changes to

$$\begin{aligned} a^*(t,k,1) = &\sum_{j=k+1}^{t}\eta_a(t,k,j,1)S_j + \sum_{j=k}^{t-1}\nu_a(t,k,j,1)Z_j + \sum_{j=0}^{k-1}\left(a_j w^a_a(t,k,j,1) + r_j w^r_a(t,k,j,1)\right) \\ &+ \delta_a(t,k)a_k \end{aligned} \tag{10}$$

$$\begin{aligned} r^*(t,k,1) = &\sum_{j=k+1}^{t}\eta_r(t,k,j,1)S_j + \sum_{j=k}^{t}\nu_r(t,k,j,1)Z_j + \sum_{j=0}^{k-1}\left(a_j w^a_r(t,k,j,1) + r_j w^r_r(t,k,j,1)\right) \\ &+ \delta_r(t,k)a_k \end{aligned} \tag{11}$$

On the left-hand side of each equation the first index t denotes the index of the rule that the player is interested in (for example he wants to know what his optimal decision-making rule looks like at the 3-rd stage of the game). The second k denotes the current stage (for example, we have just seen the flop, stage=1). The last index stands for whever the first player has made a move at this time k.

The result of this lemma is very technical. Proof can be found in appendix.

Now we can show the main theorem of this paper, which derives the optimal decision-making rules and shows the conditions for the optimum to exist, i.e. the game to have a solution.

3.2 Theorem 1

Under the linearity of perceived state condition 4, 5, a unique equilibrium exists in the biased-belief game with Utility 1 (3).

Proof:

In order to prove this, we will construct the equilibrium by sequentially deriving the optimal strategies of this extensive form game starting with the final period:

$$r_T^*(Z_T, A_{:T}, R_{:T-1}), a_T^*(S_T, A_{:T-1}, R_{:T-1}), r_{T-1}^*(Z_{T-1}, A_{:T-1}, R_{:T-2}), a_{T-1}^*(S_{T-1}, A_{:T-2}, R_{:T-2}), \dots,$$
$$r_1^*(Z_1, a_1, a_0, r_0), a_1^*(S_1, a_0, r_0), r_0^*(Z_0, A_0), a_0^*(S_0)$$

The key idea of this proof is to show, starting at $t = T$ and going backwards to $t = 0$, that all these functions are linear in their parameters, deriving the coefficients of the linear terms along the way. As we will see from the form of the utility function, linearity will guarantee that under some restrictions on the parameters of the players γ_r, $\gamma_a, \sigma_a, \sigma_r$ there exists a unique set of coefficients w defining the optimal rules at each time step t.

Following the logic we used in the first part of the paper, we start at the last period and derive the optimal response at each time going backwards. We use the fact that the problem is convex for the second player and concave for the first (that follows from the derivations below).

Proof by induction:

Base case: $t = T$

For the last time step:

$$r_T(Z_T, A_{:T}) = \arg\min_{r_T} \mathbb{E}_{S_T|A:T} \left(\sum_t a_t + r_t \right) (S_T - Z_T) + \frac{1}{2\gamma_r} r_T^2$$

Taking the derivative, we see that the second derivative is, obviously positive for positive γ_r, which means that the optimal argument of the minimization problem can be found using FOC:

$$(\mathbb{E}S|A_{:T} - Z_T) + \frac{1}{\gamma_r} r^T = 0 \implies r_T = \gamma_r Z_T - \gamma_r \left(\sum_{j=0}^{T} \phi(T,j) a_j \right) \quad (12)$$

Here we used the condition of the theorem - the linearity of perceived state 4. This shows that the optimal action $r_T(Z_T, A_{:T})$ is linear in both the known state and past actions. We have to derive the next optimal rule now, it is the last action of the first player:

$$a_T(S_T, R_{:T-1}) = \arg\max_{a_T} \mathbb{E}_{Z_T|R:T-1} \left(\sum_t a_t + r_t \right) (S_T - Z_T) + \frac{1}{2\gamma_r} r_T^2 - \frac{1}{2\gamma_a} a_T^2 =$$

Omitting the terms independent of a_T:

$$\arg\max_{a_T}(S_T - \mathbb{E}(Z_T|R\colon T-1)) \cdot a_T - \frac{1}{2\gamma_a}a_T^2 + \frac{\gamma_r}{2}\left(-2\mathbb{E}(Z_T|R\colon T-1)\alpha_T^T a_T + \alpha_T^T a_T(\sum_{t=0}^{T}\alpha_t^T a_t)\right)$$

The second derivative with respect to a_T is

$$-\frac{1}{\gamma_a} + \gamma_r(\alpha_T^T)^2$$

Assuming that it is negative (if it is positive, regardless of past action, player 1 can always win infinite amount by over-betting), from the FOC we obtain:

$$a_T^* = -\left(-\frac{1}{\gamma_a} + \gamma_r(\alpha_T^T)^2\right)^{-1}\left\{(S_T - \mathbb{E}(Z_T|R\colon T-1)) - \gamma_r\mathbb{E}(Z_T|R_{:T-1})\alpha_T^T + \frac{\alpha_T^T}{2}\left(\sum_{t=0}^{T-1}\alpha_t^T a_t\right)\right\} \quad (13)$$

As we can see, $a_T(S_T, R_{:T-1})$ is also linear in both parameters, completing the base case.
Inductive step for $t_0 < T$: By the induction step we know that the optimal rules from all the previous states $t > t_0$ are linear. Now we want to show that the next optimal rule also exist for both players and is, in fact linear. Once again, we start with the rule for receiving player:

$$r_T(Z_t, A_{:t-1}, R_{:t-1}) = \arg\min_{r_t}\mathbb{E}_{Z_T|z_t}\mathbb{E}_{S_T|A:t}\left\{\left(\sum_{i=0}^{t}a_i + r_i\right)(S_T - Z_T) + \left(\sum_{i=t+1}^{T}a_i + r_i\right)(S_T - Z_T) + \right.$$
$$\left. + \sum_{i=0}^{t}\left(\frac{1}{2\gamma_r}r_i^2 - \frac{1}{2\gamma_a}a_i^2\right) + \sum_{i=t+1}^{T}\left(\frac{1}{2\gamma_r}r_i^2 - \frac{1}{2\gamma_a}a_i^2\right)\right\} \quad (14)$$

We now can use the inductive step, from which we know that all the previous optimal rules are linear in their parameters. That, in turn, allows us to use the Main lemma. So, we rewrite 14 using the result of the main lemma:

$$r_t^*(Z_t, A_{:t}, R_{:t-1}) = \arg\min_{r_t}\mathbb{E}_{Z_T|z_t}\mathbb{E}_{S_T|A:t}\left\{\left(\sum_{i=0}^{t}a + r_i\right)(S_T - Z_T)\right.$$
$$+ \sum_{i=0}^{t}\left(\frac{1}{2\gamma_r}r_i^2 - \frac{1}{2\gamma_a}a_i^2\right) + \left(\sum_{i=t+1}^{T}a(i,t+1,0) + r(i,t+1,0)\right)(S_T - Z_T)$$
$$\left. + \sum_{i=t+1}^{T}\left(\frac{1}{2\gamma_r}r(i,t+1,0)^2 - \frac{1}{2\gamma_a}a(i,t+1,0)^2\right)\right\} \quad (15)$$

Note that we use $a(i,t+1,0), r(i,t+1,0)$ and not $a(i,t,1), r(i,t,1)$, since we are trying to solve for r_t, and it essentially becomes observable. So it should be included as an active term in the formula. Let's analyze each of the four terms one by one.

1. The first term is simply equal to

$$\left(\sum_{i=0}^{t}a + r_i\right)(\mathbb{E}S_T|A_{:t} - \mathbb{E}(Z_T|Z_t)) \text{ and the derivative is } (\mathbb{E}S_T|A_{:t} - \mathbb{E}(Z_T|Z_t))$$

2. Now to the second term, the first bracket represents a linear combination of $S_t, S_{t+1}, \dots S_T$ and $Z_t, Z_{t+1}, \dots Z_T, a_0, a_1, \dots, a_t, r_0, r_1, \dots, r_t$. Using the **linearity of perceived state condition** 4

$$\mathbb{E}(S_{t+k}|A_{:t}) = \phi_a(t,k,0)a_0 + \phi_a(t,k,1)a_1 + \cdots + \phi_a(t,k,t-1)a_{t-1} + \phi_a(t,k,t)a_t$$

and the fact that $\mathbb{E}(Z_{t+k}|Z_t)$ is known and independent of past actions. We obtain that the derivative of the second part with respect to r_t is the combination of coefficients that come from the reduction of future optimal actions:

$$\left(\sum_{i=t+1}^{T} w_a^r(i,t+1,t,0) + w_r^r(i,t+1,t,0)\right)(\mathbb{E}(S_T|A_{:t}) - \mathbb{E}(Z_T|z_t))$$

3. Once we take the derivative, the third one simplifies to $\frac{1}{\gamma_r} r^t$

4. The final term is also quadratic. We will again use the fact that the expectation of perceived state is linear in $A_{:t}$ and independent of r_t. First, rewrite each sub-term:

$$\mathbb{E}\left(r(i,t+1,0)\right)^2 = \mathbb{E}\left(\sum_{j=0}^{t} w_r^r(i,t+1,j,0)r_j + \sum_{j=0}^{t} w_r^a(i,t+1,j,0)a_j + \right.$$
$$\left. + \sum_{j=t}^{i}(S_i\eta_r(i,t+1,j,0) + Z_i\nu_r(i,t+1,j,0))\right)^2$$

by thinking of it as of

$$\mathbb{E}\left(\{r(i,t+1,0) - w_r^r(i,t+1,t,0)r_t\} + w_r^r(i,t+1,t,0)r_t\right) \times$$
$$(\{r(i,t+1,0) - w_r^r(i,t+1,t,0)r_t\} + w_r^r(i,t+1,t,0)r_t)$$

which is split into 3 parts:

$$\mathbb{E}\left(w_r^r(i,t+1,t,0)r_t\right)^2 - 2\mathbb{E}\left(w_r^r(i,t+1,t,0)r_t\right)\left(r(i,t+1,0) - w_r^r(i,t+1,t,0)r_t\right) +$$
$$\mathbb{E}\left(r(i,t+1,0) - w_r^r(i,t+1,t,0)r_t\right)^2$$

As $r(i,t+1,0) - w_r^r(i,t+1,t,0)r_t$ is independent of r_t and $w_r^r(i,t+1,t,0)r_t$ is not stochastic, we simply get one quadratic, one linear and one constant term in r_t. Thus, the derivative of the elements of this sum of the form $\mathbb{E}\left(r(i,t+1,0)\right)^2$ is

$$2\left(w_r^r(i,t+1,t,0)\right)^2 r_t - 2w_r^r(i,t+1,t,0)\mathbb{E}\left(r(i,t,0) - w_r^r(i,t+1,t,0)r_t\right)$$

Similarly for $\mathbb{E}\left(a(i,t+1,0)\right)^2$ the derivative w.r.t. r_t is:

$$2\left(w_a^r(i,t+1,t,0)\right)^2 r_t - 2w_a^r(i,t+1,t,0)\mathbb{E}\left(a(i,t,0) - w_r^a(i,t+1,t,0)r_t\right)$$

And the total derivative of the fourth term is

$$\frac{1}{\gamma_r}\sum_{i=t+1}^{T}\left(\left(w_r^r(i,t+1,t,0)\right)^2 r_t - w_r^r(i,t+1,t,0)\mathbb{E}\left(r(i,t+1,0) - w_r^r(i,t+1,t,0)r_t\right)\right) -$$
$$\frac{1}{\gamma_a}\sum_{i=t+1}^{T}\left(\left(w_a^r(i,t+1,t,0)\right)^2 r_t - w_a^r(i,t+1,t,0)\mathbb{E}\left(a(i,t+1,0) - w_a^r(i,t+1,t,0)r_t\right)\right).$$

We can conclude that the second derivative is equal to

$$n_t^r := \frac{1}{\gamma_r} + \sum_{i=t+1}^{T} \left(\frac{1}{\gamma_r} [w_r^r(i, t+1, t, 0)]^2 - \frac{1}{\gamma_a} [w_a^r(i, t+1, t, 0)]^2 \right) \tag{16}$$

which we need to be strictly positive. Whether that holds or not will be analyzed in the analysis section of the paper. We also note that in this case the FOC yields a linear solution. It is written below:

$$\begin{aligned} r_t = - (n_t^r)^{-1} \cdot \Bigg\{ & \left(1 + \sum_{i=t+1}^{T} w_a^r(i, t+1, t, 0) + w_r^r(i, t+1, t, 0) \right) (\mathbb{E}(S_T | A_{:t}) - \mathbb{E}(Z_T | z_t)) \\ & - \frac{1}{\gamma_r} \sum_{i=t+1}^{T} w_r^r(i, t+1, t, 0) \left(\mathbb{E} \left(r(i, t+1, 0) - w_r^r(i, t+1, t, 0) r_t \right) \right) + \\ & \frac{1}{\gamma_a} \sum_{i=t+1}^{T} w_a^r(i, t+1, t, 0) \left(\mathbb{E} \left(a(i, t+1, 0) - w_a^r(i, t+1, t, 0) r_t \right) \right) \Bigg\} \end{aligned} \tag{17}$$

Likewise we derive the solution for the first player's action, with the only difference that now r_t is not given and we have to reduce it as well by substituting $r(i, t, 1), a(i, t, 1)$ instead of $r(i, t+1, 0), a(i, t+1, 0)$.

$$\begin{aligned} a_t^*(Z_t, A_{:t-1}, R_{:t-1}) = & \arg\min_{a_t} \mathbb{E}_{S_T|S_t} \mathbb{E}_{Z_T|R:t-1} \left(a_t + \sum_{i=0}^{t-1} a + r_i \right) (S_T - Z_T) + \\ & + \left(r(t, t, 1) + \sum_{i=t+1}^{T} a(i, t, 1) + r(i, t, 1) \right) (S_T - Z_T) + \frac{1}{2\gamma_r} r(t, t, 1)^2 + \\ & + \sum_{i=0}^{t-1} \left(\frac{1}{2\gamma_r} r_i^2 - \frac{1}{2\gamma_a} a_i^2 \right) + \sum_{i=t+1}^{T} \left(\frac{1}{2\gamma_r} r(i, t, 1)^2 - \frac{1}{2\gamma_a} a(i, t, 1)^2 \right) - \frac{1}{2\gamma_a} a_t^2 \end{aligned} \tag{18}$$

The second (define it as n_t^a) derivative is equal to

$$n_t^a := -\frac{1}{\gamma_a} + \sum_{i=t+1}^{T} \left(\frac{1}{\gamma_r} [\delta_r(i, t)]^2 - \frac{1}{\gamma_a} [\delta_a(i, t)]^2 \right) + \frac{1}{\gamma_r} [\delta_r(t, t)]^2 \tag{19}$$

And the update rule looks like

$$\begin{aligned} a_t = - (n_T^a)^{-1} \cdot \Bigg\{ & \left(1 + \delta_r(t, t) + \sum_{i=t+1}^{T} \delta_r(i, t) + \delta_a(i, t) \right) (\mathbb{E}(S_T | S_t) - \mathbb{E}(Z_T | R_{:t-1})) + \\ & - \frac{1}{\gamma_r} \sum_{i=t}^{T} \delta_r(i, t) \left(\mathbb{E} \left(r(i, t, 1) - \delta_r(i, t) a_t \right) \right) + \frac{1}{\gamma_a} \sum_{i=t+1}^{T} \delta_a(i, t) \left(\mathbb{E} \left(a(i, t, 1) - \delta_r(i, t) a_t \right) \right) \Bigg\} \end{aligned} \tag{20}$$

■

Note that in the notation of the Main Lemma, formulas 17 and 20 define the optimal extensive form decision making rules $r(t, t, 1)$ and $a(t, t, 0)$ respectively.

3.3 Game in normal players

We define a game in normal players as a game, where each player has the following belief about the other's action given the true state:

$$\mathbb{P}(a_t|S_t) = \mathcal{N}(a_t|S_t, \sigma_a^2) \qquad \mathbb{P}(r_t|Z_t) = \mathcal{N}(r_t|Z_t, \sigma_r^2) \tag{21}$$

As we will now show, this implies that the belief about the state given past actions is linear, that is

3.4 Theorem 2

Game in normal players constitutes linearity of perceived states, i.e. 4, 5 hold.

Proof:

We show only for the first player. The proof for the second player can be found in Appendix. First we prove the equation 4 for all t, in a specific case when $k = 0$:

$$\mathbb{E}(S_t|A_{:t}) = \phi_a(t,0,0)a_0 + \phi_a(t,0,1)a_1 + \cdots + \phi_a(t,0,t-1)a_{t-1} + \phi_a(t,0,t)a_t \tag{22}$$

We show that belief is distributed normally with a mean equal to a linear combination of past opponents actions, from which it immediately follows that the expectation is linear in a. That is, we want to prove that for some set of parameters $\phi(t,j)$ and variances σ_t, the following holds:

$$\mathbb{P}(S_t|A\colon t) = \mathbb{P}(S_t|a_0, a_1, \ldots a_t) = \mathcal{N}\left(S_t \middle| \sum_{j=0}^{t} \phi(t,j)a_j, \sigma_t^2\right)$$

Proof by induction: $t = 0$

$$\mathbb{P}(S_t|A_{:0}) = \mathbb{P}(S_t|a_0) = \frac{\mathbb{P}(a_0|S_0)\mathbb{P}(S_0)}{\mathbb{P}(a_0)}$$

By conjugacy of the normal distribution with its mean, for any two normally distributed values $x.y$

$$y \sim \mathcal{N}(y|x, \sigma_y^2) \text{ and } x \sim \mathcal{N}(x|\mu, \sigma_x^2) \implies \mathbb{P}(x|\hat{y}) = \mathcal{N}\left(\frac{\sigma_y^2\mu + \sigma_x^2\hat{y}}{\sigma_x^2 + \sigma_y^2}, \frac{\sigma_y^2 \cdot \sigma_x^2}{\sigma_x^2 + \sigma_y^2}\right)$$

In our case, $x \equiv S_0$, $y \equiv a_0$, $\mu = 0$, $\sigma_x = 1$, $\sigma_y = \sigma_a$ from which we obtain

$$\mathbb{P}(S_0|a_0) = \mathcal{N}\left(S_0 \middle| a_0 \cdot \frac{1}{1+\sigma_a^2}, \frac{1}{1+\sigma_a^2}\right)$$

Proof by induction: step $t > 0$

$$\mathbb{P}(S_t|A\colon 0) = \mathbb{P}(S_t|a_0, a_1, \ldots, a_{t-1}, a_t) = \frac{\mathbb{P}(a_t|S_t, a_0, a_1, \cdots, a_{t-1})\mathbb{P}(S_t|a_0, a_1, \cdots, a_{t-1})}{\mathbb{P}(a_t|a_0, a_1, \cdots, a_{t-1})}$$

Now, everything is much more simple than it looks. By the structure of the Bayesian graph 4, we know that $\mathbb{P}(a_t|S_t, a_0, a_1, \cdots, a_{t-1}) = \mathbb{P}(a_t|S_t) = \mathcal{N}(a_t|S_t, \sigma_a)$. So, if $\mathbb{P}(S_t|a_0, a_1, \cdots, a_{t-1})$ is normal, we can easily find the distribution $\mathbb{P}(S_t|a_0, a_1, \ldots, a_{t-1}, a_t)$ using the same formula for conjugates. Let's show that it's normal and find the parameters.

$$\begin{aligned}\mathbb{P}(S_t|a_0, a_1, \cdots, a_{t-1}) &= \int \mathbb{P}(S_t|S_{t-1}, a_0, a_1, \cdots, a_{t-1})\mathbb{P}(S_{t-1}|a_0, a_1, \cdots, a_{t-1})dS_{t-1} \\ &= \int \mathbb{P}(S_t|S_{t-1})\mathbb{P}(S_{t-1}|a_0, a_1, \cdots, a_{t-1})dS_{t-1}\end{aligned}$$

The latter distribution is known from last inductive step. Thus, we can summarize,

$$\mathbb{P}(S_t|a_0, a_1, \ldots, a_{t-1}, a_t) = \mathcal{N}\left(\frac{\sigma_y^2\mu + \sigma_x^2\hat{y}}{\sigma_x^2 + \sigma_y^2}, \frac{\sigma_y^2 \cdot \sigma_x^2}{\sigma_x^2 + \sigma_y^2}\right)$$

Where

$$x := S_t, \hat{y} := a_t; \quad \sigma_y = \sigma_a; \quad \mu = \mu_{t-1} = \sum_{j=0}^{t} \phi(t-1, j)a_t; \quad \sigma_x = \sigma_{t-1}$$

This beautiful result is only possible due to the conditional conjugacy and the fact that S_t contains full information about a_t (not in reality, but in the eyes of the receiving player, i.e. in biased belief). In the end, we obtain the following chain rule:

$$S_t|A_{:t} \sim \mathcal{N}(\mu_t, \sigma_t), \quad \text{where} \tag{23}$$

$$\mu_t = \frac{\sigma_a^2\mu_{t-1} + \sigma_{t-1}^2 a_t}{\sigma_{t-1}^2 + \sigma_a^2}; \qquad \sigma_t^2 = \frac{\sigma_{t-1}^2 \cdot \sigma_a^2}{\sigma_{t-1}^2 + \sigma_a^2} \qquad \sigma_0 = 1, \mu_0 = 0 \tag{24}$$

Thus, for each T, the distribution is normal, with expectation linear in $A_{:t}$
All that is left to do, is to derive $S_{t+k}|A_{:t}$ for $k > 0$. By the law of total probability:

$$\mathbb{P}(S_{t+k}|A_{:t}) = \int \mathbb{P}(S_{t+k}|S_t)\mathbb{P}(S_t|A_{:t})dS_t$$

Since $S_{t+k}|S_t \sim \mathcal{N}(S_t, k)$ we finalize by using a very well-known result for a convolution of normal distributions:

$$S_{t+k}|A_{:t} \sim \mathcal{N}(\mu_t, \sigma_t^2 + k)$$

■ We have thus shown that for any $t, k : \mathbb{E}S_{t+k}|A_{:t} = \mathbb{E}S_t|A_{:t}$ and is linear in past actions $A_{:t}$.

The proof for the perceived state Z_t of the second player is exactly the same, wrapping up the proof.

3.5 Corollary from Theorem 1 and Theorem 2:

Game in normal players has a unique equilibrium under some conditions on $\gamma_r, \gamma_a, \sigma_a, \sigma_r$. These conditions turn out to be highly non-linear and will be discussed in the analysis section.

4 Analysis

4.1 Existence of equilibrium

After deriving the solution for the game in normal beliefs, we can show conditions on $\gamma_r, \gamma_a, \sigma_a, \sigma_r$, which lead to the existence of the game. It may be deemed important that we study cases where equilibrium exists for either player taking the active position, as in cases when, say, equilibrium only exists if player i goes first and player j follows, we can not derive which player is 'better' of the two. However, studying a more general case only benefits the insights we gain, as it tells us a story about the bias induced by going first. This bias is reflected in the the graphs we are going to look at. In particular, as we will see, a slight difference in the parameters of the second player σ_r or γ_r might be much more important than that in σ_a or γ_a.

As noted previously, while recursively solving for the extensive form solution, we need the second derivative for the second decision-maker to be positive. In case it is not, there does not exist an optimal number a player needs to bet and the optimal amount to invest is one of the $\{-\infty, +\infty\}$, as we are minimizing a concave (convex) quadratic function 16. Likewise, when the second derivative for the first-decision-maker is negative, he has to bet all the money in the world 19.

In order to analyze the potential parameters that lead to equilibrium, we run numerical experiments by fixing the number of stages T and two parameters of the $\{\gamma_r, \gamma_a, \sigma_r, \sigma_a\}$. For example, we can fix $T = 3, \gamma_r = \gamma_a = .5$ and analyze the equilibrium σ_r, σ_a in a unit square. For each pair of (σ_r, σ_a) we will plot a gray dot if equilibrium exists for this pair, a black dot if the first player can win by going all-in and a white dot if the second player has this unfair advantage.

4.1.1 Practical bounds on parameters

Before making the plots, we need to discuss the reasonable range of possible values for parameters.

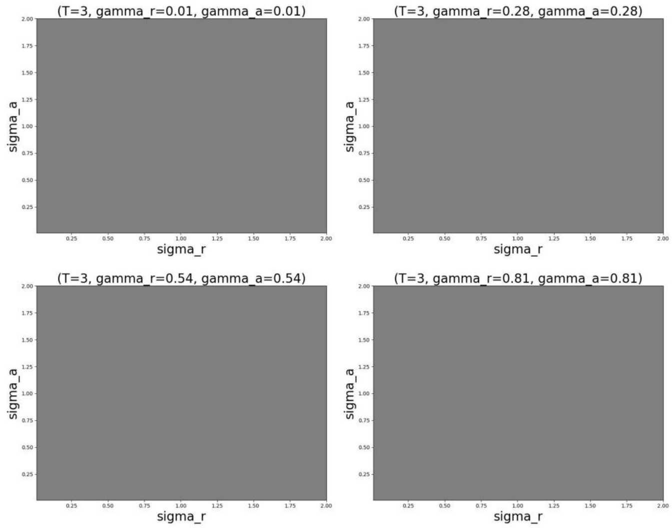
When it comes to the predictability of a player, we know the posterior distribution of the state 1. From which, trivially, the variance in the change of strength is 1. For a player to have a reasonable tightness, they would have to be predictable at the level of 2 at most, as representing an 80% change in strength every hand is unrealistic.

As for the inverse risk aversion, traditionally, betting higher then your chances to win is considered non-optimal, so $\frac{1}{\gamma} > 1$, implying that $\gamma < 1$. As we will see, higher inverse risk aversion may lead to completely unpredictable outcomes.

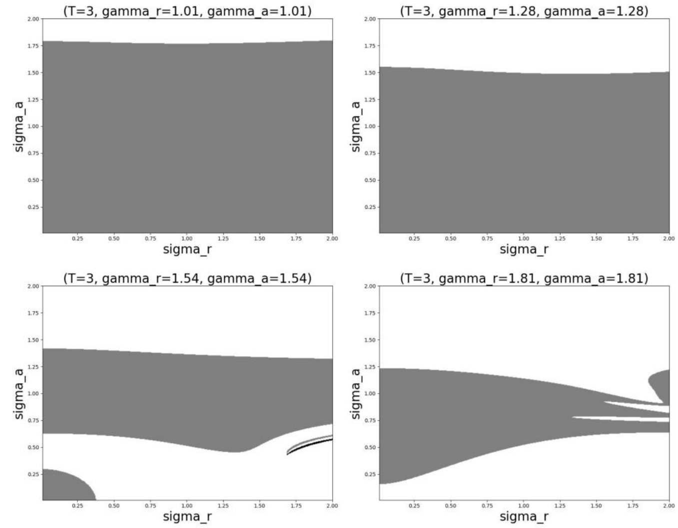
4.1.2 Analysis of equilibrium

Equal risk-aversion

Remind yourself that γ_r stands for the inverse risk-aversion. In this case, lower γ_a leads to higher risk aversion, which controls the second player's willingness to bet. What is most easily noticeable, when both risk-aversion factors are equal and low, equilibrium exists for all sensible predictability parameters 1a. On the other hand, when the risk-aversion is low ($\gamma_a = \gamma_r > 1$), the only case when we still have an equilibrium is when the first player is, in fact, predictable 1b. Once we get past a certain level of $\gamma = 1.5$ plots become chaotic. However, this is nearly not as chaotic as for extremely



(a) Varying $\gamma_a = \gamma_r$ from .01 to .81



(b) Varying $\gamma_a = \gamma_r$from 1.01 to 1.81

Figure 1: Comparison of changes when $\gamma_a = \gamma_r$.

low risk aversion ($\gamma > 10$). See **??** for beautiful chaos that erupts for longer games and extremely low risk-aversion.

Equal predictability

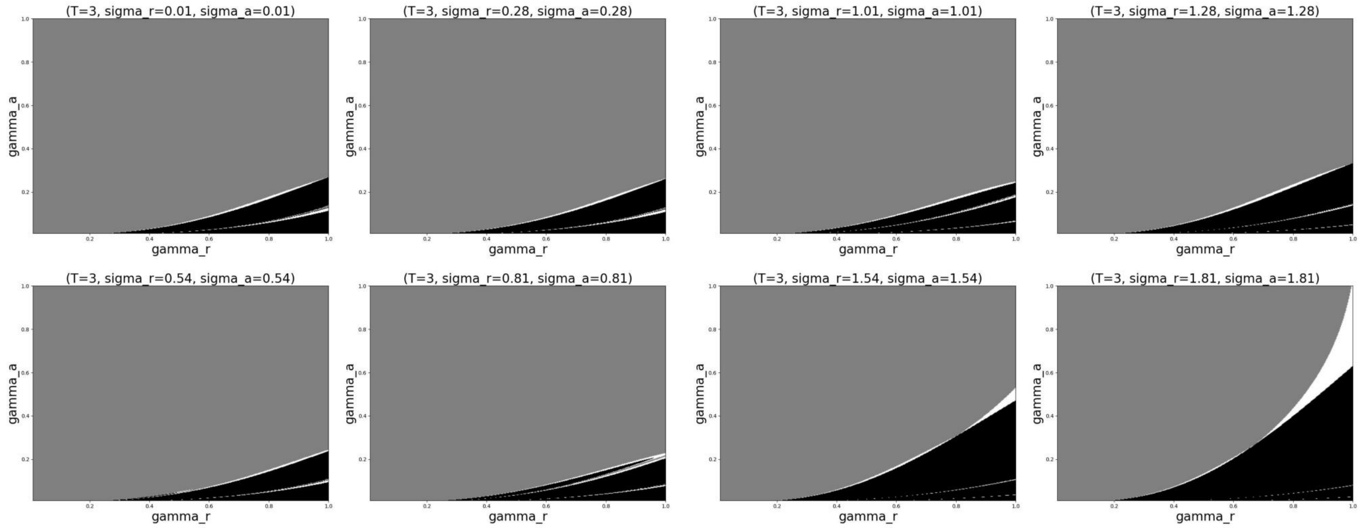
Now assume that both players are equally predictable. It turns out, that extreme values of unpredictability do not change the equilibrium outlook nearly as much as the risk-aversion 2a, 2b. There is, however a more noticeable trade-off between the risk-aversion parameters captured by this set of graphs. The recieving player can not be too loose, as it will give the first player an unfair advantage.

Risk-aversion skew

We can also take a look at how varying the γ_a, while fixing σ_a equal to 1 affects our findings. What we can see from the following graphs tells us that: Firstly, it is not clear whether γ_r or σ_r is more important for finding new equilibria, as the equilibrium zone is rather stable(see 3a). It is also reflected by the fact that increasing γ_a does not cause the same reduction in equilibrium, as increasing γ_r.

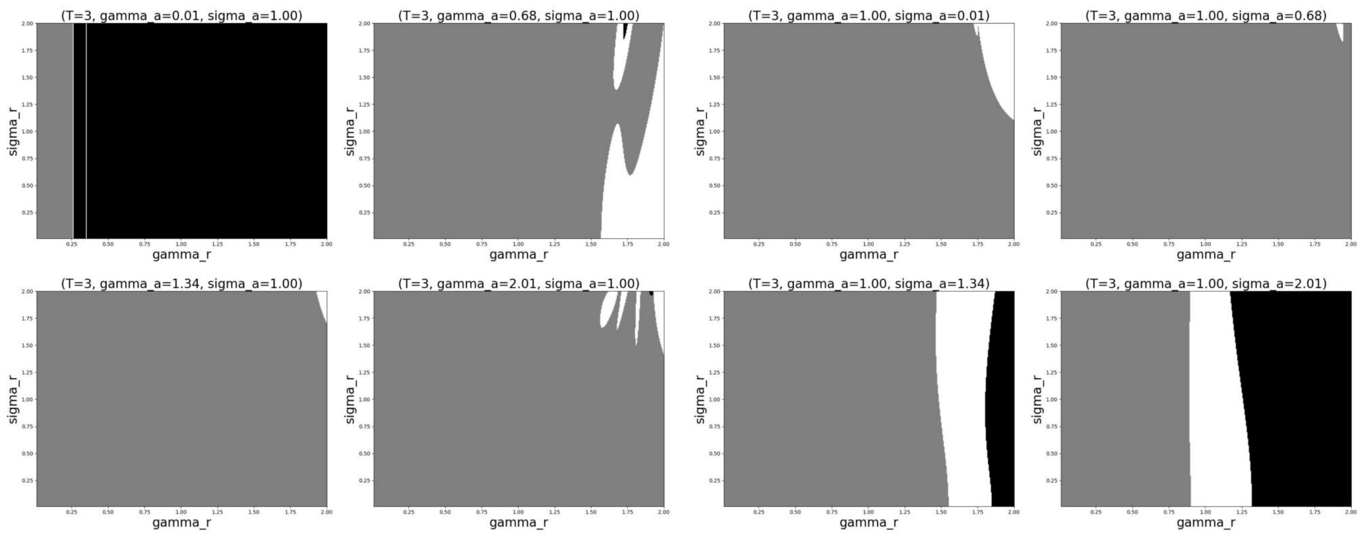
This can be interpreted in the following way: it is widely known that the second player in these types of games has a strategic advantage. He always has more information, and as such is more likely to win. When his risk aversion is low, (γ_r), it means that over-betting is much less likely to happen. On the other hand, while lower predictability of both players leads to more equilibria, being predictable does not change as much for the second player, as he has more initiative.

One the other hand, if we fix γ_a and vary σ_a, we get more easily explainable plots (3b). Being more loose first until some point gives the second player an advantage, and then turns it into a disadvantage (when $\gamma_r < .5\sigma_a$).



(a) Varying $\sigma_a = \sigma_r$ from .01 to .81 (b) Varying $\sigma_a = \sigma_r$ from 1.01 to 1.81

Figure 2: Comparison of changes when $\sigma_a = \sigma_r$.



(a) Varying γ_a from .01 to 2.01 (b) Varying σ_a from .01 to 2.01

Figure 3: Comparison of trade-offs between γ_a, σ_a.

Longer horizon games

Patterns found in this section are hard to interpret. While the normal range graphs do not change, the graphs for extreme values become more and more fractal-like 5 can be observed, with more and more games having optimal solutions, as time progresses. However, most games with low predictability of the first player now do not have a solution, as in this case, the receiving player can easier identify his strategy. *I have to say that I find these graphs truly amazing.

5 Conclusions

This paper introduces the concept of biased belief and establishes conditions for equilibrium existence in a game between two normal players, along with an analysis of players' strategic advantages.

We addressed a gap in the literature by proposing a set of assumptions that resolve the recursion problem while preserving the dynamic nature of the game with private information. We solve the problem by introducing the biased-belief and the linearity of perceived state condition, making inference feasible. We conclude by conducting an analysis related to real-world games with predictability and risk-aversion. Experiments show that model accurately predicts the advantage of the receiver under realistic parameters.

This model can be easily extended to n-player case, as under the same assumptions linearity should still hold, when the number of terms representing players' actions will increases. The only complication arises when defining the utility for a game with several players. Further analysis may include full analytic derivations for the functional form constraints on the existence of equilibria and more computationally intensive experiments, studying functional forms in long games.

6 Appendix

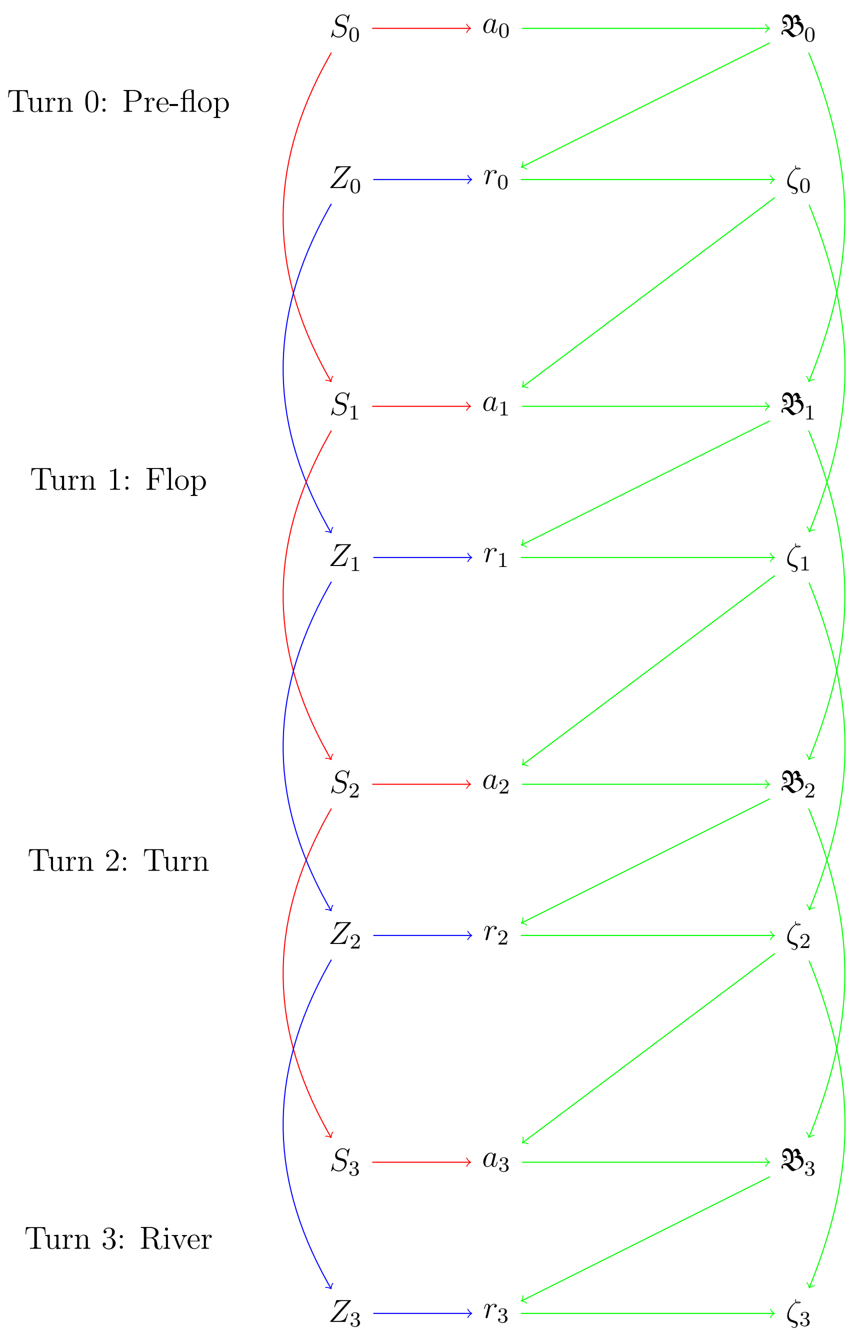


Figure 4: DAG describing the information flow in a four-period game.
Red arrows represent the change of information known to the first (active) player.
Blue arrows - for the second (receiving) player.
Green arrows highlight the flow of public information shared between two players.

6.1 Proof of the main lemma

Continue with the definitions from the formulation of the lemma: For example, in a game with T periods, the 1st player at time k, having seen only the histories $A_{:k}, R_{:k}$ might need the information about what his optimal action in round t is, given only the currently known actions. In this case, what he does is takes the original function, which we know to be linear from the condition of the lemma:

$$a_t^*(S_t, A_{:t-1}, R_{:t-1}) = \eta_t S_t + \sum_{j=0}^{k}(a_j \alpha_j^t + r_j \beta_j^t) + \sum_{j=k+1}^{t-1}(a_j \alpha_j^t + r_j \beta_j^t)$$

and reduces the last terms using the fact that the future values of a_j and r_j, $j > k$ will be chosen according to a known rule, which is linear according to the condition of the lemma.

Now, we demonstrate this reduction procedure for $k = t-1$ and then for $k = t-2$. First we start with the form of optimal decision-making rule at time t:

$$a_t^*(S_t, A_{:t-1}, R_{:t-1}) = \eta_t S_t + \sum_{j=0}^{t-2}(a_j \alpha_j^t + r_j \beta_j^t) + a_{t-1}\alpha_{t-1}^t + r_{t-1}\beta_{t-1}^t \tag{*}$$

$$r_t^*(Z_t, A_{:t}, R_{:t-1}) = \nu_t Z_t + \sum_{j=0}^{t-2}(a_j \hat{\alpha}_j^t + r_j \hat{\beta}_j^t) + a_{t-1}\hat{\alpha}_{t-1}^t + r_{t-1}\hat{\beta}_{t-1}^t + a_t \hat{\alpha}_t^t \tag{**}$$

We can immediately note that r_t^* contains a_t as it's last term. In order to correctly reduce this element, we substitute it with the optimal solution $a_t^*(S_t, A_{:t-1}, R_{:t-1})$. We do not make any assumptions here, as although the value of S_t will not be accessible to the second decision-maker, he understands the S_t as a hidden random state. This will further be reflected when he takes expectation over the belief about S_t. Technically, by substituting a_t for the optimal a_t^* we obtain:

$$r_t^*(Z_t, A_{:t-1}, R_{:t-1}, S_t) = \nu_t Z_t + \sum_{j=0}^{t-2}(a_j \hat{\alpha}_j^t + r_j \hat{\beta}_j^t) + a_{t-1}\hat{\alpha}_{t-1}^t + r_{t-1}\hat{\beta}_{t-1}^t +$$

$$+\hat{\alpha}_t^t \left(\eta_t S_t + \sum_{j=0}^{t-2}(a_j \alpha_j^t + r_j \beta_j^t) + a_{t-1}\alpha_{t-1}^t + r_{t-1}\beta_{t-1}^t \right) \tag{**'}$$

Once again, the way you would think of this last equation is the following: After making his move at time $t-1$ and before observing a_t, player 2 guesses the optimal move of the first player, as a function of a known S_t. Once the move turns, he reminds himself that he does not observe S_t and will substitute the realized value of a_t instead of a_t^* as represented by the last term. It is crucial that no expectations are taken at this point. Players simply calculate the optimal trajectories, given the reduced observable history.

Similarly, now imagine player 1 after move $t-2$. Having observed the realized values $a_0, a_1, \ldots, a_{t-2}, r_0, r_1, \ldots, r_{t-2}$, he guesses what his optimal move would be on the t-th turn, in terms of values that have been observed and future values S_{t-1}, Z_{t-1}, S_t.

$$
\begin{aligned}
a_t^*(S_t, A_{:t-1}, R_{:t-1}) &= \eta_t S_t + \sum_{j=0}^{t-2}(a_j\alpha_j^t + r_j\beta_j^t) + a_{t-1}\alpha_{t-1}^t + r_{t-1}\beta_{t-1}^t \\
&= \eta_t S_t + \sum_{j=0}^{t-2}(a_j\alpha_j^t + r_j\beta_j^t) + a_{t-1}^*\alpha_{t-1}^t + r_{t-1}^*\beta_{t-1}^t
\end{aligned}
$$

This is where the derivations get tricky, and we introduce the following 4-dimensional arrays of coefficients for actions: $w_a^a(t,k,j,b), w_r^a(t,k,j,b), w_a^r(t,k,j,b), w_r^r(t,k,j,b)$ and for states $\eta_a(t,k,j,b), \eta_r(t,k,j,b), \nu_a(t,k,j,b), \nu_r(t,k,j,b)$. They represent the coefficients for the reduced forms of optimal solutions. We start with the case of the optimal solutions themselves:

$$
r_t^* \equiv r(t,t,1) \qquad = \nu_r(t,t,t,1)Z_t + \sum_{j=0}^{t-1}\left(w_r^a(t,t,j,1)a_j + w_r^r(t,t,j,1)r_j\right) + \delta_r(t,t)a_t \tag{25}
$$

$$
a_t^* \equiv a(t,t,0) \qquad = \eta_a(t,t,t,0)S_t + \sum_{j=0}^{t-1}\left(w_a^a(t,t,j,0)a_j + w_a^r(t,t,j,0)r_j\right) \tag{26}
$$

For reduction timestamps k lower than t we define first the coefficients when a_k is not observed with $A_{:k-1}, R_{k-1}$ and then when it is:

$$
r^*(t,k,0) = \sum_{j=k}^{t}\eta_r(t,k,j,0)S_j + \sum_{j=k}^{t}\nu_r(t,k,j,0)Z_j + \sum_{j=0}^{k-1}\left(w_r^a(t,k,j,0)a_j + w_r^r(t,k,j,0)r_j\right) \tag{27}
$$

$$
a^*(t,k,0) = \sum_{j=k}^{t}\eta_a(t,k,j,0)S_j + \sum_{j=k}^{t-1}\nu_a(t,k,j,0)Z_j + \sum_{j=0}^{k-1}\left(w_a^a(t,k,j,0)a_j + w_a^r(t,k,j,0)r_j\right) \tag{28}
$$

$$
\begin{aligned}
a^*(t,k,1) = &\sum_{j=k+1}^{t}\eta_a(t,k,j,1)S_j + \sum_{j=k}^{t-1}\nu_a(t,k,j,1)Z_j + \sum_{j=0}^{k-1}\left(a_j w_a^a(t,k,j,1) + r_j w_a^r(t,k,j,1)\right) \\
&+ \delta_a(t,k)a_k
\end{aligned} \tag{29}
$$

$$
\begin{aligned}
r^*(t,k,1) = &\sum_{j=k+1}^{t}\eta_r(t,k,j,1)S_j + \sum_{j=k}^{t}\nu_r(t,k,j,1)Z_j + \sum_{j=0}^{k-1}\left(a_j w_r^a(t,k,j,1) + r_j w_r^r(t,k,j,1)\right) \\
&+ \delta_r(t,k)a_k
\end{aligned} \tag{30}
$$

Reduction rules

By the condition of the lemma, we have all optimal solutions for times $t > t_0$. In our notation it means that we have obtained the coefficient forming all of the $r^*(t,t,1), a^*(t,t,0), t > t_0$.

Step 1: Firstly, we obtain $r^*(t,t,0)$ by substituting a_t for $a^*(t,t,0)$ in $r^*(t,t,1)$. After some algebra:

$$
\begin{aligned}
\eta_r(t,t,t,0) &= \delta_r(t,t)\eta_a(t,t,t,0) \qquad & w_r^a(t,t,j,0) &= w_r^a(t,t,j,1) + \delta_r(t,t)w_a^a(t,t,j,0) \\
\nu_r(t,t,t,0) &= \nu_r(t,t,t,1) \qquad & w_r^r(t,t,j,0) &= w_r^r(t,t,j,1) + \delta_r(t,t)w_a^r(t,t,j,0)
\end{aligned} \tag{31}
$$

Step 2: Now assume we have started with $r^*(t,t,0), a^*(t,t,0)$ and reduced them down to $r^*(t,k,0), a^*(t,k,0)$ $, k \leq t$. Let's reduce them by one more step. In order to obtain $r^*(t,k-$

$1,1), a^*(t,k-1,1)$, we simply substitute the r_{k-1} for the optimal value defined by 25 for $r(k-1,k-1,1)$, obtaining:

$$
\begin{aligned}
\nu_a(t,k-1,k-1,1) &= w_a^r(t,k,k-1,0)\nu_r(k-1,k-1,k-1,1) \\
\nu_r(t,k-1,k-1,1) &= w_r^r(t,k,k-1,0)\nu_r(k-1,k-1,k-1,1) \\
\nu(t,k-1,j,1) &= \nu(t,k,j,0) \text{ only for } j > k-1 \\
\eta(t,k-1,j,1) &= \eta(t,k,j,0) \\
w_a^a(t,k-1,j,1) &= w_a^a(t,k,j,0) + w_a^r(t,k,k-1,0)w_r^a(k-1,k-1,j,1) \\
w_a^r(t,k-1,j,1) &= w_a^r(t,k,j,0) + w_a^r(t,k,k-1,0)w_r^r(k-1,k-1,j,1) \\
w_r^a(t,k-1,j,1) &= w_r^a(t,k,j,0) + w_r^r(t,k,k-1,0)w_r^a(k-1,k-1,j,1) \\
w_r^r(t,k-1,j,1) &= w_r^r(t,k,j,0) + w_r^r(t,k,k-1,0)w_r^r(k-1,k-1,j,1) \\
\delta_a(t,k-1) &= w_a^a(t,k,k-1,0) \\
\delta_r(t,k-1) &= w_r^a(t,k,k-1,0)
\end{aligned} \tag{32}
$$

Step 3: Now, to calculate the coefficients for $r^*(t,k-1,0), a^*(t,k-1,0)$ we need to reduce the a_{k-1} from $r^*(t,k-1,1), a^*(t,k-1,1)$ by substituting it for $a^*(k-1,k-1,0)$:

$$
\begin{aligned}
\eta_a(t,k-1,k-1,0) &= \delta_a(t,k-1)\eta_a(k-1,k-1,k-1,0) \\
\eta_r(t,k-1,k-1,0) &= \delta_r(t,k-1)\eta_a(k-1,k-1,k-1,0) \\
\eta(t,k-1,j,0) &= \eta(t,k-1,j,1) \text{ only for } j > k-1 \\
\nu(t,k-1,j,0) &= \nu(t,k-1,j,1) \\
w_a^a(t,k-1,j,0) &= w_a^a(t,k-1,j,1) + \delta_a(t,k-1)w_a^a(k-1,k-1,j,0) \\
w_a^r(t,k-1,j,0) &= w_a^r(t,k-1,j,1) + \delta_a(t,k-1)w_a^r(k-1,k-1,j,0) \\
w_r^a(t,k-1,j,0) &= w_r^a(t,k-1,j,1) + \delta_r(t,k-1)w_a^a(k-1,k-1,j,0) \\
w_r^r(t,k-1,j,0) &= w_r^r(t,k-1,j,1) + \delta_r(t,k-1)w_a^r(k-1,k-1,j,0)
\end{aligned} \tag{33}
$$

Using these formulas 32 - 33 and given the initial optimal solutions $r(t,t,1), a(t,t,0)$ for $t \geq t_0, k > t_0$ we can derive the new values forms $r(t,k,0), a(t,k,0), r(t,k,1), a(t,k,1)$.

Formally, this finishes the proof, as we have shown that given the linearity of optimal actions and linearity of higher order reduced optimal actions, reducing them one more step preserves the linearity. That is, the reduced form is linear. However, this does not finish the proof of theorem 1, as we have not shown that the linearity of $r(t,k,0), a(t,k,0), r(t,k,1), a(t,k,1)$ for $\tau < k < t$ implies the linearity of the optimal solutions $r(\tau,\tau,1), a(\tau,\tau,1)$.

6.2 All decision-making formulas

This section of the Appendix repeats the main formulas derived in the paper, that define the flow of information for two players and derives the last formulas needed to calculate the optimal rules.

6.2.1 Calculating the perceived expected state

We start with the formulas for updating beliefs:

$$S_t|A_{:t} \sim \mathcal{N}(\mu_t^S, (\sigma_t^S)^2) \qquad \mu_t^S = \frac{\sigma_a^2 \mu_{t-1}^S + (\sigma_{t-1}^S)^2 a_t}{(\sigma_{t-1}^S)^2 + \sigma_a^2}; \qquad (\sigma_t^S)^2 = \frac{(\sigma_{t-1}^S)^2 \cdot \sigma_a^2}{(\sigma_{t-1}^S)^2 + \sigma_a^2} \tag{34}$$

$$S_{t+k}|A_{:t} \sim \mathcal{N}(\mu_t^S, (\sigma_t^S)^2 + k)$$

$$Z_t|R_{:t-1} \sim \mathcal{N}(\mu_t^Z, (\sigma_t^Z)^2) \qquad \mu_t^Z = \frac{\sigma_r^2 \mu_{t-1}^Z + (\sigma_{t-1}^Z)^2 r_{t-1}}{(\sigma_{t-1}^Z)^2 + \sigma_r^2}; \qquad (\sigma_t^Z)^2 = \frac{(\sigma_{t-1}^Z)^2 \cdot \sigma_r^2}{(\sigma_{t-1}^Z)^2 + \sigma_r^2} \tag{35}$$

$$Z_{t+k}|R_{:t-1} \sim \mathcal{N}(\mu_t^Z, (\sigma_t^Z)^2 + k)$$

Where for both of them $\sigma_0 = 1, \mu_0 = 0$ and σ_r, σ_a are players' characteristics. This set of equation allows us to derive the coefficients for the linear form in conditional expectation:

$$\mathbb{E}(S_{t+k}|A_{:t}) = \phi_a(t,k,0)a_0 + \phi_a(t,k,1)a_1 + \cdots + \phi_a(t,k,t-1)a_{t-1} + \phi_a(t,k,t)a_t \tag{36}$$

$$\mathbb{E}(Z_{t+k}|R_{:t-1}) = \phi_r(t,k,0)r_0 + \phi_r(t,k,1)r_1 + \cdots + \phi_r(t,k,t-1)r_{t-1} \tag{37}$$

Where, trivially, $\phi(t,k,j) = \phi(t,j)$, as both means are independent of k. The boundary conditions:

$$\phi_r(0,j) = 0 \forall j \qquad \phi_a(0,j) = 1, \text{ if } j = 0 \tag{38}$$

The evolutionary formulas are:

$$\phi_r(t,t-1) = \frac{(\sigma_{t-1}^Z)^2}{(\sigma_{t-1}^Z)^2 + \sigma_r^2} \qquad \phi_a(t,t) = \frac{(\sigma_{t-1}^S)^2}{(\sigma_{t-1}^S)^2 + \sigma_a^2}, \tag{39}$$

$$\phi_r(t,j-1) = \phi_r(t-1,j-1)\frac{\sigma_r^2}{(\sigma_{t-1}^Z)^2 + \sigma_r^2} \qquad \phi_a(t,j) = \phi_a(t-1,j)\frac{\sigma_a^2}{(\sigma_{t-1}^S)^2 + \sigma_a^2}, \qquad \text{if } j < t \tag{40}$$

Using 34, 35 with 39, 40 one can derive the linear form, calculating $\phi_r(t,j), \phi_a(t,j) \ \ \forall t, \ j \leq t$.

An attentive reader will notice that this procedure leads to almost equal-weighted memory of past actions.

6.2.2 Calculating the coefficients for optimal decision-making rules

What we want to obtain are the formulas 25, 26 for all t. In order to do that, we will need to derive the 29, 30 for all $t, k : \ t \leq k$.

First we find the coefficients for the last period, i.e. we identify $r(T,T,1), a(T,T,0)$. The next equation derived in Theorem 1 constructs the optimal rule in the final period of the game:

$$r_T^* = \gamma_r Z_T - \gamma_r \left(\sum_{t=0}^{T} \phi_a(T,t)a_t \right) \tag{41}$$

$$a_T^* = -\left(-\frac{1}{\gamma_a} + \gamma_r(\phi_a(T,T))^2 \right)^{-1} \Bigg\{ (S_T - \mathbb{E}(Z_T|R_{:T-1})) - \gamma_r \mathbb{E}(Z_T|R_{:T-1})\phi_a(T,T) + \frac{\phi_a(T,T)}{2} \left(\sum_{t=0}^{T-1} \phi_a(T,t)a_t \right) \Bigg\} \tag{42}$$

Following the derivations of the theorem 1, we know that the utility is quadratic in actions.

$$U_1 = aa_t^2 + ba_t + c, \text{ where } a, b, c \text{ are independent of } a_t$$

Which implies that the optimal a_t is simply equal to $-\frac{b}{2a}$. To simplify further derivations, we denote the 2 times the quadratic coefficient (which is a function of all the previous actions) as n_t^S and the linear coefficient as m_t^S for the r-step. For the a-step we denote them n_t^Z and m_t^Z respectively. Then:

$$r(t,t,1) = -\frac{m_t^S}{n_t^S} \qquad a(t,t,0) = -\frac{m_t^Z}{n_t^Z}$$

Remind yourself that in our terminology $r(T,T,1) = r_T^* \qquad a(T,T,0) = a_T^*$, so we can derive the coefficients $w(T,T,j)$ from these two forms. They form the four boundary equations:

$$w_r^a(T,T,j,1) = -\gamma_r \phi_a(T,j) \qquad w_r^r(T,T,j,1) = 0 \tag{43}$$

$$w_a^a(T,T,j,0) = -\frac{\phi_a(T,T)}{2n_t^Z}\phi_a(T,j) \qquad w_a^r(T,T,j,0) = \frac{1+\gamma_r\phi_a(T,T)}{n_t^Z}\phi_r(T,j) \tag{44}$$

The further progression is more complicated and uses two different types of descent formulas. **The first one** tackles the problem of reduction, when one predicts the future perceived state given the current information: it fully follows the formulas 32-33, which we have previously derived.

The second formula, as derived by the theorem, calculates the coefficients for the optimal decision-making rules, given the coefficients of all reduced forms, as exactly described by equations 17, 20. The only thing left to do is to rewrite them in terms of $w(t,k,j,b)$.

First, let's rewrite them by splitting the linear part (nominator) into several parts:

$$\begin{aligned} r_t = -\left(n_t^r\right)^{-1} \cdot \Bigg\{ &\left(1 + \sum_{i=t+1}^{T} w_a^r(i,t+1,t,0) + w_r^r(i,t+1,t,0)\right)\left(\sum_{j=0}^{t}\phi(t,j)a_j - z_t\right) \\ &- \frac{1}{\gamma_r}\sum_{i=t+1}^{T} w_r^r(i,t+1,t,0)\left(\mathbb{E}\left(r(i,t+1,0) - w_r^r(i,t+1,t,0)r_t\right)\right) + \\ &\frac{1}{\gamma_a}\sum_{i=t+1}^{T} w_a^r(i,t+1,t,0)\left(\mathbb{E}\left(a(i,t+1,0) - w_a^r(i,t+1,t,0)r_t\right)\right) \Bigg\} \end{aligned} \tag{45}$$

1. Let's open the $\mathbb{E}\left(r(i,t+1,0) - w_r^r(i,t+1,t,0)r_t\right)$ for each $i > t$

$$\begin{aligned} &\mathbb{E}\left(\sum_{j=t+1}^{i}\nu_r(i,t+1,j,0)Z_j + \sum_{j=t+1}^{i}\eta_r(i,t+1,j,0)S_j\right) + \\ &\qquad \sum_{j=0}^{t-1}\left(w_r^a(i,t+1,j,0)a_j + w_r^r(i,t+1,j,0)r_j\right) + w_r^a(i,t+1,t,0)a_t = \end{aligned}$$

$$\begin{aligned} &\sum_{j=t+1}^{i}\nu_r(i,t+1,j,0)Z_t + \left(\sum_{j=t+1}^{i}\eta_r(i,t+1,j,0)\right)\left(\sum_{j=0}^{t}\phi_a(t,j)a_j\right) + \\ &\qquad \sum_{j=0}^{t-1}\left(w_r^a(i,t+1,j,0)a_j + w_r^r(i,t+1,j,0)r_j\right) + w_r^a(i,t+1,t,0)a_t = \end{aligned}$$

$$\textbf{define} \qquad \hat{\eta}_r(i,t,0) := \sum_{k=t}^{i} \eta_r(i,t,k,0) \qquad \hat{\nu}_r(i,t,0) := \sum_{k=t}^{i} \nu_r(i,t,k,0)$$

$$\textbf{and} \qquad \hat{\eta}_a(i,t,0) = \sum_{k=t}^{i} \eta_a(i,t,k,0) \qquad \hat{\nu}_a(i,t,0) := \sum_{k=t}^{i-1} \nu_a(i,t,k,0)$$

Then this term can be rewritten as:

$$= \hat{\nu}_r(i,t+1,0)z_t + \sum_{j=0}^{t} \Big(\{w_r^a(i,t+1,j,0) + \hat{\eta}_r(i,t+1,0)\phi_a(t,j)\}a_j + w_r^r(i,t+1,j,0)r_j \Big) +$$
$$+ (w_r^a(i,t+1,t,0) + \hat{\eta}_r(i,t+1,0)\phi_a(t,t))a_t$$

2. Let's open the $\mathbb{E}\left(a(i,t+1,0) - w_a^r(i,t+1,t)r_t\right)$

$$\mathbb{E}\left(\sum_{j=t+1}^{i} \eta_a(i,t+1,j,0)S_j + \sum_{j=t+1}^{i-1} \nu_a(i,t+1,j,0)Z_j + \sum_{j=0}^{t} (w_a^a(i,t+1,j,0)a_j + w_a^r(i,t+1,j,0)r_j) \right) =$$

Here, the most important step to note is that, although, we are substituting $a(i,t+1,1)$ as a function of a known state S_i, it is not known to the first player. In this case, when we take expectation, we take expectation over $S_t|A:t$, as S_t is unobservable to the second player. Similarly to the first term, we obtain:

$$= \hat{\nu}_a(i,t+1,0)z_t + \sum_{j=0}^{t} ((w_a^a(i,t+1,j,0) + \hat{\eta}_a(i,t+1,0)\phi_a(t,j))a_j + w_a^r(i,t+1,j,0)r_j) +$$
$$+ (w_a^a(i,t+1,t,0) + \hat{\eta}_a(i,t+1,0)\phi_a(t,t))a_t$$

From which we finally obtain the dependence between optimal solutions at time $r(t+1,t+1,1), a(t+1,t+1,0)$ and $r(t,t,1)$:

$$n_t^r := \frac{1}{\gamma_r} + \sum_{i=t+1}^{T} \left(\frac{1}{\gamma_r}[w_r^r(i,t+1,t,0)]^2 - \frac{1}{\gamma_a}[w_a^r(i,t+1,t,0)]^2 \right) \tag{46}$$

$$\begin{aligned}\nu_r(t,t,t,1) = -\frac{1}{n_t^r}\Bigg[&\left(1 + \sum_{i=t+1}^{T} [w_a^r(i,t+1,t,0) + w_r^r(i,t+1,t,0)]\right) \\ &- \frac{1}{\gamma_r}\sum_{i=t+1}^{T} w_r^r(i,t+1,t,0)\hat{\nu}_r(i,t+1,0) + \frac{1}{\gamma_a}\sum_{i=t+1}^{T} w_r^a(i,t+1,t,0)\hat{\nu}_a(i,t+1,0)\Bigg]\end{aligned} \tag{47}$$

$$\begin{aligned}w_r^a(t,t,j,1) = -\frac{1}{n_t^r}\Bigg\{ &\left(1 + \sum_{i=t+1}^{T} [w_a^r(i,t+1,t,0) + w_r^r(i,t+1,t,0)]\right)\phi_a(t,j) \\ &- \frac{1}{\gamma_r}\sum_{i=t+1}^{T} w_r^r(i,t+1,t,0)[w_r^a(i,t+1,j,0) + \hat{\eta}_r(i,t+1,0)\phi_a(t,j)] \\ &+ \frac{1}{\gamma_a}\sum_{i=t+1}^{T} w_r^a(i,t+1,t,0)[w_a^a(i,t+1,j,0) + \hat{\eta}_a(i,t+1,0)\phi_a(t,j)]\Bigg\}\end{aligned} \tag{48}$$

$$w_r^r(t,t,j,1) = -\frac{1}{n_t^r}\Bigg\{ -\frac{1}{\gamma_r}\sum_{i=t+1}^{T} w_r^r(i,t+1,t,0)w_r^r(i,t+1,j,0) + \frac{1}{\gamma_a}\sum_{i=t+1}^{T} w_r^a(i,t+1,t,0)w_a^r(i,t+1,j,0)\Bigg\} \quad (49)$$

These 4 formulas 46-49 finalize the reduction step for calculating optimal $r(t,t,1)$.
Finally, we use the same logic to derive the last set of formulas, defining the optimal action of the second player at time t — $a(t,t,0)$:

$$a_t = -\left(n_T^a\right)^{-1}\cdot\Bigg\{\left(1+\delta_r(t,t)+\sum_{i=t+1}^{T}\delta_r(i,t)+\delta_a(i,t)\right)\left(\mathbb{E}(S_T|S_t)-\mathbb{E}(Z_T|R_{:t-1})\right)+ -\frac{1}{\gamma_r}\sum_{i=t}^{T}\delta_r(i,t)\left(\mathbb{E}\left(r(i,t,1)-\delta_r(i,t)a_t\right)\right)+\frac{1}{\gamma_a}\sum_{i=t+1}^{T}\delta_a(i,t)\left(\mathbb{E}\left(a(i,t,1)-\delta_r(i,t)a_t\right)\right)\Bigg\}$$

Note that $\delta(t,t)$ is separated from other terms, as it has a different meaning. Using the same rules to simplify the formulas

$$\mathbb{E}S_T|S_t = S_t; \qquad \mathbb{E}Z_T|R_{:t-1} = \sum_{j=0}^{t-1}\phi_r(t,j)r_j$$

and the definition of reduced-form optimal actions, we repeat the calculations:

$$\textbf{define} \qquad \overline{\eta}_r(i,t,1) := \sum_{k=t+1}^{i}\eta_r(i,t,k,1) \qquad \overline{\nu}_r(i,t,1) := \sum_{k=t}^{i}\nu_r(i,t,k,1)$$

$$\textbf{and} \qquad \overline{\eta}_a(i,t,1) = \sum_{k=t+1}^{i}\eta_a(i,t,k,1) \qquad \overline{\nu}_a(i,t,1) := \sum_{k=t}^{i-1}\nu_a(i,t,k,1)$$

3. Let's open $\mathbb{E}\ r(i,t,1)-\delta_r(i,t)a_t$, trivially,

$$= \sum_{j=t+1}^{i}\eta_r(i,t,j,1)S_t + \left(\sum_{j=t}^{i}\nu_r(i,t,j,1)\right)\left(\sum_{j=0}^{t-1}\phi_r(t,j)r_j\right) + \sum_{j=0}^{t-1}\left(a_jw_r^a(i,t,j,1)+r_jw_r^r(i,t,j,1)\right)$$
$$= \overline{\eta}_r(i,t,1)S_t + \sum_{j=0}^{t-1}\left(a_jw_r^a(i,t,j,1)+r_j(w_r^r(i,t,j,1)+\phi(t,j)\overline{\nu}_r(t,j,1)\right)$$

4. Let's open $\mathbb{E}\ a(i,t,1)-\delta_a(i,t)a_t$, trivially,

$$= \sum_{j=t+1}^{i}\eta_a(i,t,j,1)S_t + \left(\sum_{j=t}^{i-1}\nu_a(i,t,j,1)\right)\left(\sum_{j=0}^{t-1}\phi_r(t,j)r_j\right) + \sum_{j=0}^{t-1}\left(a_jw_a^a(i,t,j,1)+r_jw_a^r(i,t,j,1)\right)$$
$$= \overline{\eta}_a(i,t,1)S_t + \sum_{j=0}^{t-1}\left(a_jw_a^a(i,t,j,1)+r_j\{w_a^r(i,t,j,1)+\phi(t,j)\overline{\nu}_a(i,t,1)\}\right)$$

This allows us to finalize the coefficients:

$$n_t^a = -\frac{1}{\gamma_a} + \sum_{i=t+1}^{T}\left(\frac{1}{\gamma_r}[\delta_r(i,t)]^2 - \frac{1}{\gamma_a}[\delta_a(i,t)]^2\right) + \frac{1}{\gamma_r}[\delta_r(t,t)]^2 \tag{50}$$

$$\begin{aligned}\eta_a(t,t,t,0) = -\frac{1}{n_t^a}\Bigg[&\left(1 + \delta_r(t,t) + \sum_{i=t+1}^{T}[\delta_r(i,t) + \delta_a(i,t)]\right) \\ &- \frac{1}{\gamma_r}\sum_{i=t}^{T}\delta_r(i,t)\overline{\eta}_r(i,t,1) + \frac{1}{\gamma_a}\sum_{i=t+1}^{T}\delta_a(i,t)\overline{\eta}_a(i,t,1)\Bigg]\end{aligned} \tag{51}$$

$$w_a^a(t,t,j,0) = -\frac{1}{n_t^a}\left[-\frac{1}{\gamma_r}\sum_{i=t}^{T}\delta_r(i,t)w_r^a(i,t,j,1) + \frac{1}{\gamma_a}\sum_{i=t+1}^{T}\delta_a(i,t)w_a^a(i,t,j,1)\right] \tag{52}$$

$$\begin{aligned}w_a^r(t,t,j,0) = -\frac{1}{n_t^a}\Bigg[&-\frac{1}{\gamma_r}\sum_{i=t}^{T}\delta_r(i,t)\big(w_r^r(i,t,j,1) + \phi_r(t,j)\overline{\nu}_r(i,t,1)\big) \\ &+ \frac{1}{\gamma_a}\sum_{i=t+1}^{T}\delta_a(i,t)\big(w_a^r(i,t,j,1) + \phi_r(t,j)\overline{\nu}_a(i,t,1)\big)\Bigg]\end{aligned} \tag{53}$$

6.3 Solution Algorithm

Algorithm 1 Solution Algorithm for Computing r^* and a^* given $\sigma_a, \sigma_r, \gamma_a, \gamma_r$

Compute all σ_t^S and σ_t^Z, by applying 34, 35

Use σ_t^S and σ_t^Z to calculate $\phi_a(t, j)$ and $\phi_r(t, j)$ for all $t, j : j < t$ following 38, 39, 40

Compute $r(T, T, 1)$ and $a(T, T, 0)$ as given by 43, 44

Compute $r(T, T, 0)$ using $r(T, T, 1)$ and $a(T, T, 0)$ as given by 31

At this point, we have obtained the boundary conditions. Next steps will only use evolutionary formulas.

for all t_0 starting from $t_0 = T - 1$ to $t_0 = 0$ **do**
    Compute $r(t_0, t_0, 1)$ as given by 46, 47, 48, 49
    for all $t > t_0$ **do**
        Compute $r(t, t_0, 1)$ and $a(t, t_0, 1)$ using $r(t_0, t_0, 1)$ as given by 32
    end for
    Compute $a(t_0, t_0, 0)$ using $r(t, t_0, 1), t \geq t_0$ and $a(t, t_0, 1), t > t_0$ as given by 50, 51, 52, 53.
    for all $t > t_0$ **do**
        Compute $r(t, t_0, 0)$ and $a(t, t_0, 0)$ as given by 33
    end for
    Finally, obtain $r(t_0, t_0, 0)$ as given by 31
end for

6.4 Figures

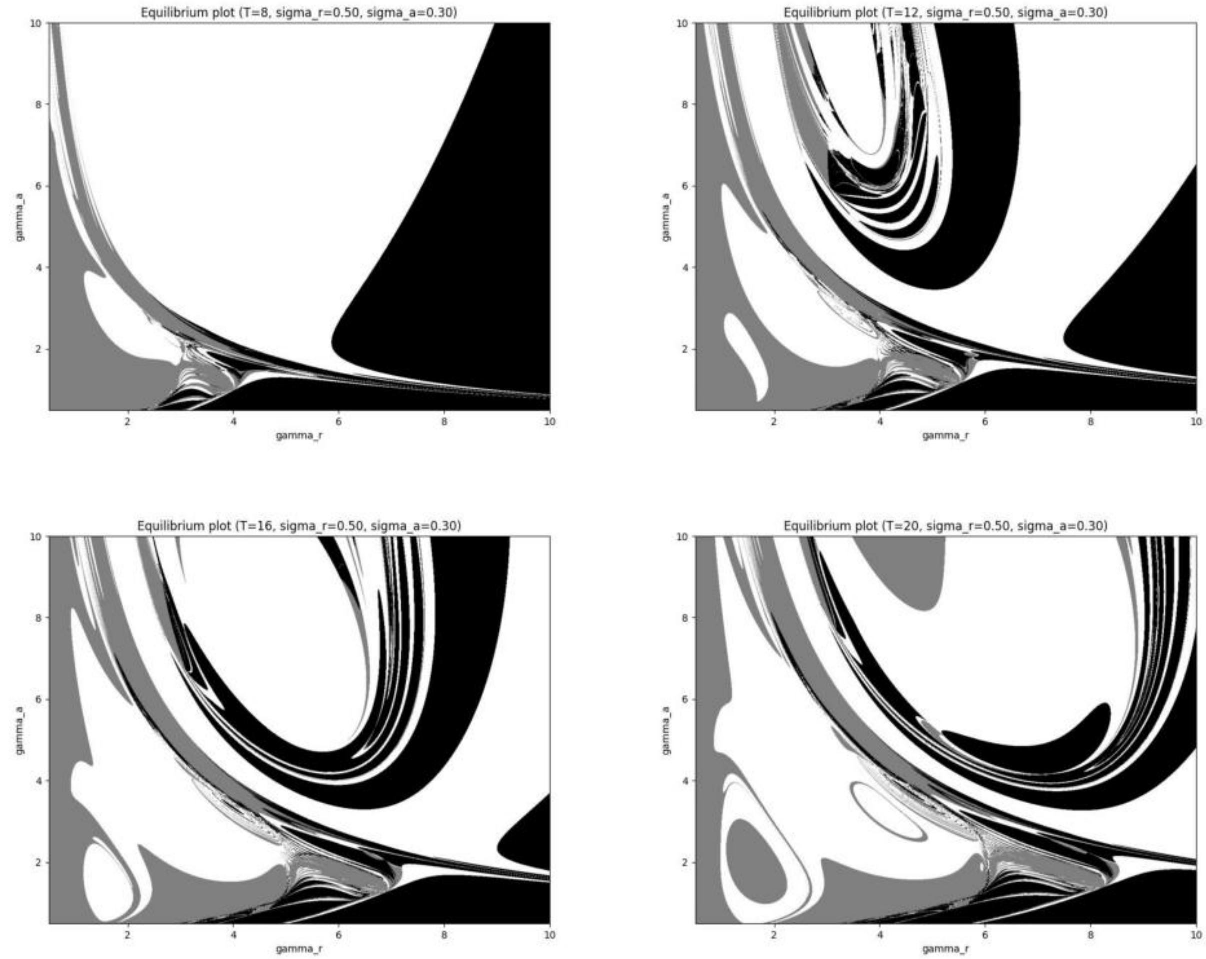


Figure 5: Different plots for regions of equilibrium existence.
γ_a is on the y-axis; γ_a is on the x-axis, both changing from 0 to 10

References

Al-Nowaihi, A., & Dhami, S. (2015). Evidential equilibria: Heuristics and biases in static games of complete information. *Games*, *6*(4), 637–676.

Billings, D., Burch, N., Davidson, A., Holte, R., Schaeffer, J., Schauenberg, T., & Szafron, D. (2003). Approximating game-theoretic optimal strategies for full-scale poker. *IJCAI*, *3*, 661.

Bordon, P., & Fu, C. (2015). College-major choice to college-then-major choice. *The Review of economic studies*, *82*(4), 1247–1288.

Bowling, M., Burch, N., Johanson, M., & Tammelin, O. (2015). Heads-up limit hold'em poker is solved. *Science*, *347*(6218), 145–149.

Caragiannis, I., Kurokawa, D., & Procaccia, A. (2014). Biased games. *Proceedings of the AAAI Conference on Artificial Intelligence*, *28*(1).

Cole, H. L., & Kocherlakota, N. (2001). Dynamic games with hidden actions and hidden states. *Journal of Economic Theory*, *98*(1), 114–126.

Doepke, M., & Townsend, R. M. (2006). Dynamic mechanism design with hidden income and hidden actions. *Journal of Economic Theory*, *126*(1), 235–285.

Fershtman, C., & Pakes, A. (2012). Dynamic games with asymmetric information: A framework for empirical work. *The Quarterly Journal of Economics*, *127*(4), 1611–1661.

Haan, M. A., & Hauck, D. (2023). Games with possibly naive present-biased players. *Theory and Decision*, *95*(2), 173–203.

Heller, Y., & Winter, E. (2020). Biased-belief equilibrium. *American Economic Journal: Microeconomics*, *12*(2), 1–40.

Kamenica, E., & Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, *101*(6), 2590–2615.

Reif, J. H. (1979). Universal games of incomplete information. *Proceedings of the eleventh annual ACM symposium on Theory of computing*, 288–308.

Yamamoto, Y. (2019). Stochastic games with hidden states. *Theoretical Economics*, *14*(3), 1115–1167.