Preventive Wars

KLAUS ABBINK, LU DONG, LINGBO HUANG*

Abstract

The rise of a new power may lead the dominant power to seek a preventive war. We study this scenario in an experimental two-stage bargaining game. In each stage, the rising power makes a bargaining offer and the declining power must choose whether to accept it or fight. Between the two stages, the winning probability shifts towards the rising power. We find fewer preventive wars when the power shift is smaller and when the rising state has the commitment power. Communication and repeated interaction decrease the likelihood of preventive wars. High fighting costs almost eliminate such wars when the rising power's first-stage offer is sufficiently large.

Keywords

power shift, commitment, bargaining, conflict, communication

JEL codes

C72, C91, F51, N40

* Abbink: Monash Business School, Clayton, VIC, Australia, klaus.abbink@monash.edu. Dong: Economics Experimental Laboratory, Nanjing Audit University, Nanjing 211815, China, lu.dong@outlook.com. Huang: Economics Experimental Laboratory, Nanjing Audit University, Nanjing 211815, China, lingbo.huang@outlook.com. Financial support from the Australian Research Council (DP1411900) and National Natural Science Foundation of China (Grant No. 71903092) is gratefully acknowledged.

1. Introduction

When the Cold War ended, it looked like the world had become unipolar, with the United States as the only superpower left. The subsequent rise of China took place predominantly in the economic arena, while the Chinese government still followed Deng Xiaoping's doctrine of keeping a low profile in the geopolitical sphere. This changed when Xi Jinping took over as the Chairman of the Chinese Communist Party in 2012. China is now well on course towards parity with the U.S. in economic strength, and has started flexing its geostrategic muscles by investing heavily in its military. The election of Donald Trump as the U.S. president in 2016 did not improve Sino-American relations. Though Trump's rhetoric and actions mainly revolved around trade issues rather than geopolitics, the U.S. attitude towards China became increasingly icy during Trump's term in the office, and the Biden administration does not show much intent to change that. Ferguson (2019) argues that a new Cold War has already begun, though he does not expect it to become hot in the near future.

The rise of a new power challenging the incumbent hegemon's position inevitably leads to tensions. Allison (2017) coined the term 'Thucydides Trap' after the ancient Greek historian who wrote on the Peloponnesian War (431−404 BCE): "It was the rise of Athens and the fear that this instilled in Sparta that made war inevitable".[1] Allison analyzed 16 modern historical events in which a rising power challenged a ruling power and found that war was the outcome in 12 of those instances. Most famously, the rise of the German empire in the late 19th century threatened a still dominant, but slowly declining British empire. It can be argued that this shift in power was the underlying cause of World War I, which was eventually triggered by a minor event, the assassination of Archduke Franz Ferdinand by a Serbian terrorist.[2]

Historical examples can only ever go so far; for every aspect that parallels the historical blueprint there is another one that deviates from it. Concurring with this view, Allison is adamant that war between the U.S. and China is *not* inevitable. Nonetheless, there are many fault lines along which conflict can break out. Allison (2017) lays out four realistic scenarios of how a minor incident, an accident or a targeted provocation, can escalate into full-fledged war. In addition, both sides have developed significant capabilities for cyber warfare. A comprehensive shutdown of the other side's communication networks can be

[1] It has been doubted whether this translation and interpretation of the original source is entirely accurate. We view the term Thucydides Trap as a catchy, if somewhat hard-to-pronounce expression for a more general concept of an interplay between a rising and a ruling power.

[2] Germany, being a rising power, feared that another rising power, Russia, could gain in strength and become a threat on Germany's eastern border (the two empires were neighbors at the time). This perceived risk prompted strategists in the German government to consider an attack on Russia as a preventive war, as discussed in detail by Clark (2014). Other than wars between major states, the outbreak of many other conflicts of limited scale has been attributed to the logic of preventive wars. For example, in the Gulf War, the U.S. attacked Iraq because it feared that Iraq would gain more monopoly power over oil supplies and thus more bargaining leverage due to the U.S. reliance on oil.

as devastating as a nuclear strike, but the psychological threshold for actually attempting such an attack might be lower than that for unleashing a widely feared nuclear exchange.

Shifts in relative power are also dangerous because they create an incentive for the *ruling* power to initiate conflict, to destroy the rising power before it can become too big. Such type of conflict, termed *preventive war*, is theoretically inevitable (Fearon 1995; Powell 1999) because the rising power faces a commitment problem stemming from bargaining divisions spread over multiple stages. Once the rising power has become superior, it has every reason to renege on any promise about future bargaining agreements. Foreseeing this adverse outcome, the declining power must launch an early attack—preventive war—while it is still stronger than the rising power. A sufficiently large shift in power is at the core of the commitment problem, and preventive wars are launched precisely to forestall a shift in the balance of power in favor of the opponent and avoid losing any bargaining advantages in the future.

This outcome, however, is inefficient because there would be bargaining solutions in which both powers would be better off, as dictated by the Coase theorem. To achieve them, the rising power would have to commit to a bargaining division at the second stage even though, once this stage is reached, it could improve its outcome by reneging on the promised division. If it could commit, the declining power could faithfully abstain from a preventive war and achieve an outcome that is overall more beneficial for it than the risky outcome of a preventive war. However, since this is not possible, a costly war is inevitable. This argument holds even in the absence of any information asymmetries, which is the standard explanation for the failure of bargaining and the outbreak of inefficient conflicts (Brito and Intriligator (1985); see reviews by Ausubel, Cramton, and Deneckere (2002) and Sanchez-Pages (2012)).

In the present study, we employ a highly parsimonious model to capture the essence of the logic of preventive wars, building on the seminal works by Fearon (1995) and Powell (1999). In this simple scenario, there are two players, a rising and a declining player, who interact for at most two bargaining stages. In each stage, the rising player makes a bargaining offer to the declining player who then decides whether to accept the offer or go to war. However, the war is costly and inefficient. If war breaks out in the first stage, a lottery strongly in favor of the declining player will determine who wins the total pie (i.e., all payoffs across two stages) and the game ends. If war occurs in the second stage, however, a lottery strongly in favor of the rising player will determine who attains all the gains. Thus, the exogenously varied probability of winning captures the shift in power over the two stages. Finally, if peace is kept in both stages, both players simply receive their respective bargaining payoffs.

It is worth noting that our simple model allows only the rising player to make the bargaining offer and only the declining player has the right to declare a war. The deprivation of the rising player's right to declare war emphasizes that the occurrence of war does not rely on the fear of being attacked but rather on the unease with "the peace it will have to accept after the rival has grown stronger" (Fearon 1995, p.406). The reason for isolating the fear of being attacked is that it is typically associated with preemptive war, which is often confused with preventive war. A preemptive war is launched in anticipation of an imminent attack by the opponent whose intentions are unknown (Schelling 1960; Jervis 1976; Abbink, Dong, and Huang 2021). By contrast, the aim of a preventive war is stopping an opponent from developing the capability to attack. In his just war theory, Walzer (1977) justifies a preemptive war as a form of self-defense, while no such argument can be made for a preventive war, as the mere capability to attack does not imply any intention to do so. Thus, our modeling choice is intended to differentiate the logic of preventive war from that of preemptive war as clearly as possible.

In the theoretical model we adopt here, war is inevitable. In practice, agents may find a way out of the Thucydides Trap. To test whether, and if so how, the trap can be avoided we carry the game theoretic model into the experimental laboratory, as it is an ideal environment for testing the behavioral validity of game theoretic predictions. Countless experiments on social dilemma games, ultimatum games, trust games and others have shown that even clear game theoretic predictions often break down when they are at odds with human instincts. Often, experimental subjects choose to cooperate where game theory says they should not, and achieve much more efficient outcomes, and this is particularly the case in public good and trust games. Sometimes, however, we observe a tendency toward much greater aggression than theory would imply. This is true for many rent-seeking games (Dechenaux, Kovenock, and Sheremeta 2015).

In different treatments, we study the role of shift in power and commitment capability in reducing the likelihood of preventive wars. First, we compare a treatment in which theory predicts war (*PreventiveWar*) with a treatment which is similar, but the parameters are set up so that war should not happen in theory (*NoWar*). Second, we examine the commitment problem more directly, allowing the rising player to commit to the bargaining offer at the second stage (*RP-Commit*). In theory, this scenario should lead to peace. In the experiment, we indeed observe a lower conflict rate, but the separation is far less sharp than the theory predicts.

We then investigate the role of two mechanisms that are often invoked to resolve conflict, one of which is direct communication between the two parties who may attempt to resolve conflict via diplomatic means. To test the role of communication, in addition to the baseline treatment, we also allow the two players to send free-form messages to each other before the game starts in each round (the *Chat* treatment). Another mechanism involves repeated interaction which may encourage the players to avoid conflict by building reputation via their bargaining and fighting decisions. To test its role, we turn to a partner matching protocol and fix the players' roles for each pair throughout the experiment (the *Repeat* treatment). Compared to the baseline treatments, both conflict resolution mechanisms help reduce the frequency of preventive wars, thus instilling more confidence in Allison's (2017) claim that the Thucydides Trap is indeed not inevitable.

Under these conditions, we find that treatment differences are rather small, especially compared with the clearly demarcated theoretical predictions. We therefore conduct two further treatments in which we make fighting more or less attractive than in the previous treatments. We find that, if the costs of fighting are very high, preventive wars can almost be eliminated if the rising player offers as much as possible in the first stage. Thus, this treatment shines a light of hope that war is indeed not inevitable especially between major states with devastating military powers such as nuclear weapons. Low fighting costs, however, increase the likelihood of conflict.

Our study marks one of the first attempts to thoroughly examine the logic of preventive wars through an experimental game. The work that most closely aligns with our research was Tingley (2011) who conducted an experiment on international relations. Using an indefinite repeated experimental game, Tingley (2011) tested whether a larger discount factor leads to more preventive wars and found evidence supporting this outcome. The approach adopted in our study is substantially different as, instead of varying the discount factor, we manipulate a host of other parameters (i.e. probability of winning and fighting cost) and situational conditions (i.e. commitment capability, communication and repeated interaction). This strategy allows us to test more directly and comprehensively the underlying logic of and the remedial mechanisms to preventive wars.

By contrast, the literature on dynamic power shifts and associated commitment problems in conflict has been almost exclusively theoretical. The formal models of preventive wars were originally analyzed by Fearon (1995) and Powell (1999). Using infinitely repeated games, they demonstrate that states' inability to commit to bargaining divisions over multiple stages combined with a large discontinuous change in the distribution of military power can lead to war when both parties have complete information. Powell (2004) shows that lack of commitment power and large shifts in power are two common threads across several seemingly diverse studies, the topics of which were originally thought to be unrelated to preventive wars. These include Fearon's (1996) analysis of prolonged civil wars, Acemoglu and Robinson's (2000, 2001) model of costly coups and political transitions, and de Figueiredo's (2002) account of limited term of office and inefficient policy insulation. Powell (2006) further includes Fearon's (1996) model of bargaining over issues that affect future bargaining power as another case where the same mechanism is at work even when the shifting distribution of power is endogenous to past concessions in bargaining. Baliga and Sjöström (2013) provide a theoretical review of several formal models of bargaining and conflict, including ones about commitment problems. In a more recent work, using a markedly different setup, Baliga and Sjöström (2020) study a type of two-sided commitment in which both parties can commit to challenge the status quo and the party that unilaterally commits will gain a first-mover advantage in the ensuing conflict. In their model, a power shift can also lead to greater conflict when the rising state is initially weak; however, when the rising power becomes sufficiently strong, a further power shift would actually make conflict less likely.[3]

[3] The relationship between outbreak of war and commitment was also discussed by Beviá and Corchón (2010), Chadefaux (2011), Wolford, Reiter, and Carrubba (2011) and Krainin (2017).

Our investigation is also related to the experimental studies about conflict resolution mechanisms such as side payments in Ultimatum bargaining games (Coursey and Stanley 1988; Kimbrough and Sheremeta 2013, 2014; Kimbrough, Sheremeta, and Shields 2014; Kimbrough et al. 2015) and in Nash demand games (Herbst, Konrad, and Morath 2017), which may be followed by a conflict stage. In most of these studies, side payments or other bargaining arrangements are binding, thereby implicitly eliminating the commitment problems. As an exception, Kimbrough and Sheremeta (2013) compare the effect of binding and non-binding side payments on reducing conflict. They find that non-binding side payments do not help reduce conflict, whereas binding ones do. Kimbrough et al. (2015) also study commitment problems, albeit in a different setup. In their game, a coin flip serves as a conflict resolution mechanism, the outcome of which, if not accepted by both players, would lead to a conflict stage where players exert costly effort to win a prize. They find that this non-binding coin flip may nevertheless reduce conflict rates. The commitment problem we study is, however, fundamentally different because it pertains to the inability to commit to bargaining allocations across *multiple* stages, which alongside the power shift may create the incentive to strike early by the untrusty opponent. In this scenario, conflict arises in anticipation of the commitment problem and is also theoretically inevitable. By contrast, in the setup adopted by Kimbrough et al. (2015), conflict follows immediately from a breakdown of commitment and does not convey the sense of inevitability.

2. Theoretical Framework

We model the preventive war scenario as a two-stage two-player bargaining game. In each stage, two players bargain over a pie worth one unit. In Stage 1, a rising player (henceforth referred to as RP) makes an offer to a declining player (henceforth referred to as DP), $x_1^{DP} \in [0,1]$. DP then observes x_1^{DP} and responds with either "accept" or "fight". If DP chooses "accept", the game moves on to Stage 2. If DP chooses "fight", there will be no Stage 2. In Stage 2, the two players bargain over another pie worth one unit. RP makes an offer $x_2^{DP} \in [0,1]$. DP then observes x_2^{DP} and again responds with either "accept" or "fight". If DP chooses "accept" again, the game ends with both players receiving their respective share of the total bargaining pie according to RP's allocation decisions in the two stages. If "fight" occurs in either stage, any agreement about offers in Stage 1 will be nullified. In particular, suppose "fight" occurs in Stage 2, the offer in Stage 1 will not count and players' payoffs will only depend on the outcome of the fight. If the two players fight in either stage, they will compete for a prize worth two units, which is the total payoff across both stages.[4]

[4] An alternative modeling choice is that the fight would only affect the division of the current and future total bargaining pie, but not previous pie divisions. We favor the current setup largely because it allows for a simpler model that can still reproduce the preventive war logic. Otherwise, the preventive war scenario could only arise when the future bargaining pie or the shadow of the future (as embodied in an infinitely repeated game with a discount) is large enough.

To model the inefficiency of war as opposed to peace, each player also has to incur an irrevocable cost of fighting, C, which we set to $C < 1$ to avoid trivial outcomes. Each player's probability of winning the prize depends on their relative strength in the current stage. Importantly, compared to Stage 1, RP becomes stronger while DP becomes weaker in Stage 2. We model the relative strength directly as DP's probability of winning in a stage, which is P_1^{DP} in Stage 1 and P_2^{DP} in Stage 2 whereby $0 < P_2^{DP} < 0.5 < P_1^{DP} < 1$. Correspondingly, RP's probabilities of winning in the two stages satisfy $0 < P_1^{RP} < 0.5 < P_2^{RP} < 1$, $P_1^{RP} + P_1^{DP} = 1$ and $P_2^{RP} + P_2^{DP} = 1$. Figure 1 displays the timing, decisions and payoffs of the game.

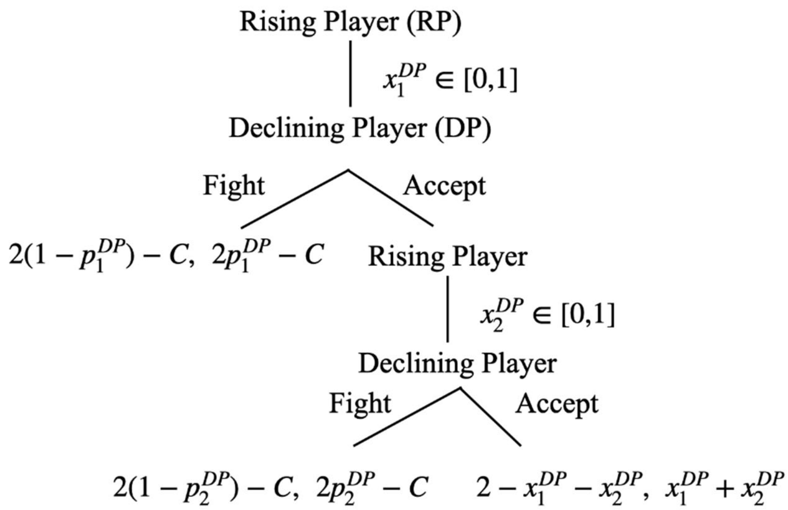

Figure 1: The game tree

We apply the subgame perfect equilibrium (SPE) concept to solve the game. We need to distinguish two fundamentally different parameter constellations that lead to distinct SPE outcomes. Whether the SPE prediction is war or peace depends on whether or not it is possible to make an offer in Stage 1 that keeps the declining player satisfied.

Proposition. If $2P_1^{DP} - C \leq 1$, then every SPE involves peace. If $2P_1^{DP} - C > 1$, then every SPE involves war.

Proof. Let us first examine the latter case, in which war is strategically inevitable. We solve the game by backward induction. The declining player will fight in Stage 2 if

$$x_1^{DP*} + x_2^{DP} < 2P_2^{DP} - C \qquad (1)$$

If the first-stage offer has been sufficiently high, the declining player has no interest in fighting. Hence, if $x_1^{DP*} \geq 2P_2^{DP} - C$, the declining player will accept the offer even if he is offered zero payoff in Stage 2. Note that this amount is less than one. The declining player foresees that he (or she, but for brevity we will use 'he' and 'she' interchangeably without inferring any particular gender) will not be offered a positive amount in Stage 2 if the first-stage offer matches his expected payoff from fighting in the second stage. His total payoff will thus be no greater than 1. His payoff from fighting in Stage 1 is $2P_1^{DP} - C > 1$, whereas his payoff from accepting is $x_1^{DP} + 0 \leq 1$. The problem here is that, even if the rising player offers everything he can offer in Stage 1 the declining player cannot be satisfied. Thus, in the absence of an enforcement mechanism, no commitment the rising player makes in Stage 1 would be credible. It is therefore more advantageous for the declining player to seek war in Stage 1.

If $2P_1^{DP} - C \leq 1$, fight at any stage is not an SPE outcome. As in the previous scenario, the declining player will fight if $x_1^{DP*} + x_2^{DP} < 2P_2^{DP} - C$. In an SPE, Stage 2 is only reached if the rising player's offer in Stage 1 was sufficiently high to deter the declining player from fighting at this stage. Thus $x_1^{DP*} \geq 2P_1^{DP} - C$. Since $2P_1^{DP} - C \leq 1$ this is now a feasible offer. Consequently, this condition together with (1) implies that the declining player will not fight even if offered zero payoff in the second stage. Hence, in an SPE, the rising player will offer zero in Stage 2.

It remains for us to show that it is in the rising player's interest to make a first-stage offer of at least $x_1^{DP*} = 2P_1^{DP} - C$. Recall that $P_1^{RP} < 0.5$. If the rising player offers less than that, the declining player will fight in Stage 1, and the rising player's payoff will be $2P_1^{RP} - C < 1$. If the rising player offers at least $x_1^{DP*} = 2P_1^{DP} - C$, the declining player will refrain from fighting and the game proceeds to Stage 2. As in this stage alone, the rising player receives a payoff of 1 (because $x_2^{DP*} = 0$), it is in the rising player's interest to move the game further to Stage 2. The rising player offers the smallest amount that ensures game continuation, hence $x_1^{DP*} = 2P_1^{DP} - C$. ■

To demonstrate that commitment is the key issue in preventive wars, we retain the condition $2P_1^{DP} - C > 1$ and assume that RP has the commitment power. Specifically, in Stage 1, in addition to making the actual offer x_1^{DP} for this stage, RP also announces the plan of offer $\widehat{x_2^{DP}} \in [0,1]$ in Stage 2, which is binding, i.e. $x_2^{DP} = \widehat{x_2^{DP}}$. Knowing that DP's war payoff in Stage 1 is $2P_1^{DP} - C$, RP's offers must satisfy $x_1^{DP} + x_2^{DP} \geq 2P_1^{DP} - C$. That is, any combination of x_1^{DP} and x_2^{DP} which is no less than DP's war payoff in Stage 1 will satisfy DP (who will choose "accept" in both stages) and will thus preclude the costly war. In the SPE, RP's payoff is $2(1 - P_1^{DP}) + C$ and DP's payoff is $2P_1^{DP} - C$. By allowing DP to commit to an offer in Stage 2, the game is essentially reduced to a single-stage game. As long as RP offers DP at least DP's war payoff, fighting will not occur and the commitment problem is irrelevant.

While lack of commitment should generally be considered as a real-world feature that is difficult to resolve, the proposed scenario is still reminiscent of some real-world situations in which at least partial commitment is possible. For example, RP's credibility of committing to future offers provides an analog to situations in which a nation's ability to renege on an agreement may be constrained by reputational concerns or anticipation of domestic or international backlash against repudiating the deal.

3. Experimental Design and Hypotheses

3.1. Treatments

3.1.1. The PreventiveWar treatment (baseline)

The basic experimental design follows the model setup and commences with the baseline *PreventiveWar* treatment, whereby each session consists of 20 rounds. In each round, two players are randomly matched and randomly assigned to the role of RP and DP (which are referred to as Person A and Person B respectively in the experiment). Each player starts with an endowment of 5 tokens per round, which can be used to cover any potential loss from the game. Each round has two stages. In Stage 1, RP makes an offer out of a pie worth 10 tokens as well as announces the plan for his offer in Stage 2 (fraction of another pie worth 10 tokens). Upon considering RP's offer and plan, DP decides whether to "accept" or "reject" in this stage.[5] Both players' decisions and plans made in Stage 1 will be revealed after this stage is completed. The players are informed in the instructions as well as on their computer screens that their plans are not binding, and therefore, they will be free to change their decisions whenever the game proceeds to Stage 2.

If fighting occurs (i.e. DP chooses "reject") in Stage 1, RP's offer will not be implemented and the two players will compete for a prize worth 20 tokens, whereby participation in conflict incurs a cost of 5 tokens per person. The computer will decide who receives the prize according to their probability of winning in Stage 1: 20% for RP and 80% DP ($P_1^{DP} = 0.8$). The round will end without proceeding to Stage 2.

When the game proceeds to Stage 2 (i.e. DP chooses "accept" in Stage 1), RP makes an offer out of another pie worth 10 tokens. DP observes RP's offer and decides whether to "accept" or "reject" it. If DP chooses "reject", this will nullify RP's offers in *both* stages, which will thus not be implemented and the two players will compete for a total prize worth 20 tokens, each incurring a cost of 5 tokens. The computer will decide who receives the prize according to their probability of winning in Stage 2, which is set to 70% for RP and 30% for DP ($P_2^{DP} = 0.3$). This scenario represents a large power shift favoring RP in Stage 2. If, however, DP chooses "accept", both players will receive their respective payoffs corresponding to RP's offers in the two stages.

According to the theoretical analysis in the previous section, DP will choose "reject" in Stage 1 regardless of RP's offer and plan. Thus, DP's expected equilibrium payoff in a round is 20 * 0.8 – 5 = 11 tokens and RP's expected equilibrium payoff is 20 * (1 – 0.8) –

[5] In addition to deciding whether to accept or reject the first-stage offer, DP is also asked to make a non-binding announcement of her plan for Stage 2. If she chooses "accept", she announces whether she would be willing to "accept" or "reject" each of the possible RP's offers in Stage 2. If DP chooses "reject" in Stage 1, such an announcement becomes superfluous.

5 = –1 token. Hence, the baseline PreventiveWar treatment illustrates the scenario of preventive wars.

3.1.2. The NoWar treatment

To examine the influence of power shifts (which cause the commitment problem), we implemented the *NoWar* treatment which differs from the PreventiveWar treatment only in terms of RP's winning probability in Stage 1, which is increased from 20% to 30%, thereby reducing DP's winning probability to 70%. This seemingly small parameter change eliminates the commitment problem because DP's expected war payoff in Stage 1 is now 20 * 0.7 – 5 = 9 tokens, which is less than the maximum possible offer (10 tokens) RP can make in Stage 1. Therefore, in the SPE, RP's offer is exactly 9 tokens in Stage 1 and 0 token in Stage 2. In equilibrium, DP chooses to accept the offer in both stages and receives 9 tokens while RP's payoff is 11 tokens. Figure 2 displays the game trees pertaining to the PreventiveWar and NoWar treatments with their respective parameterization.

Owing to a slight difference in RP's winning probability in Stage 1 (as a sort of "discontinuity" design), the PreventiveWar and NoWar treatments produce sharply contrasting theoretical predictions. However, the comparison between these two treatments provides a rather conservative test of the logic behind preventive wars since the change in incentive seems subtle. In Section 5.2, we will introduce two additional treatments with a much stronger separation and additionally test whether the likelihood of preventive wars is sensitive to the fighting cost.

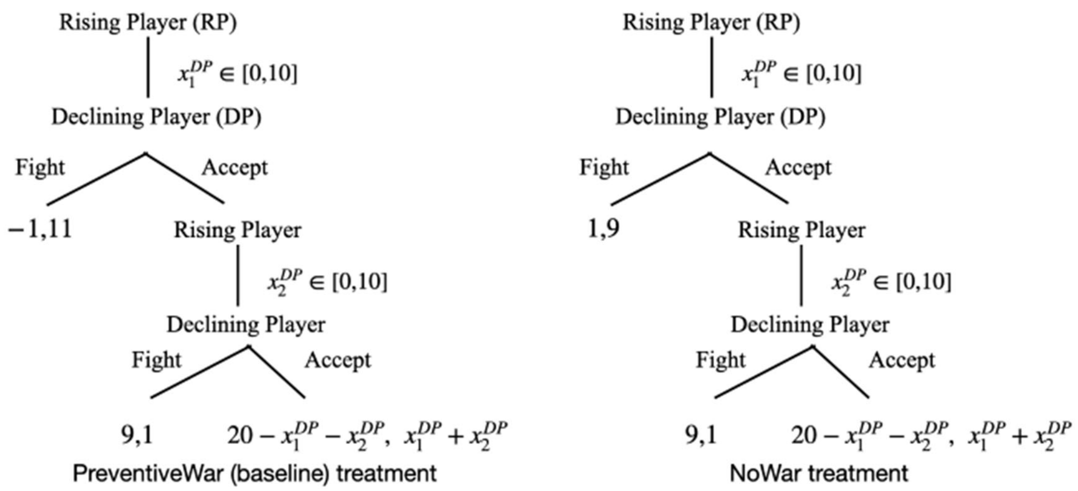

Figure 2: Parameterization in the PreventiveWar and NoWar treatments

Notes: The fighting payoffs are the expected payoffs given the winning probabilities in each situation. The payoffs are chosen such that theory predicts preventive war (fighting at Stage 1) in the baseline treatment and peace in the NoWar treatment.

3.1.3. The RP-Commit treatment

While the NoWar treatment is designed to test for the commitment problem through change in the parameter and therefore the incentive for launching a preventive war, in the *RP-Commit* treatment we manipulate directly the rising player's ability to commit to his plan. Specifically, RP must commit to his plan for Stage 2. In the instructions, RP is told that the computer will automatically implement his plan if the game proceeds to Stage 2. Thus, RP essentially makes decisions for both stages at the same time and the game essentially reduces to a one-stage game. All other parts of the design are the same as in the PreventiveWar treatment. The SPE predicts that the sum of RP's demands in both stages will be 20 * (1 – 0.8) + 5 = 9 tokens; DP chooses to accept in both stages and receives 20 * 0.8 – 5 = 11 tokens.

The upper panel of Table 1 summarizes the key features and equilibrium predictions of the three main treatments.

Table 1: The experimental design

Main treatments	**DP's prob. of winning in stage 1 (P_1^{DP})**	**Fighting cost (C)**	**Preventive war? (SPE)**	**Equilibrium payoff [a]**	**No. of participants**
PreventiveWar	0.8	5	Yes	$\pi_{RP} = -1$ $\pi_{DP} = 11$	80
NoWar	0.7	5	No	$\pi_{RP} = 11$ $\pi_{DP} = 9$	80
RP-Commit	0.8	5	No	$\pi_{RP} = 9$ $\pi_{DP} = 11$	70
Additional treatments					
Chat	0.8	5	Yes	$\pi_{RP} = -1$ $\pi_{DP} = 11$	80
Repeat	0.8	5	Yes	$\pi_{RP} = -1$ $\pi_{DP} = 11$	80
HighCost	0.8	9	No	$\pi_{RP} = 13$ $\pi_{DP} = 7$	40
LowCost	0.8	1	Yes	$\pi_{RP} = 3$ $\pi_{DP} = 15$	40

Notes: [a] All payoffs exclude a 5-token per-round endowment. The specifics of additional treatments are introduced in Section 5. In the *Chat* treatment, players could send free-form messages to each other for two minutes at the beginning of each round. In the *Repeat* treatment, each RP−DP pair is always matched and their roles are fixed across all rounds.

3.2. Hypotheses

Our main hypotheses regarding whether DP initiates the preventive war follow immediately from the SPE prediction:

Hypothesis 1: *DP is more likely to choose to fight in Stage 1 in PreventiveWar than in NoWar treatment.*

Remark 1: According to the SPE, we also expect that DP's likelihood to fight in Stage 1 is not conditional on RP's offer in PreventiveWar. However, we expect to observe a sharp decline in this likelihood when RP's offer in Stage 1 is at least 9 tokens in NoWar. Furthermore, we expect RP's non-binding announcement of his plan to be a mere cheap talk and thus it should not have any impact on DP's decision. If this is the case, this would provide indirect evidence supporting our theoretical claim that commitment is the key issue.

Hypothesis 2: *DP is more likely to choose to fight in Stage 1 in PreventiveWar than in RP-Commit treatment.*

Remark 2: In RP-Commit, according to the SPE, RP's combined offer across both stages should be 11 tokens. Consequently, we expect to observe a sharp decline in DP's likelihood of fighting in Stage 1 when RP's total offer is lower than 11 tokens.[6]

3.3. Experimental procedure

The experiment was programmed in z-Tree (Fischbacher 2007) and was conducted at the Economics Experimental Laboratory of the Nanjing Audit University (NEXL). We recruited 230 participants from a university-wide student pool and organized 23 sessions, each with 10 participants. The participants were randomly assigned to partitioned computer terminals upon arrival. They received experimental instructions (see Appendix A) in written form, which were also read aloud by the experimenter at the start of each session. The experiment started once all participants completed their comprehension quiz questions about the instructions. At the end of the experiment, participants completed a survey inquiring into their demographics and strategies used in the game. Participants were paid 1 RMB for every 5 tokens they accumulated in all rounds, in addition to 15 RMB for taking part in the experiment (with decimals in the final amount rounded to the nearest tenth). A typical session lasted about one hour with average earnings of 64.1 RMB (approximately US$9.9).

4. Experimental Results

[6] It is worth noting that, in equilibrium, DP's expected payoff in RP-Commit is the same as in PreventiveWar. Thus, DP could have chosen to fight in RP-Commit and would have obtained the same expected payoff. However, the resolution via fighting is both socially costlier and more uncertain than that via bargaining. Hence, if DP is risk-averse, one may expect that she would prefer not to initiate the preventive war when RP has the commitment power. However, risk aversion does not undermine the logic of preventive wars because the effectiveness of commitment power lies precisely in resolving the uncertainty about RP's offer in Stage 2. In the absence of this logic, risk aversion alone is unlikely to explain any differences in treatment outcomes.

4.1. Comparing PreventiveWar and NoWar: Does a larger power shift lead to more preventive wars?

Figure 3 shows the frequency of preventive wars, i.e. the likelihood of declining players' decision to fight in Stage 1 across treatments. As shown in Column (1) of Table 2, on aggregate, DP was significantly more likely to initiate preventive wars in PreventiveWar than in NoWar (56.6% vs. 41.0%, see hypothesis test H0: $\beta_0 = 0$, $p = 0.004$),[7] in line with Hypothesis 1. Figure B1 in Appendix B shows the frequency of preventive wars across rounds, suggesting a lower frequency in NoWar than in PreventiveWar in almost all rounds as well as a somewhat widened treatment difference in later rounds.

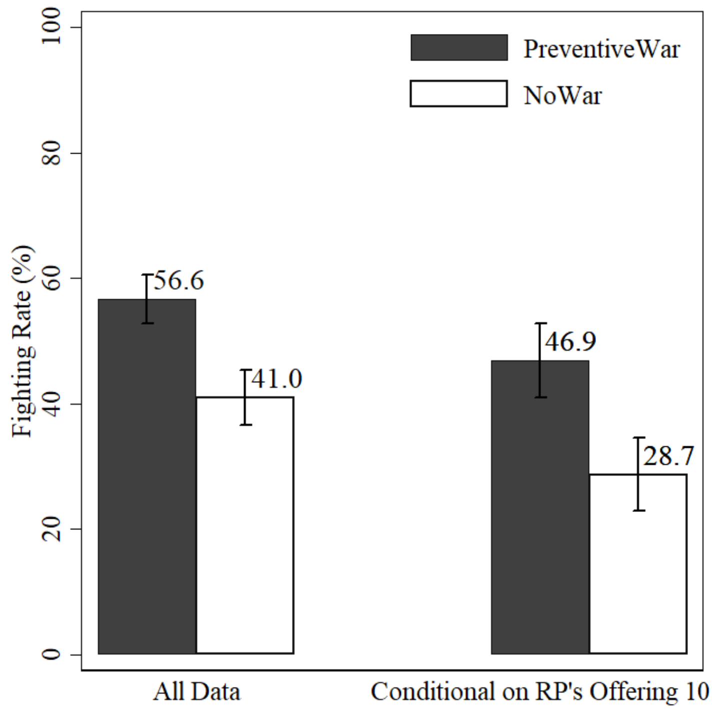

Figure 3: The frequency of preventive wars

Notes: This figure shows the frequency of preventive wars using the full sample as well as conditional on RP's offering 10 tokens at Stage 1. Error bars represent one standard error of the mean clustered at the session level.

Figure 4 shows rising players' offers in both stages. In the NoWar treatment, RP's offer of at least 9 tokens in Stage 1 should theoretically satisfy DP who should then refrain from fighting. However, in the PreventiveWar treatment, offering all 10 tokens, which is the most RP can do in Stage 1, should still be theoretically insufficient. We indeed observe that the amount offered in Stage 1 increases as the rounds progress and tends to the

[7] Unless otherwise stated, all p values in this subsection relate to the coefficient estimates from random effects probit regressions reported in Table 2.

maximum amount that RP could offer in Stage 1. In the NoWar treatment, RP offered on average 7.7 tokens to DP, with 9 tokens accounting for 9.4% and 10 tokens for 41.4% of cases. In the PreventiveWar treatment, the average offer was 8.3 tokens, with 9 tokens accounting for 7.5% and 10 tokens for 54.3% instances. Thus, offering 10 to DP in Stage 1 was the modal behavior in both treatments. On the other hand, if the game proceeds to Stage 2, RP should optimally offer 0 in both stages. Indeed, RP's Stage 2 offer declined to almost zero. In the NoWar and PreventiveWar treatments, the average offer was 0.78 and 0.47 tokens, respectively. There is no evidence of treatment difference in offers made in either stage.[8] These results strengthen our confidence that our subjects correctly interpreted the underlying incentive: those in the RP role tried their best to satisfy DP in Stage 1 even when it was impossible to do so, especially in the PreventiveWar treatment. However, they seized the chance to secure almost all tokens in Stage 2 when they were more powerful than DP.

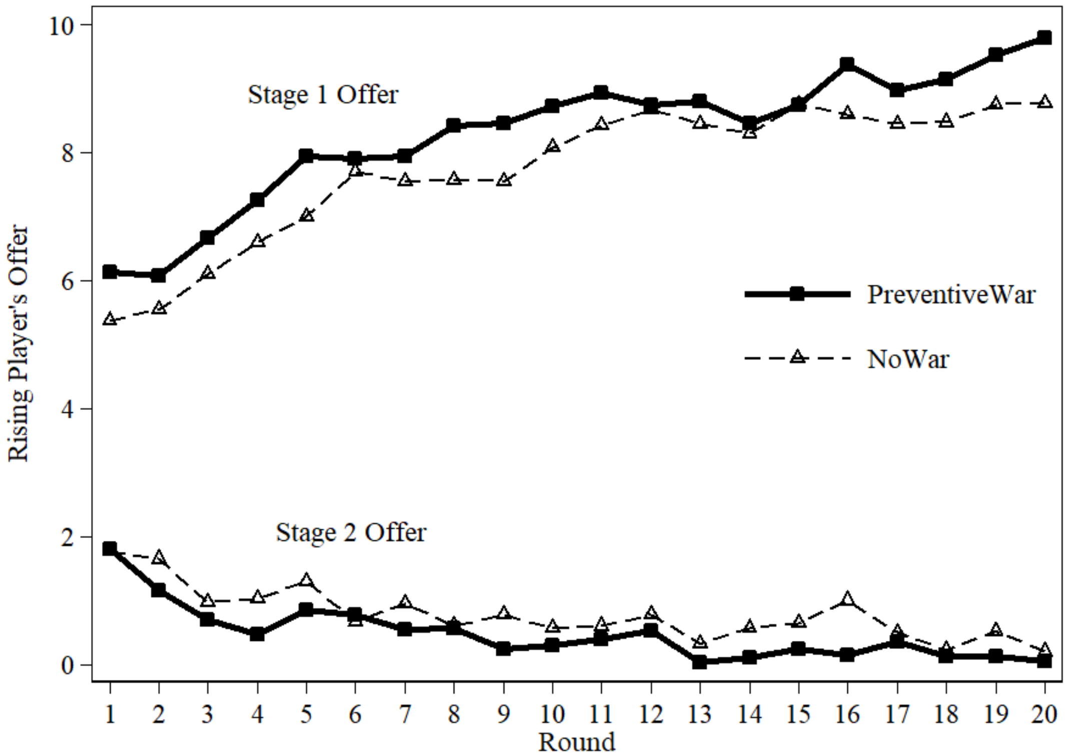

Figure 4: Rising players' offers in both stages over rounds

Next, we conduct conditional analyses to examine the influence of RP's Stage 1 offer on DP's decision.[9] Conditional on RP offering 10 tokens, which is theoretically sufficient to

[8] The test is conducted using a random effects regression by regressing either Stage 1 or Stage 2 offer on the "treatment" variable and "round" variable with standard errors clustered at the session level.
[9] Figure B2 in Appendix B shows the distribution of RP's Stage 1 offer and the frequency of preventive wars conditional on each possible offer in this stage.

avoid preventive wars in NoWar but not in PreventiveWar, we find a somewhat widened and significantly higher frequency of fighting in PreventiveWar than in NoWar (46.8% vs. 25.1%, see Figure 3, Column (2) in Table 2, and hypothesis test H0: $\beta_0 + \beta_2 = 0$, $p = 0.001$).[10] This is again consistent with Hypothesis 1. Conversely, conditional on RP offering fewer than 9 tokens, which is theoretically insufficient to avoid preventive wars in either treatment, the frequency of fighting was 69.3% in PreventiveWar and 53.6% in NoWar. The difference is not statistically significant (see Column (3) in Table 2 and hypothesis test H0: $\beta_0 + \beta_4 = 0$, $p = 0.142$). Thus, as predicted, when RP's Stage 1 offer could not satisfy DP (as it is below 9 tokens), DP's propensity to initiate preventive wars was not significantly different between the two treatments. Furthermore, the significant estimate of the "round" variable indicates that, in PreventiveWar, DP rejected RP's insufficient offer in Stage 1 more often in later rounds, suggesting that DP learned to adopt the equilibrium strategy more often.

It is worth noting that DP's propensity to initiate preventive wars significantly increased even in PreventiveWar when RP's offer was sufficiently low (see Column (3) in Table 2 and hypothesis test H0: $\beta_3 = 0$, $p < 0.001$). Thus, contrary to the theoretical prediction, DP's decision was conditional on RP's Stage 1 offer even when preventive wars were theoretically unavoidable. This observation might be attributed to DP's risk aversion, since securing 9 or 10 tokens might be more attractive to DP than receiving 11 tokens in expectation through the preventive war. It might also be ascribed to the focal behavior when the Stage 1 offer was 10, which—albeit not theoretically sufficient for DP to avoid fight—was perceived as fair allocation and was thus accepted with high likelihood.

The theory also predicts that RP's announcement of the allocation plan for Stage 2 should not have any impact on DP's decision in either treatment. This is confirmed in the data (see Column (4) in Table 2, where β_5 and β_6 estimates are insignificant), providing further evidence that our subjects understood the commitment problem and treated such announcement as a mere cheap talk.

Next, we briefly examine players' decisions when the game proceeds to Stage 2 (i.e. when the preventive war did not occur). In PreventiveWar, DP chose to fight in Stage 2 in 11.0% of the cases, almost exclusively when RP's offered fewer than 10 tokens in total. A similar pattern was observed in NoWar, in which DP chose to fight in Stage 2 in 12.3% cases, again mostly in response to the total offer of fewer than 10 tokens. We also find that in both treatments RP's total offer was never lower than 3 tokens. Given that DP's expected payoff from fighting in Stage 2 (excluding the 5-token endowment) was only one token, fighting in this stage was never a materially optimal decision for DP. Thus, although the Stage 1 fighting rate is generally low, the fact that DP sometimes chose to fight suggests that they might care about fair allocation even at their own cost.

[10] Theoretically, RP's offer of 9 tokens is sufficient to avoid preventive wars in the NoWar treatment. The regression results are robust to this offer amount in Stage 1 as the independent variable.

Table 2: Random effects probit regressions of preventive wars (without commitment)

Dependent Variable: fight in Stage 1 = 1	(1)	(2)	(3)	(4)
β_0: NoWar	-0.189***	-0.193**	-0.271***	-0.224***
	(0.055)	(0.096)	(0.075)	(0.085)
β_1: **1**[RP's Stage 1 offer = 10]		-0.396***		0.396***
		(0.055)		(0.054)
β_2: NoWar × **1**[RP's Stage 1 offer = 10]		-0.102		-0.097
		(0.090)		(0.084)
β_3: **1**[RP's Stage 1 offer < 9]			0.381***	
			(0.070)	
β_4: NoWar × **1**[RP's Stage 1 offer < 9]			0.104	
			(0.102)	
β_5: RP's Stage 2 plan				-0.003
				(0.006)
β_6: NoWar × RP's Stage 2 plan				0.007
				(0.011)
Round	-0.001	0.016***	0.017***	0.016***
	(0.003)	(0.004)	(0.004)	(0.004)
H0: $\beta_0 + \beta_2 = 0$		$p = 0.001$		$p = 0.002$
H0: $\beta_0 + \beta_4 = 0$			$p = 0.142$	
Observations	1600	1600	1600	1600
Clusters	16	16	16	16

Notes: Standard errors clustered at the session level are given in parentheses. Average marginal effects are reported. PreventiveWar serves as the benchmark. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Last, because of the less frequent war at Stage 1 in NoWar but otherwise similar behavior in Stage 2, RP's final payoff (excluding the endowment) was significantly higher in NoWar than in PreventiveWar (6.1 vs. 3.7, see Figure 5 and Column (1) of Table B1 in Appendix B for statistical evidence yielded by a random effects regression, H0: $\beta_0 = 0$, $p = 0.003$). By contrast, DP's final payoff was significantly lower in NoWar than in PreventiveWar (9.1 vs. 10.2, see Column (1) of Table B1, H0: $\beta_0 + \beta_3$, $p = 0.004$). These results are qualitatively consistent with the theoretical prediction. However, RP still earned much more than DP, let alone reversing the earning gap in NoWar as theory would predict (see Table 1).

In sum, the results related to the PreventiveWar and NoWar treatments suggest that the initiation of preventive wars is not due to DP's unconditional propensity to fight. In contrast to the NoWar treatment, RP is simply unable to satisfy DP at Stage 1 in the PreventiveWar treatment. Observing more fighting at Stage 1 in the PreventiveWar treatment confirms our theoretical claim that the inability to commit to one's decision in the future stage is one of the key reasons for the preventive war.

Result 1: *The greater power shift, which results in RP's inability to satisfy DP within one stage, causes more preventive wars. As a result, RP earned less while DP earned more compared to the scenario in which RP could satisfy DP within one stage.*

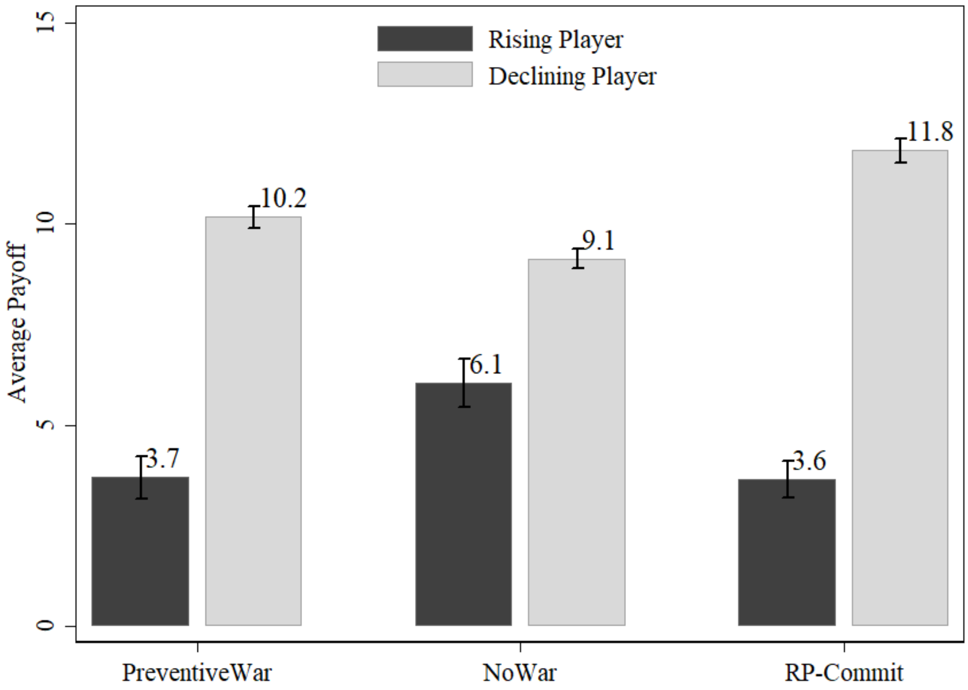

Figure 5: RP's and DP's average payoffs (excluding 5-token endowment)

Note: Error bars represent one standard error of the mean clustered at the session level.

4.2. Comparing PreventiveWar and RP-Commit: Does commitment capability lead to fewer preventive wars?

To provide more direct evidence for the commitment problem, in the RP-Commit treatment we allowed RP to commit to offers in both stages at the beginning of the game. On aggregate, DP was less likely to initiate the preventive war compared to PreventiveWar (44.3% vs. 56.6%, p = 0.003, see Column (1) in Table 3 and hypothesis test H0: $\beta_0 = 0$, p = 0.007),[11] in line with Hypothesis 2. Thus, the commitment power does help reduce the

[11] Unless otherwise stated, all p values in this subsection relate to the coefficient estimates from random effects probit regressions reported in Table 3.

likelihood of the preventive war. Figure B1 in Appendix B shows the frequency of preventive wars across rounds, suggesting a lower frequency in RP-Commit than in PreventiveWar in most rounds. The frequency also tends to decrease across successive rounds in RP-Commit.

To examine in more detail whether RP used his commitment power to his advantage, we next test how DP's decision depends on RP's offers in two stages.[12] In RP-Commit, RP offered 10 tokens in Stage 1 in only 24.1% of the cases because RP tended to even out the offers across two stages. Given RP's commitment power, it is thus more informative to directly examine RP's total offer, which steadily increased across successive rounds and generally exceeded the theoretical optimal amount of 11 tokens. Conditional on RP offering at least 11 tokens in total, which is theoretically sufficient to avoid preventive wars, the frequency of fighting was 31.8%. By contrast, conditional on RP offering fewer than 11 tokens in total, the frequency was 59.8%. The difference is statistically significant (see Column (2) in Table 3 and hypothesis test H0: $\beta_1 = 0$, $p < 0.001$). These findings suggest that commitment helps reduce the likelihood of preventive war whenever RP's total offer is sufficiently high according to the theoretical threshold. However, RP did not always use his commitment power to avoid the war, as his total offer was still lower than 11 tokens in 42.2% of the cases (in which preventive wars were prevalent, as noted earlier).

Table 3: Random effects probit regressions of preventive wars (with commitment)

Dependent Variable: fight in Stage 1 = 1	(1)	(2)
β_0: RP-Commit	-0.150***	
	(0.056)	
β_1: **1**[RP's total offer $\geq$ 11]		-0.366***
		(0.039)
Round	-0.005	-0.003
	(0.004)	(0.005)
Observations	1500	700
Clusters	15	7

Notes: Standard errors clustered at the session level are given in parentheses. Average marginal effects are reported. PreventiveWar serves as the benchmark in Column (1), and Column (2) only includes RP-Commit. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In addition to the issue of making inadequate offer, RP sometimes offered more than 11 tokens, which helped avoid conflict. Figure 5 shows how RP's offer-making pattern

[12] Figure B3 in Appendix B shows the distribution of RP's total offer and the frequency of preventive wars conditional on each possible total offer.

affected his final payoff. RP's final payoff in RP-Commit was almost the same as that in PreventiveWar (3.6 vs. 3.7, see Column (2) of Table B1 in Appendix B, H0: $\beta_1 = 0$, p = 0.943). On the other hand, DP's final payoff was significantly higher in RP-Commit than in PreventiveWar (11.8 vs. 10.2, see Column (2) of Table B1 in Appendix B, H0: H0: $\beta_1 + \beta_4 = 0$, $p < 0.001$). Thus, while RP used his commitment power to reduce the frequency of preventive wars, he failed to generate material advantage. In the end, perhaps surprisingly, it was DP who benefited from RP's commitment power.

Result 2: *Endowing RP with the power of committing to the future offer led to less frequent preventive wars. However, RP's offer sometimes exceeded the equilibrium level, resulting in no improvement in his final payoff.*

5. Additional Treatments

5.1. Do other mechanisms help reduce preventive wars in the absence of commitment capability?

The main three treatments have shown that the lack of power to commit to bargaining allocations over multiple stages is the key to the logic of preventive wars. However, absence of commitment power is a real-world feature which we cannot simply assume away or expect some party to either self-commit or be forced to commit. In reality, states with conflicting interests usually strive to maintain peace through other mechanisms. One diplomatic strategy commonly adopted in this case is resolving conflict of interest through bargaining via direct communication. A large body of behavioral economics literature has shown that communication, even cheap talk, can promote cooperation and achieve higher efficiency (e.g. Dawes, McTavish, and Shaklee 1977; Balliet 2009). Moreover, throughout history, potential conflicts such as the Cuban Missile Crisis were indeed avoided via diplomatic means.

Another conflict-prevention mechanism involves building reputation by maintaining a tough position during the bargaining process or even by engaging in limited conflict. Reputation is, however, often developed gradually and requires repeated and consistent interaction. A considerable effort has been devoted in the extant theoretical and experimental literature to understanding how two interacting parties maintain cooperation through history-dependent strategies such as tit-for-tat (e.g., Axelrod 1984; Dal Bó and Fréchette 2018).

In this section, we investigate whether communication and repeated interaction help decrease the risk of preventive wars in the absence of commitment capability (note that both mechanisms might convey "soft" commitment). We thus designed two additional treatments based on the baseline PreventiveWar treatment.

To investigate the effect of communication, we introduce the *Chat* treatment, which differs from PreventiveWar only in that RP and DP could send free-form messages to each other for two minutes at the beginning of each round.

To investigate the effect of repeated interaction, we introduce the *Repeat* treatment, which differs from PreventiveWar only in that a pair of RP and DP would interact in all 20 rounds and each participant would assume the same role as either RP or DP throughout the session. All other aspects of the experimental procedure are the same as in the PreventiveWar treatment.

We conducted eight sessions for each treatment, for which 160 participants were recruited from the same student pool. The lower panel of Table 1 summarizes the design of the Chat and Repeat treatments. Note that, as shown in Table 1, in both treatments, the SPE predicts the occurrence of preventive wars.

5.1.1. Main results

Figure 6 shows the frequency of preventive wars in the Chat and Repeat treatments. Compared to PreventiveWar, DP was significantly less likely to initiate preventive war both when she could talk to RP in the Chat treatment (56.6% vs. 28.9%, see Column (1) in Table 4, H0: $\beta_0 = 0$, $p < 0.001$) and when she could repeatedly play with the same RP in the Repeat treatment (56.6% vs. 37.4%, see Column (1) in Table 4, H0: $\beta_1 = 0$, $p < 0.001$). Furthermore, comparing Chat and Repeat, the communication mechanism was not significantly more effective in curbing preventive wars than the reputation mechanism ($\beta_0 = \beta_1$, $p = 0.288$). Figure B4 in Appendix B shows the frequency of preventive wars across rounds, suggesting a lower frequency in Chat and Repeat than in PreventiveWar in almost all rounds.[13]

Next, we turn to the conditional analysis and test how DP's decision depends on RP's Stage 1 offer.[14] Conditional on RP offering 10 tokens, the frequency of fighting was 23.1% in the Chat treatment and 32.5% in the Repeat treatment, both of which are significantly lower than that in the PreventiveWar treatment (see Figure 6 and Column (2) in Table 4, H0: $\beta_0 + \beta_3 = 0$, $p < 0.001$, H0: $\beta_0 + \beta_4 = 0$, $p = 0.001$). Conditional on RP's offer of fewer than 9 tokens, the frequency of fighting was 42.3% and 38.3% in the Chat and Repeat treatments, respectively, both of which were again significantly lower than that in PreventiveWar treatment (Column (3) in Table 4, H0: $\beta_0 + \beta_6 = 0$, $p < 0.001$, H0: $\beta_0 + \beta_7 = 0$, $p < 0.001$). Thus, compared to PreventiveWar, both mechanisms help reduce the likelihood of

[13] Figure B5 in Appendix B shows rising players' offers in both stages, exhibiting a similar pattern in the Repeat treatment as in the PreventiveWar treatment: Stage 1 offer was increasing across successive rounds and approached the maximum amount that RP could offer in Stage 1, whereas RP's Stage 2 offer was close to zero across all rounds. In the Chat treatment, we observe that RP often offered either 10 or 5 tokens in Stage 1. Interestingly, the offer of 5 tokens was rejected only slightly more frequently (31.2%, see Figure B6) than offer of 0 tokens. This pattern was different from that in PreventiveWar in which 5-token offer was rejected 66.1% of the time. Thus, it appears that players coordinated on one of the two types of division of the total prize amount. One is the most typical combination of RP offering 10 tokens in Stage 1 and 0 tokens in Stage 2. The other is RP's offer of 5 tokens in both stages. Both types of division would result in an equal split of the total winnings. Content analysis of players' communication messages provides supporting evidence of players' coordination on the above divisions. Further details are available upon request.

[14] Figure B6 in Appendix B shows the distribution of RP's Stage 1 offer and the frequency of preventive wars conditional on each possible Stage 1 offer in the Chat and Repeat treatments.

preventive wars regardless of RP's Stage 1 offer. Nonetheless, the higher the Stage 1 offer, the lower the frequency of preventive wars.[15]

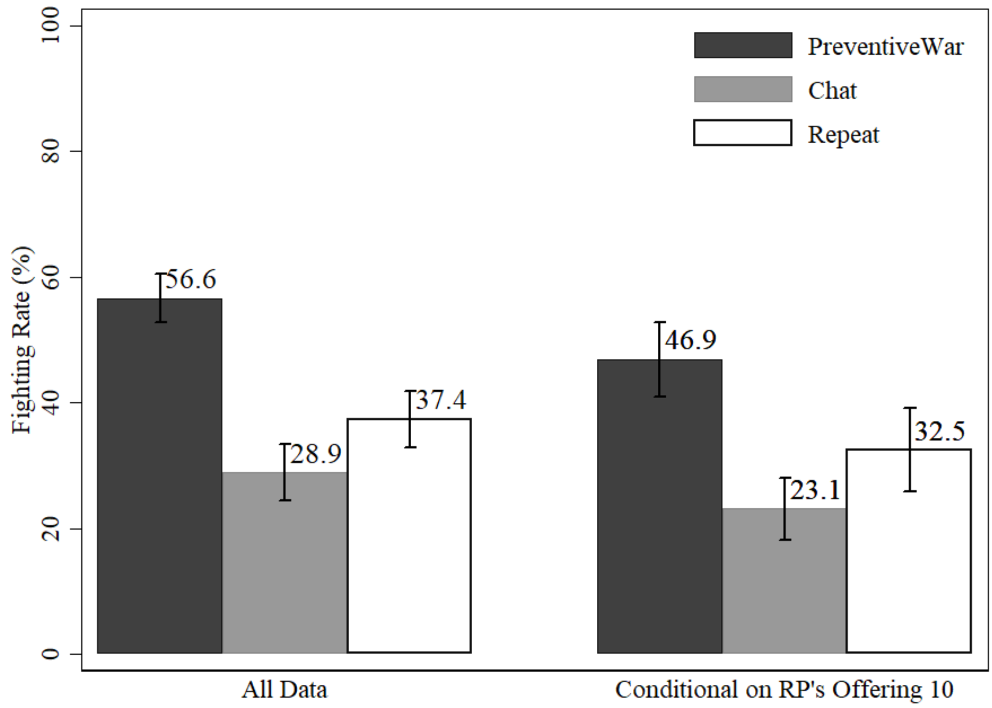

Figure 6: The frequency of preventive wars (Chat and Repeat treatments)

Notes: This figure shows the frequency of preventive wars using the full sample as well as conditional on RP offering 10 tokens at Stage 1. Error bars represent one standard error of the mean clustered at the session level.

Next, in line with the theory and the findings from the PreventiveWar treatment, RP's announcement of his plan for Stage 2 had little impact on DP's decision in either Chat or Repeat treatments (see Column (4) in Table 4, where β_8, β_9 and β_{10} estimates are insignificant). This result also suggests that the effectiveness of the two mechanisms cannot be attributed to DP's confidence that RP would execute his plan when making Stage 2 allocation.

[15] It is worth noting that the decision to initiate preventive war is clearly path-dependent in the Repeat treatment: DP was significantly more likely to fight in Stage 1 if preventive wars occurred in previous rounds (see Table B4 in Appendix B for probit regression analysis probing into the effect of the occurrence of preventive wars in previous rounds on the likelihood of preventive wars in the current round). After accounting for the path-dependent effect, the size of Stage 1 offer no longer seemed to be relevant.

Table 4: Random effects probit regressions of preventive wars (Chat and Repeat treatments)

Dependent Variable: reject in Stage 1 = 1	(1)	(2)	(3)	(4)
β_0: Chat	-0.326***	-0.280***	-0.323***	-0.245***
	(0.053)	(0.072)	(0.071)	(0.070)
β_1: Repeat	-0.243***	-0.306***	-0.281***	-0.355***
	(0.065)	(0.080)	(0.082)	(0.086)
β_2: **1**[RP's Stage 1 offer = 10]		-0.358***		-0.359***
		(0.066)		(0.065)
β_3: Chat × **1**[RP's Stage 1 offer = 10]		-0.018		-0.052
		(0.085)		(0.092)
β_4: Repeat × **1**[RP's Stage 1 offer = 10]		-0.008		-0.010
		(0.091)		(0.090)
β_5: **1**[RP's Stage 1 offer < 9]			0.326***	
			(0.073)	
β_6: Chat × **1**[RP's Stage 1 offer < 9]			0.025	
			(0.091)	
β_7: Repeat × **1**[RP's Stage 1 offer < 9]			-0.030	
			(0.094)	
β_8: RP's Stage 2 plan				-0.002
				(0.006)
β_9: Chat × RP's Stage 2 plan				-0.008
				(0.010)
β_{10}: Repeat × RP's Stage 2 plan				0.010
				(0.011)
Round	-0.001	0.012***	0.010***	0.012***
	(0.002)	(0.003)	(0.003)	(0.003)
H0: $\beta_0 + \beta_3 = 0$		$p < 0.001$		
H0: $\beta_0 + \beta_4 = 0$		$p = 0.001$		
H0: $\beta_0 + \beta_6 = 0$			$p < 0.001$	$p < 0.001$
H0: $\beta_0 + \beta_7 = 0$			$p < 0.001$	$p < 0.001$
Observations	2400	2400	2400	2400
Clusters	56	56	56	56

Notes: Standard errors are clustered at the session level for PreventiveWar and Chat, and at the group level for Repeat. Average marginal effects are reported. PreventiveWar serves as the benchmark. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Finally, because of the lower frequency of preventive wars and the low frequency of rejection in Stage 2,[16] both mechanisms help increase RP's payoffs. Compared to PreventiveWar, RP's payoff was significantly higher in both the Chat treatment (6.4 vs. 3.7, see Figure B7 and Column (1) of Table B2 in Appendix B, H0: $\beta_0 = 0$, $p = 0.001$) and the Repeat treatment (5.6 vs. 3.7, see Column (2) of Table B2 in Appendix B, H0: $\beta_1 = 0$, $p = 0.010$). Neither mechanism, however, decreased DP's payoff (in Chat, 10.4 vs. 10.2, see Column (1) of Table B2, H0: $\beta_0 + \beta_3 = 0$, $p = 0.586$; in Repeat, 10.1 vs. 10.2, see Column (2) of Table B2, H0: $\beta_1 + \beta_4 = 0$, $p = 0.865$).

Result 3: *Both communication and reputation mechanisms helped reduce the frequency of preventive wars and led to higher RP's final payoff.*

5.2. Is the likelihood of preventive wars sensitive to the fighting cost?

In two of our main treatments (PreventiveWar and NoWar), we only slightly changed the expected payoff of preventive wars so that the theory predicts preventive war in one case and peace in the other. Given such parameterization, however, these treatments provide a rather conservative test of the theoretical prediction. The results also suggest that DP was far from making theoretically optimal decision in either treatment (the frequency of preventive wars was 56.6% instead of 100% in PreventiveWar, and 41.0% instead of 0% in NoWar). Here, we introduce two additional treatments that allow a greater treatment separation by widening the gap in the expected payoffs of preventive wars still further. We achieve this goal by changing the fighting cost C (see Figure 1).

In the *HighCost* treatment, $C = 9$ and the DP's expected payoff of fighting in Stage 1 is only 7 tokens. Therefore, as in the NoWar treatment, RP could satisfy DP in Stage 1 and preventive wars could be avoided in theory. By contrast, in the *LowCost* treatment, we adopt $C = 1$ and the DP's expected payoff of fighting in Stage 1 increases to 15 tokens. Therefore, as in the PreventiveWar treatment, as RP cannot satisfy DP in Stage 1, preventive wars are unavoidable. Note that the gap in the expected payoff of preventive wars between HighCost and LowCost increases to 8 tokens compared to merely 2 tokens between NoWar and PreventiveWar.

We conducted four sessions for each treatment, recruiting a total of 80 participants from the same student pool. The lower panel of Table 1 summarizes the design of the HighCost and LowCost treatments. Figure 7 displays the game trees of the two treatments with their respective parameterization.

[16] Similar to PreventiveWar, in Chat, DP fought in Stage 2 in 4.4% of the cases, and only when the total offer consisted of fewer than 10 tokens. A similar pattern was observed in Repeat, where DP fought in Stage 2 8.0% of the time, but sometimes for the total offer exceeding 10 tokens.

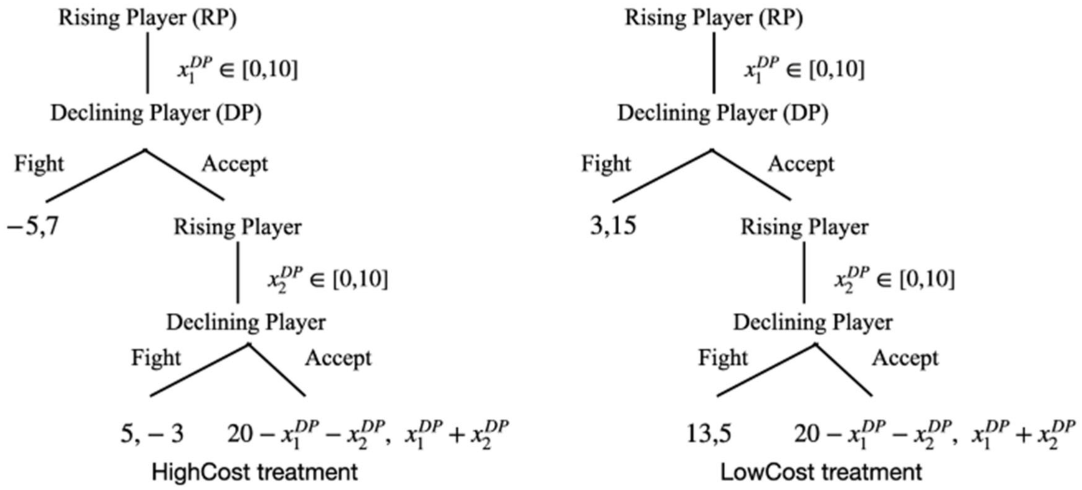

Figure 7: Parameterization in the HighCost and LowCost treatments

Notes: The fighting payoffs are the expected payoffs given the winning probabilities and fighting cost ($C = 9$ in *HighCost* and $C = 1$ in *LowCost*) in each situation. The payoffs are chosen such that theory predicts preventive war (fighting at Stage 1) in the HighCost treatment and peace in the LowCost treatment.

5.2.1. Main results

Figure 8 shows the frequency of preventive wars in the HighCost and LowCost treatments. Compared to PreventiveWar, DP was significantly less likely to initiate the preventive war when the fighting cost was high (56.6% vs. 26.8%, see Column (1) in Table 5, H0: $\beta_0 = 0$, $p < 0.001$). By contrast, the frequency of preventive wars was significantly higher when the fighting cost was low (56.6% vs. 71.8%, see Column (1) in Table 5, H0: $\beta_1 = 0$, $p = 0.006$). This finding is consistent with the theoretical prediction. Figure B8 in Appendix B shows the frequency of preventive wars across rounds, revealing a consistently higher fighting rate in the HighCost treatment and a lower fighting rate in the LowCost treatment compared to the PreventiveWar treatment. In particular, the preventive war frequency in the HighCost treatment was close to zero in later rounds. Thus, DP's decision was indeed sensitive to the fighting cost parameter, leading to a much greater treatment difference and almost eliminating preventive wars when the fighting cost was high enough.[17]

[17] Figure B9 in Appendix B shows rising players' offers in both stages, exhibiting a very similar pattern in both treatments as in the PreventiveWar treatment.

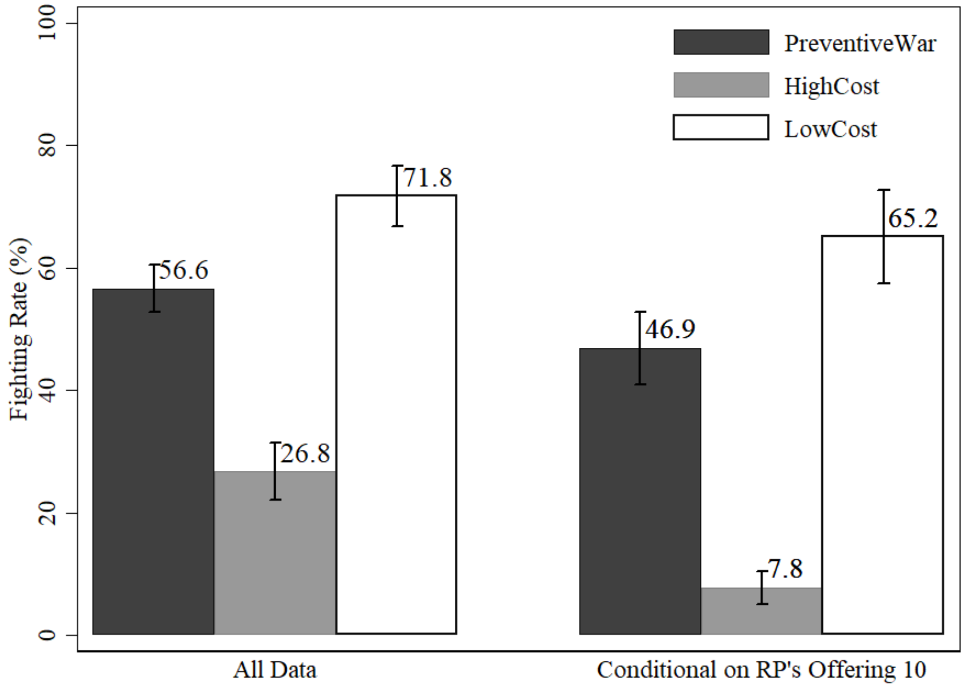

Figure 8: The frequency of preventive wars (HighCost and LowCost treatments)

Notes: This figure shows the frequency of preventive wars using the full sample as well as conditional on RP offering 10 tokens at Stage 1. Error bars represent one standard error of the mean clustered at the session level.

Turning to the conditional analysis, conditional on RP offering 10 tokens, the frequency of preventive wars was only 7.8% in the HighCost treatment, which was significantly lower than that in the PreventiveWar treatment (see Figure 8 and Column (2) in Table 5, H0: $\beta_0 + \beta_3 = 0$, $p < 0.001$). The frequency remained high (65.2%) in the LowCost treatment, and was significantly higher than that in the PreventiveWar treatment (see Column (2) in Table 5, H0: $\beta_0 + \beta_4 = 0$, $p = 0.014$).[18] It is worth noting that, when the fighting cost was low, the frequency of preventive wars was no longer sensitive to the Stage 1 offer: the frequency did not differ between cases in which the offer was 10 and less than 10 tokens (see Column (2) in Table 5, H0: $\beta_4 = 0$, $p = 0.468$). This result implies that the equal split of all resources is not necessarily considered as a focal behavior, which appears to have contributed to the low frequency of preventive wars when the Stage 1 offer was 10 tokens in the PreventiveWar treatment. Instead, DP considered the size of the expected payoff of preventive wars, as implied by the logic of preventive wars.

[18] Figure B10 in Appendix B shows the distribution of RP's stage 1 offer and the frequency of preventive wars conditional on each possible Stage 1 offer in the HighCost and LowCost treatments.

Table 5: Random effects probit regressions of preventive wars (HighCost and LowCost treatments)

Dependent Variable: reject in Stage 1 = 1	(1)	(2)
β_0: HighCost	-0.328***	-0.025
	(0.061)	(0.067)
β_1: LowCost	0.190***	0.158***
	(0.069)	(0.053)
β_2: **1**[RP's Stage 1 offer = 10]		-0.348***
		(0.049)
β_3: HighCost × **1**[RP's Stage 1 offer = 10]		-0.422***
		(0.095)
β_4: LowCost × **1**[RP's Stage 1 offer = 10]		0.073
		(0.101)
Round	-0.004	0.016***
	(0.006)	(0.004)
H0: $\beta_0 + \beta_3 = 0$		$p < 0.001$
H0: $\beta_0 + \beta_4 = 0$		$p = 0.014$
Observations	1600	1600
Clusters	16	16

Notes: Standard errors are clustered at the session level. Average marginal effects are reported. PreventiveWar serves as the benchmark. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

However, when the fighting cost was high, as shown in Figure B10 in Appendix B, the preventive war frequency was particularly sensitive to whether the Stage 1 offer was 10 tokens (see Column (2) in Table 5, H0: $\beta_3 = 0$, $p < 0.001$). Thus, DP did not apply the logic of preventive wars "asymmetrically" but rather insisted on fair allocation although a Stage 1 offer exceeding 7 tokens was theoretically sufficient to avoid preventive wars. This behavior might be explained by DP's willingness to teach RP a hard "lesson" at her own cost. This strategy appears to work well for both players since the Stage 1 offer was almost always 10 tokens and the preventive wars was almost eliminated in later rounds.

Finally, as expected, because of the rather low frequency of preventive wars in the HighCost treatment, RP's and DP's payoffs became more equalized. RP's payoff was significantly higher than that in the PreventiveWar treatment (5.8 vs. 3.7, see Figure B11 and Column (1) of Table B3 in Appendix B, H0: $\beta_0 = 0$, $p = 0.021$), whereas DP's payoff was significantly lower than that in the PreventiveWar treatment (9.0 vs. 10.2, see Column (1) of Table B3, H0: $\beta_0 + \beta_3 = 0$, $p < 0.001$). By contrast, because of the high frequency of preventive wars in the LowCost treatment, RP's and DP's payoffs became even more unequal. However, due to the lower fighting cost, RP's payoff was still marginally significantly higher than that in the PreventiveWar treatment (4.8 vs. 3.7, see Figure B11 and Column (2) of Table B3 in Appendix B, H0: $\beta_1 = 0$, $p = 0.059$). On the other hand, DP's payoff was significantly higher than that in the PreventiveWar treatment (13.7 vs. 10.2, see Column (2) of Table B3, H0: $\beta_1 + \beta_4 = 0$, $p < 0.001$).

Result 4: *A higher fighting cost led to much lower frequency of preventive wars and almost eliminated them in later rounds. By contrast, a lower fighting cost led to even higher frequency of preventive wars which was insensitive to the Stage 1 offer.*

6. Summary and Conclusions

Emergence of a new power carries the risk that the incumbent hegemon would seek conflict to prevent the newcomer from taking over its position. We model this scenario as a two-stage bargaining game in which the rising power makes an offer to the declining power as to how to split a pie. The declining power can either accept the offer or fight. This process is repeated, whereby the probability of winning the fight shifts towards the rising power between the stages. Depending on the parameter selection, the subgame perfect equilibrium involves either war or peace.

In this study, we tested (1) the game theoretic predictions in the laboratory, and (2) institutions that are often shown to be effective in resolving conflict. Overall, our results support the comparative statics of the theory. In the treatments in which war is predicted, conflict rates are higher than in those for which theory predicts peace. However, the expected sharp separation—only war or only peace—does not materialize. Conflict rates in the baseline treatments are surprisingly similar, which has both positive and negative ramifications. On a positive note, even where war should be inevitable, subjects in the lab often find ways to avoid it. The downside is that, even where peace should prevail, the situation is still volatile and many players fail to implement the more efficient peaceful solution. Wars virtually disappear only when fighting costs are so high that conflict becomes very unattractive. The policy implications of these observations are both obvious and counter-intuitive. They are readily apparent because they imply that significant fighting costs would deter wars. Yet, they are also counter-intuitive because fighting costs are most easily increased by enhancing the destructive potential of war through massive military build-up, which is not what we would naturally regard as a recipe for peace.

Still, as was shown here, several remedial mechanisms can be employed to reduce the likelihood of conflict. Both repeated interaction (which may serve as a proxy for an intensive exchange between powers) and pre-play communication lead to lower conflict rates. These findings imply that diplomacy can play a decisive role in conflicts between a rising and a declining power. That said, diplomacy is unlikely to result in miracles as, while it led to a significant reduction in the likelihood of conflict, its impact was quantitatively less substantial than one might hope for.

Since, as suggested by Powell (2004, 2006), the same mechanism may be at work not only in international conflicts, but may be involved in a wide range of diverse phenomena such as civil wars, revolutions, policy insulation, strikes, and predatory pricing, the evidence presented in this work may also be useful to informing research in these areas. In future studies in this domain, our experimental framework may be extended to allow for endogenous winning probabilities. For example, a higher today's bargaining gain increases one's military capacity tomorrow. We expect two opposing effects depending on how much the rising state is willing to concede today. The declining state would have less reason to attack today when the concessions offered by the rising state are sufficient because of reduced fear associated with a smaller (endogenous) power shift. Conversely, the declining state would have more reason to strike today when the rising state fails to offer adequate concessions due to stronger fear associated with a greater power shift. Intuitively, an endogenous shift in power should be more conducive to preventive wars compared to the exogenous one considered in the present study.

Of course, our study has limitations which should be noted when interpreting our findings. The most obvious one is the small scale of our experimental scenario compared to an armed conflict in the real world, especially if both powers have substantial nuclear arsenals. Losing a few yuan in the lab is substantial for a university student, but it is not the same as losing hundreds of millions of lives in an all-out war between, say, the U.S. and China. Unfortunately, we cannot wait for a sufficiently large set of real-life observations to appear that would allow us to run empirical analysis beyond studying historical anecdotes, as too many lives would be lost in the process. In addition, in contrast to historical studies, we can control the experimental conditions and can ensure that everything else is equal when required. The laboratory, with its controlled, replicable and damage-free environment is thus a valuable tool for testing various war scenarios. Laboratory experiments should therefore be seen as a complement to historical analysis, as both have their particular strengths and weaknesses.

References

Abbink, Klaus, Lu Dong, and Lingbo Huang. 2021. "Arms Races and Conflict: Experimental Evidence." *The Economic Journal* 131 (637): 1883–1904.

Acemoglu, Daron, and James A. Robinson. 2000. "Why Did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective." *Quarterly Journal of Economics* 115 (4): 1167–99.

Acemoglu, Daron, and James A Robinson. 2001. "A Theory of Political Transitions." *American Economic Review* 91 (4): 938–63.

Allison, Graham. 2017. *Destined for War: Can America and China Escape Thucydides's Trap?* Houghton Mifflin Harcourt.

Ausubel, L. M., P. Cramton, and R. J. Deneckere. 2002. “Bargaining with Incomplete Information.” In *Handbook of Game Theory*, edited by Robert J Aumann and S Hart, 1897–1945. Elsevier.

Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York, NY: Basic Books.

Baliga, Sandeep, and Tomas Sjöström. 2013. “Bargaining and War: A Review of Some Formal Models.” *Korean Economic Review* 29 (2): 235–66.

———. 2020. “The Strategy and Technology of Conflict.” *Journal of Political Economy* 128 (8): 3186–3219. https://doi.org/10.1086/707767.

Balliet, Daniel. 2009. “Communication and Cooperation in Social Dilemmas: A Meta-Analytic Review.” *Journal of Conflict Resolution* 54 (1): 39–57.

Beviá, Carmen, and Luis C. Corchón. 2010. “Peace Agreements without Commitment.” *Games and Economic Behavior* 68 (2): 469–87.

Brito, Dagobert L, and Michael D Intriligator. 1985. “Conflict, War and Redistribution.” *American Political Science Review* 79 (4): 943–57.

Chadefaux, Thomas. 2011. “Bargaining over Power: When Do Shifts in Power Lead to War?” *International Theory* 3 (2): 228–53.

Clark, Christopher. 2014. *The Sleepwalkers: How Europe Went to War in 1914*. Harper Perennial.

Coursey, Don L., and Linda R. Stanley. 1988. “Pretrial Bargaining Behavior within the Shadow of the Law: Theory and Experimental Evidence.” *International Review of Law and Economics* 8 (2): 161–79.

Dal Bó, Pedro, and Guillaume R. Fréchette. 2018. “On the Determinants of Cooperation in Infinitely Repeated Games: A Survey.” *Journal of Economic Literature* 56 (1): 60–114.

Dawes, Robyn M., Jeanne McTavish, and Harriet Shaklee. 1977. “Behavior, Communication, and Assumptions about Other People’s Behavior in a Commons Dilemma Situation.” *Journal of Personality and Social Psychology* 35 (1): 1–11.

Dechenaux, Emmanuel, Dan Kovenock, and Roman M Sheremeta. 2015. “A Survey of Experimental Research on Contests, All-Pay Auctions and Tournaments.” *Experimental Economics* 18 (4): 609–69.

Fearon, James D. 1995. “Rationalist Explanations for War.” *International Organization* 49 (3): 379–414.

———. 1996. “Bargaining over Objects That Influence Future Bargaining Power.” mimeo, Standard University.

Ferguson, Niall. 2019. “The New Cold War? It’s With China, and It Has Already Begun.” *The New York Times*, December 2, 2019. https://www.nytimes.com/2019/12/02/opinion/china-cold-war.html.

Figueiredo, Rui J.P. De. 2002. “Electoral Competition, Political Uncertainty, and Policy Insulation.” *American Political Science Review* 96 (2): 321–33.

Fischbacher, Urs. 2007. “Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments.” *Experimental Economics* 10 (2): 171–78.

Herbst, Luisa, Kai A. Konrad, and Florian Morath. 2017. “Balance of Power and the Propensity of Conflict.” *Games and Economic Behavior* 103 (May): 168–84.

Jervis, Robert. 1976. *Perception and Misperception in International Politics*. Princeton, NJ: Princeton University Press.

Kimbrough, Erik O., Jared Rubin, Roman M. Sheremeta, and Timothy W. Shields. 2015. “Commitment Problems in Conflict Resolution.” *Journal of Economic Behavior & Organization* 112 (April): 33–45.

Kimbrough, Erik O., and Roman M. Sheremeta. 2013. “Side-Payments and the Costs of Conflict.” *International Journal of Industrial Organization* 31 (3): 278–86.

———. 2014. “Why Can’t We Be Friends? Entitlements and the Costs of Conflict.” *Journal of Peace Research* 51 (4): 487–500.

Kimbrough, Erik O., Roman M. Sheremeta, and Timothy W. Shields. 2014. “When Parity Promotes Peace: Resolving Conflict between Asymmetric Agents.” *Journal of Economic Behavior & Organization* 99 (March): 96–108.

Krainin, Colin. 2017. “Preventive War as a Result of Long-Term Shifts in Power.” *Political Science Research and Methods* 5 (1): 103–21.

Powell, Robert. 1999. *In the Shadow of Power: States and Strategies in International Politics*. Princeton, NJ: Princeton University Press.

———. 2004. “The Inefficient Use of Power: Costly Conflict with Complete Information.” *American Political Science Review* 98 (2): 231–41.

———. 2006. “War as a Commitment Problem.” *International Organization* 60 (1): 169–203.

Sanchez-Pages, Santiago. 2012. “Bargaining and Conflict with Incomplete Information.” In *Oxford Handbook of the Economics of Peace and Conflict*, edited by Michelle R Garfinkel and Stergios Skaperdas. Oxford University Press.

Schelling, Thomas C. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Tingley, Dustin H. 2011. “The Dark Side of the Future: An Experimental Test of Commitment Problems in Bargaining1.” *International Studies Quarterly* 55 (2): 521–44.

Walzer, Michael. 1977. *Just And Unjust Wars*. Basic Books.

Wolford, Scott, Dan Reiter, and Clifford J. Carrubba. 2011. “Information, Commitment, and War.” *Journal of Conflict Resolution* 55 (4): 556–79.

Appendix A. Experimental instructions (English translation)

[The PreventiveWar treatment; NoWar only differs in players' probabiliy of winning in stage 1, and thus is omitted here.]

Welcome! You are taking part in a decision making experiment. You have earned 15 RMB for showing up on time. In addition, you can earn more money in this experiment. The amount of money you earn will depend upon the decisions you make and on the decisions other people make. Your earnings in this experiment are expressed in EXPERIMENTAL CURRENCY UNITS, which we will refer to as ECUs. At the end of the experiment you will be paid IN CASH using a conversion rate of **1 RMB for every 5 ECUs** of earnings from the experiment (final payment will be rounded to the nearest 10 cents). Please do not communicate with each other during the experiment. If you have a question, feel free to raise your hand, and an experimenter will come to help you.

The experiment consists of 20 rounds. At the beginning of each round, you will be randomly matched with another participant in this room. You will be either a Person A or Person B. If you are Person A (or B), the other participant will be Person B (or A). Both the pairing and role will be randomly determined from round to round.

At the beginning of each round, each player receives an endowment of 5 ECUs. Each round has two stages. In each stage, Person A will decide how to split 10 ECUs between herself and Person B. Person B will then decide to either "accept" or "reject". In the following, we will explain their respective decisions in more detail.

<u>Stage 1</u>

Person A will move first and decide how to split 10 ECUs (in integer) between herself and Person B. Person A will also announce her plan of splitting another 10 ECUs in stage 2. But she does not have to execute this plan in stage 2.

After observing Person A's decision for stage 1 and plan for stage 2, Person B will move next and decide whether to "accept" or "reject". If he chooses to "accept", then he will receive the number of ECUs determined by Person A. Moreover, Person B will announce his plan of whether to "accept" or "reject" in response to each of the possible splitting decision made by Person A in stage 2. He also does not have to execute this plan in stage 2. After Person B has made both his decision for stage 1 and plan for stage 2, these decisions will be revealed to Person A. The pair will then proceed to stage 2.

If Person B chooses to "reject", the computer will determine who will receive a prize worth 20 ECUs. **Person A's probability of winning the prize is 20% whereas Person B's probability of winning is 80%.** In addition, each player has to pay 5 ECUs irrespective of the outcome of the competition (they do not need to pay this amount if Person B chooses to "accept"). There will be no stage 2 if the pair competes in stage 1.

<u>Stage 2</u>

Similar to stage 1, Person A will move first and decide how to split 10 ECUs (in integer) between herself and Person B.

Person B will move next and decide whether to "accept" or "reject". If he chooses to "accept", then he will receive the number of ECUs determined by Person A.

If Person B chooses to "reject", the ECUs received in stage 1 will be eliminated and the computer will determine who will receive a prize worth 20 ECUs. **Person A's probability of winning the prize is 70% whereas Person B's probability of winning is 30%.** In addition, each player has to pay 5 ECUs irrespective of the outcome of the competition (they do not need to pay this amount if Person B chooses to "accept").

Payoff

If Person B chooses to "accept" in both stages, the total ECUs (additional to endowment) Person B will receive in the round will be the sum of his shares determined by Person A in both stages. The total ECUs Person A will receive will be the sum of her shares determined by herself in both stages.

If Person B chooses to "reject" in either stage, each player's earnings in the round will NOT depend on A's decisions in either stage and will only depend on whether he/she wins the prize in that stage, subtracted by 5 ECUs. **Note that while Person B will have a higher probability of winning in stage 1, Person A will have a higher probability of winning in stage 2.**

Your total earnings for the experiment will be the sum of the earnings in all rounds.

This completes the instructions. Before we begin the experiment, to make sure that every participant understands the instructions, please answer several review questions on your screen.

[The RP-Commit treatment]

Welcome! You are taking part in a decision making experiment. You have earned 15 RMB for showing up on time. In addition, you can earn more money in this experiment. The amount of money you earn will depend upon the decisions you make and on the decisions other people make. Your earnings in this experiment are expressed in EXPERIMENTAL CURRENCY UNITS, which we will refer to as ECUs. At the end of the experiment you will be paid IN CASH using a conversion rate of **1 RMB for every 5 ECUs** of earnings from the experiment (final payment will be rounded to the nearest 10 cents). Please do not communicate with each other during the experiment. If you have a question, feel free to raise your hand, and an experimenter will come to help you.

The experiment consists of 20 rounds. At the beginning of each round, you will be randomly matched with another participant in this room. You will be either a Person A or Person B. If you are Person A (or B), the other participant will be Person B (or A). Both the pairing and role will be randomly determined from round to round.

At the beginning of each round, each player receives an endowment of 5 ECUs. Each round has two stages. In each stage, Person A will decide how to split 10 ECUs between herself and Person B. Person B will then decide to either "accept" or "reject". In the following, we will explain their respective decisions in more detail.

<u>Stage 1</u>

Person A will move first and decide how to split 10 ECUs (in integer) between herself and Person B. Person A will also announce her plan of splitting another 10 ECUs in stage 2. And she will have to execute this plan in stage 2.

After observing Person A's decision for stage 1 and plan for stage 2, Person B will move next and decide whether to "accept" or "reject". If he chooses to "accept", then he will receive the number of ECUs determined by Person A. After Person B has made his decision for stage 1, it will be revealed to Person A. The pair will then proceed to stage 2.

If Person B chooses to "reject", the computer will determine who will receive a prize worth 20 ECUs. **Person A's probability of winning the prize is 20% whereas Person B's probability of winning is 80%.** In addition, each player has to pay 5 ECUs irrespective of the outcome of the competition (they do not need to pay this amount if Person B chooses to "accept"). There will be no stage 2 if the pair competes in stage 1.

<u>Stage 2</u>

Similar to stage 1, Person A will move first and decide how to split 10 ECUs (in integer) between herself and Person B. And she must split according to the plan made in stage 1 (the computer will automatically implement the decision according to Person A's plan).

Person B will move next and decide whether to "accept" or "reject". If he chooses to "accept", then he will receive the number of ECUs determined by Person A.

If Person B chooses to “reject”, the ECUs received in stage 1 will be eliminated and the computer will determine who will receive a prize worth 20 ECUs. **Person A’s probability of winning the prize is 70% whereas Person B’s probability of winning is 30%.** In addition, each player has to pay 5 ECUs irrespective of the outcome of the competition (they do not need to pay this amount if Person B chooses to “accept”).

Payoff

If Person B chooses to “accept” in both stages, the total ECUs (additional to endowment) Person B will receive in the round will be the sum of his shares determined by Person A in both stages. The total ECUs Person A will receive will be the sum of her shares determined by herself in both stages.

If Person B chooses to “reject” in either stage, each player’s earnings in the round will NOT depend on A’s decisions in either stage and will only depend on whether he/she wins the prize in that stage, subtracted by 5 ECUs. **Note that while Person B will have a higher probability of winning in stage 1, Person A will have a higher probability of winning in stage 2.**

Your total earnings for the experiment will be the sum of the earnings in all rounds.

This completes the instructions. Before we begin the experiment, to make sure that every participant understands the instructions, please answer several review questions on your screen.

Appendix B. Additional tables and figures

Table B1: Random effects regression of final payoff (main treatments)

Dependent Variable: Final payoff	(1)	(2)
β_0: NoWar	2.359***	
	(0.783)	
β_1: RP-Commit		-0.049
		(0.683)
β_2: DP	6.499***	6.489***
	(0.737)	(0.729)
β_3: NoWar × DP	-3.405***	
	(1.070)	
β_4: RP-Commit × DP		1.692*
		(1.000)
Constant	3.682***	3.687***
	(0.525)	(0.522)
H0: $\beta_0 + \beta_3 = 0$	$p = 0.004$	
H0: $\beta_1 + \beta_4 = 0$		$p = 0.000$
Observations	3200	3000
Clusters	16	15

*Notes: Standard errors clustered at the session level are in parentheses. RP in PreventiveWar serves as the benchmark. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Table B2: Random effects regression of final payoff (Chat and Repeat treatments)

Dependent Variable: Final payoff	(1)	(2)
β_0: Chat	2.747***	
	(0.802)	
β_1: Repeat		1.977***
		(0.768)
β_2: DP	6.489***	6.524***
	(0.726)	(0.746)
β_3: Chat × DP	-2.557**	
	(1.081)	
β_4: Repeat × DP		-2.054*
		(1.108)
Constant	3.687***	3.669***
	(0.520)	(0.528)
H0: $\beta_0 + \beta_3 = 0$	$p = 0.586$	
H0: $\beta_1 + \beta_4 = 0$		$p = 0.865$
Observations	3200	3200
Clusters	16	48

*Notes: Standard errors are clustered at the session level for PreventiveWar and Chat, and at the group level for Repeat. RP in PreventiveWar serves as the benchmark. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Table B3: Random effects regression of final payoff (HighCost and LowCost treatments)

Dependent Variable: Final payoff	(1)	(2)
β_0: HighCost	2.133**	
	(0.926)	
β_1: LowCost		1.161*
		(0.614)
β_2: DP	6.496***	6.494***
	(0.742)	(0.740)
β_3: HighCost × DP	-3.304***	
	(1.068)	
β_4: LowCost × DP		2.335***
		(0.909)
Constant	3.683***	3.684***
	(0.321)	(0.528)
H0: $\beta_0 + \beta_3 = 0$	$p = 0.000$	
H0: $\beta_1 + \beta_4 = 0$		$p = 0.000$
Observations	2400	2400
Clusters	12	12

Notes: Standard errors are clustered at the session level. RP in PreventiveWar serves as the benchmark. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B4: Probit regressions of preventive wars in the Repeat treatment

Dependent Variable: fight in stage 1 = 1	Average marginal effects (1)	(2)
L1. preventive war	0.370***	0.313***
	(0.065)	(0.049)
L2. preventive war		0.297***
		(0.053)
1[RP's stage 1 offer < 9]	0.043	0.059
	(0.058)	(0.050)
Round	-0.004	-0.002***
	(0.003)	(0.002)
Observations	760	720
Clusters	40	40

*Notes: Standard errors are clustered at the group level. Average marginal effects are reported. *** p < 0.01.*

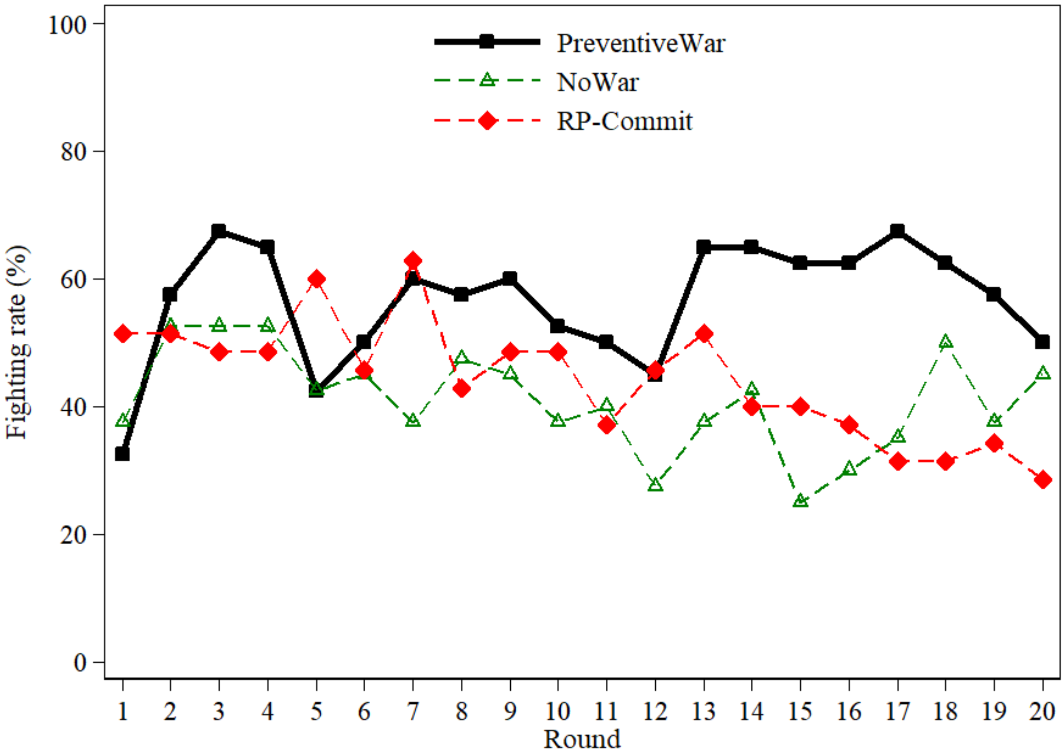

Figure B1: The frequency of preventive wars across rounds (main treatments)

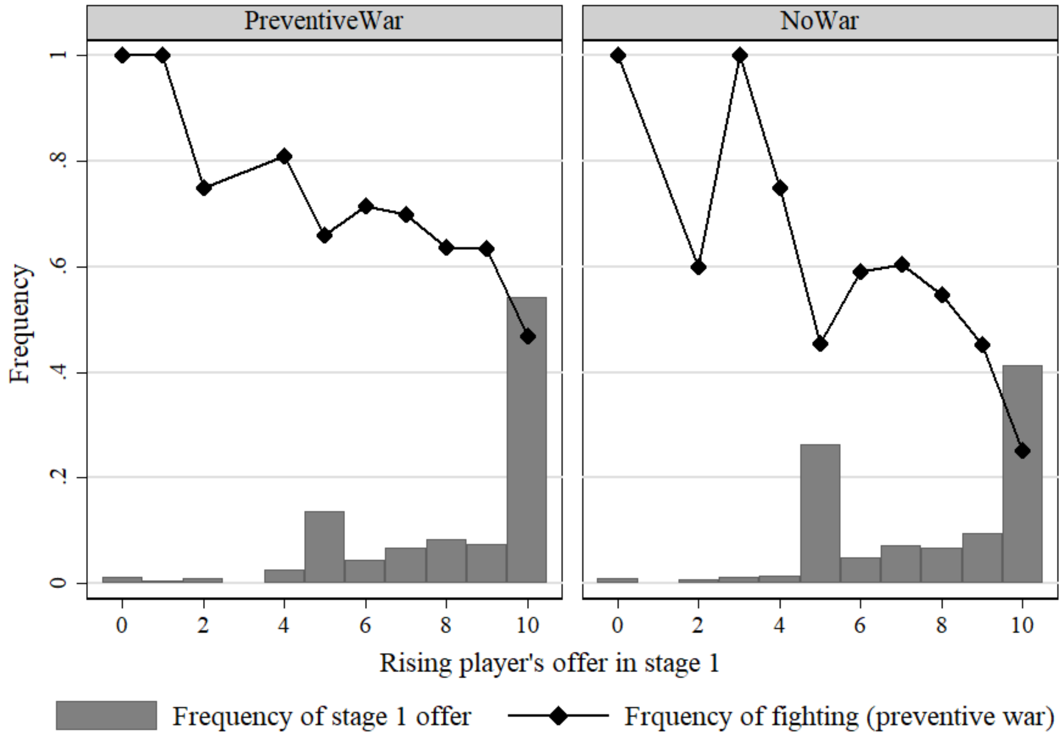

Figure B2: The frequency of preventive wars conditional on RP's stage 1 offer (main treatments)

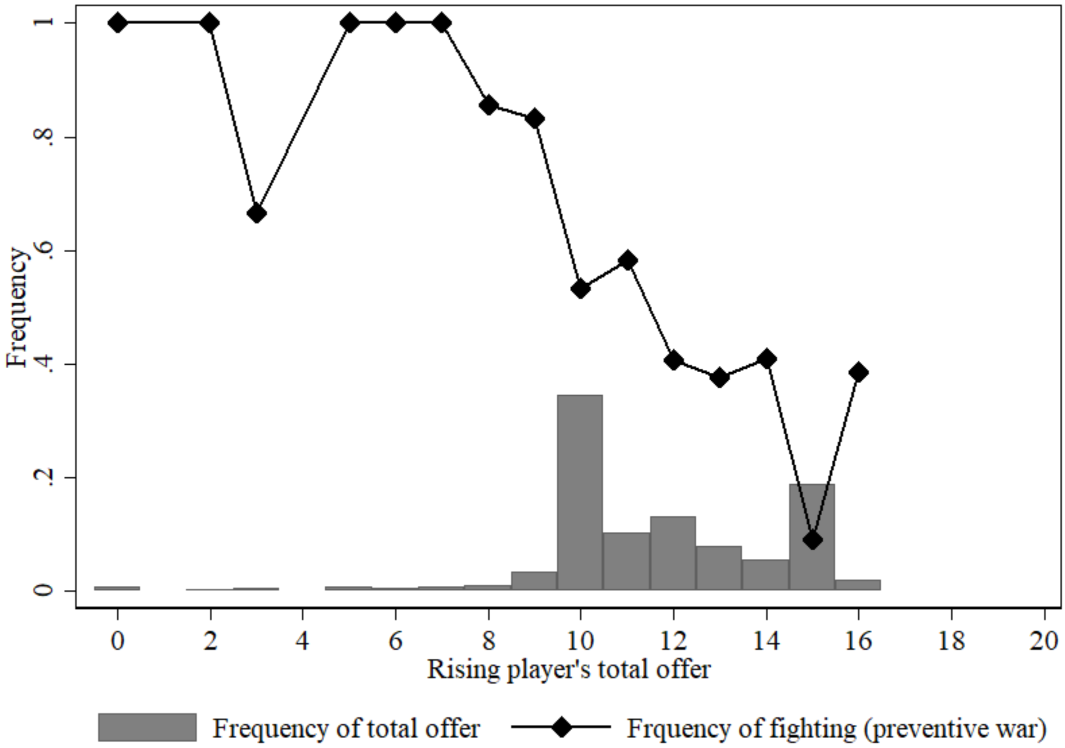

Figure B3: The frequency of preventive wars conditional on RP's total offer (RP-Commit treatment)

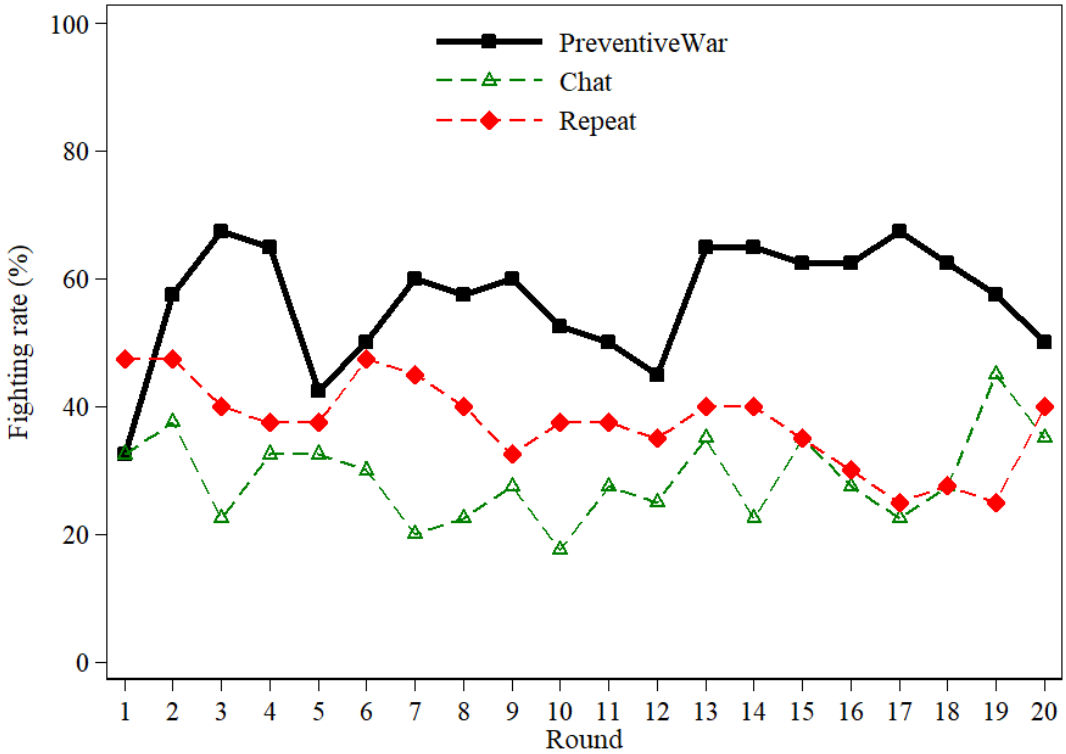

Figure B4: The frequency of preventive wars across rounds (Chat and Repeat treatments)

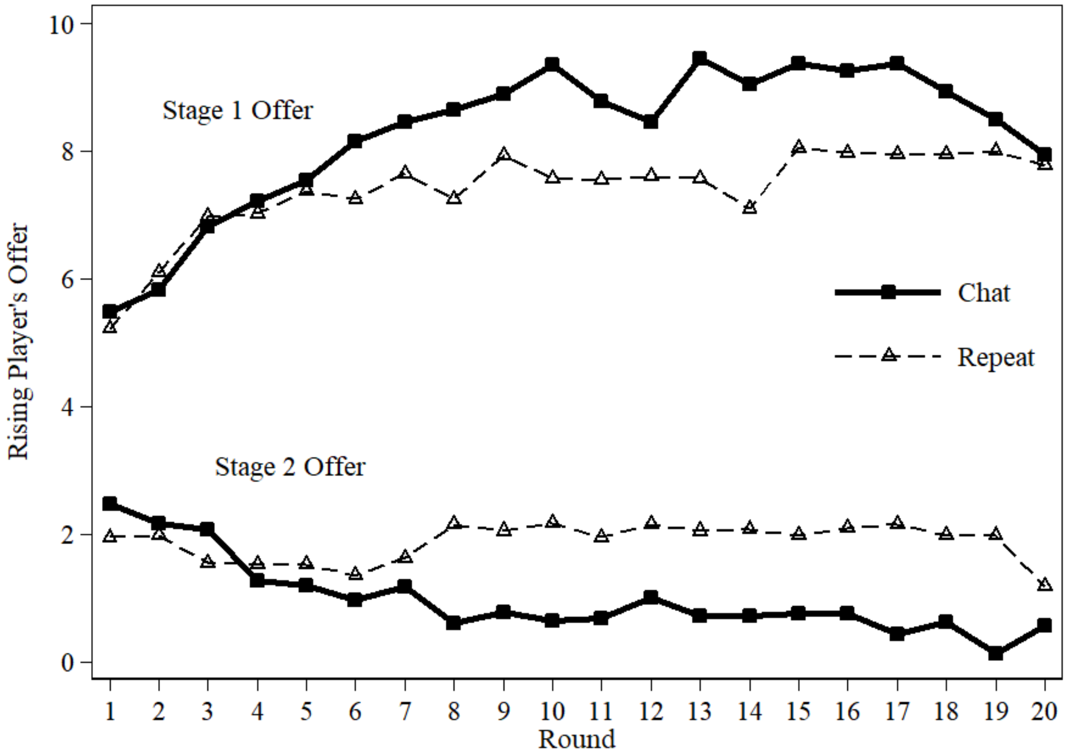

Figure B5: Rising players' offers in both stages over round (Chat and Repeat treatments)

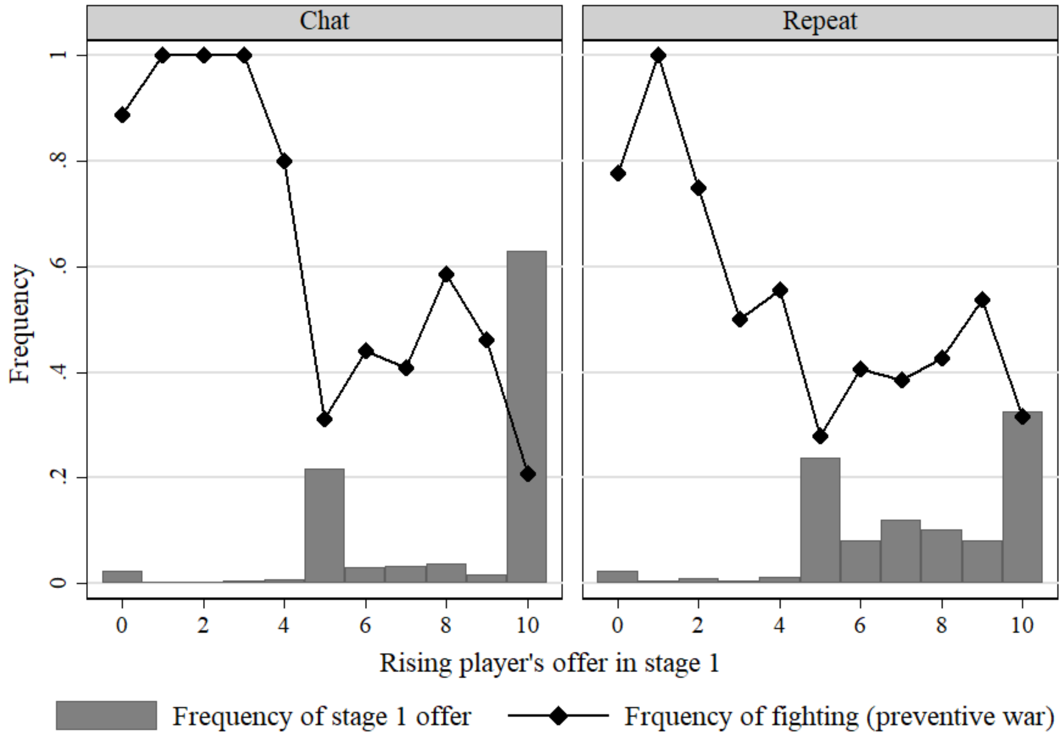

Figure B6: The frequency of preventive wars conditional on RP's stage 1 offer (Chat and Repeat treatments)

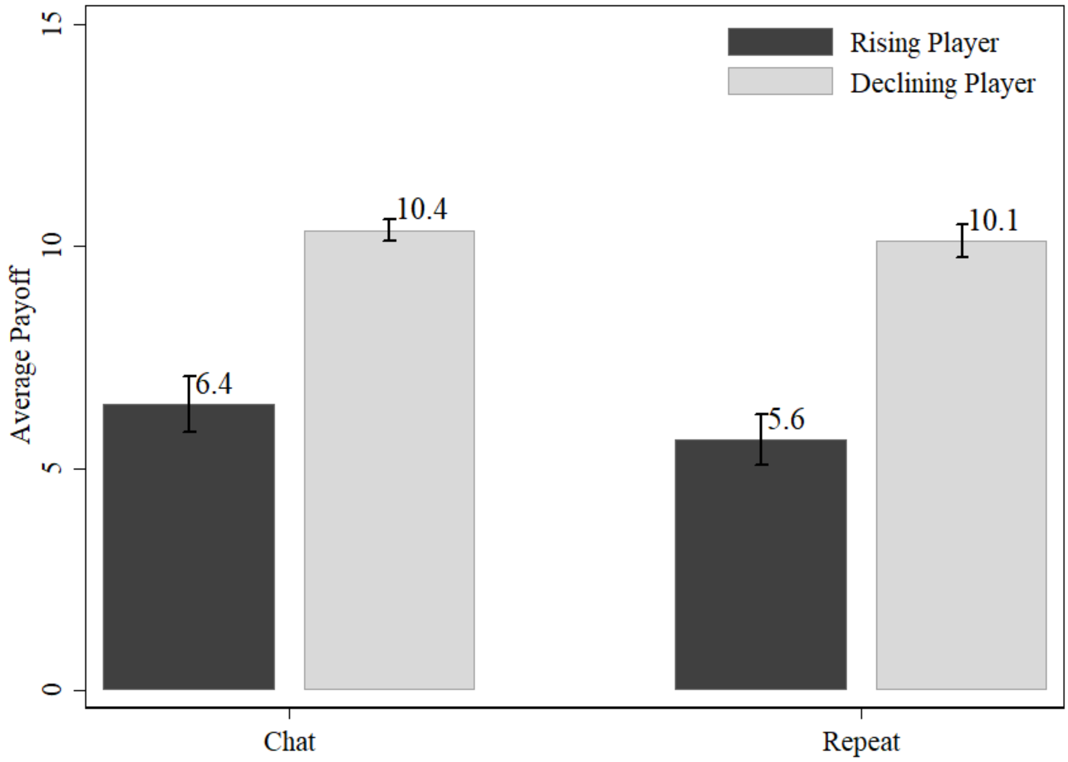

Figure B7: RP's and DP's average payoffs (excluding 5-token endowment; Chat and Repeat treatments)

Note: Error bars represent one standard error of the mean clustered at the session level.

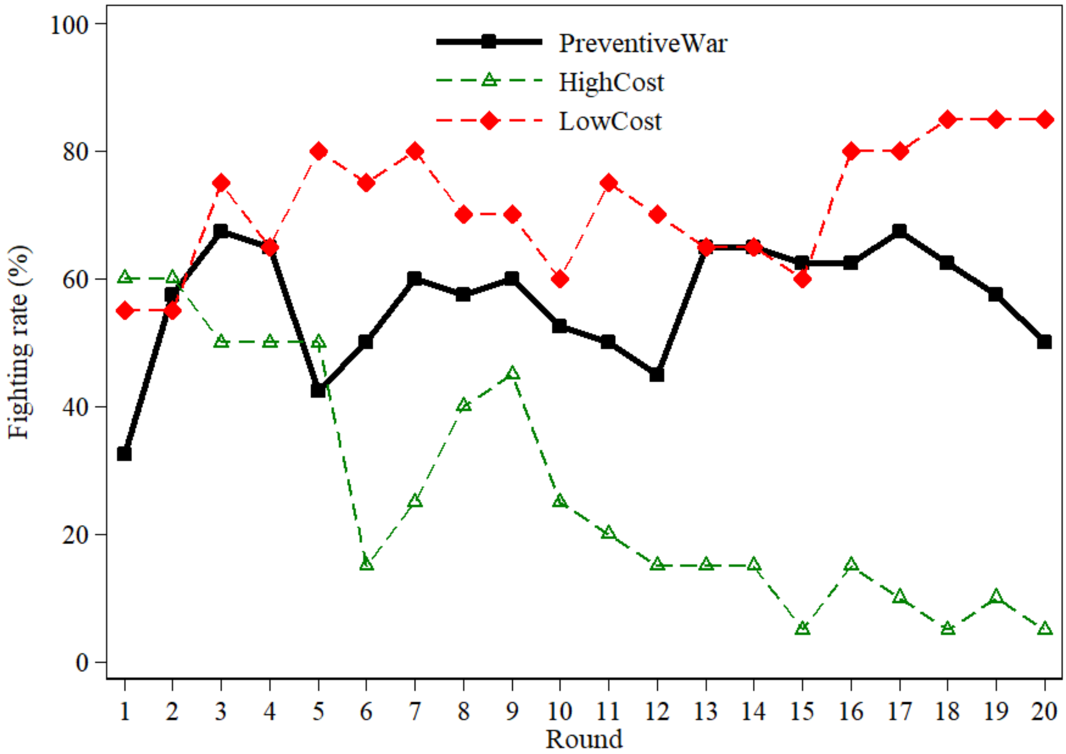

Figure B8: The frequency of preventive wars across rounds (HighCost and LowCost treatments)

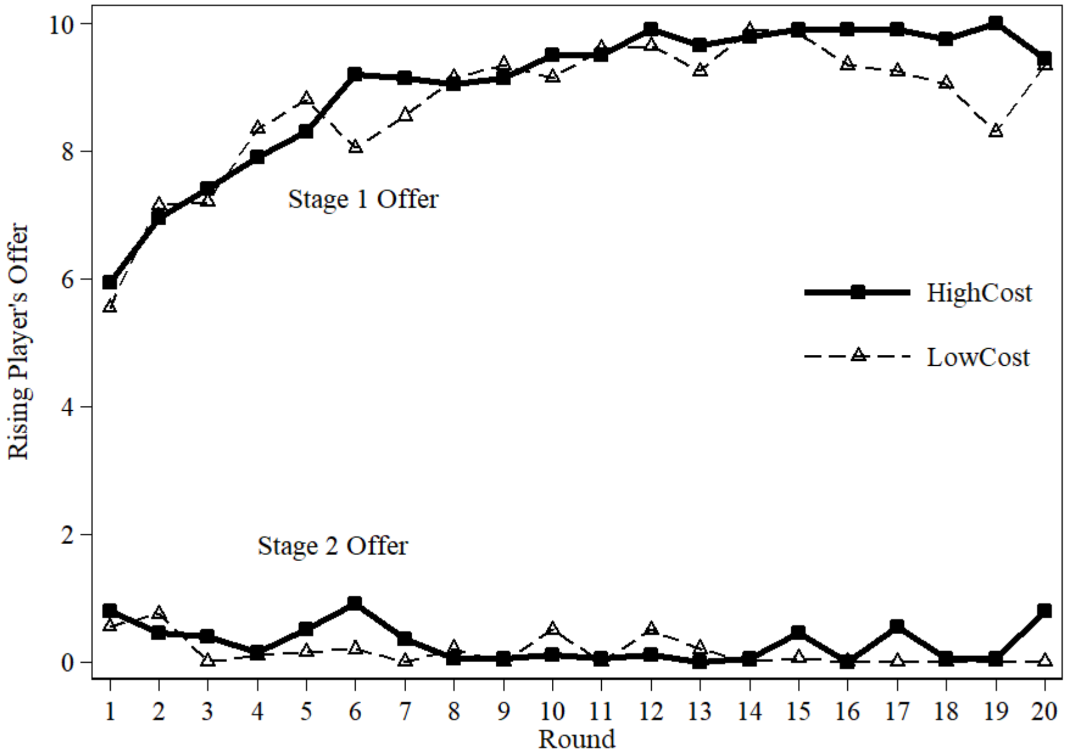

Figure B9: Rising players' offers in both stages over round (HighCost and LowCost treatments)

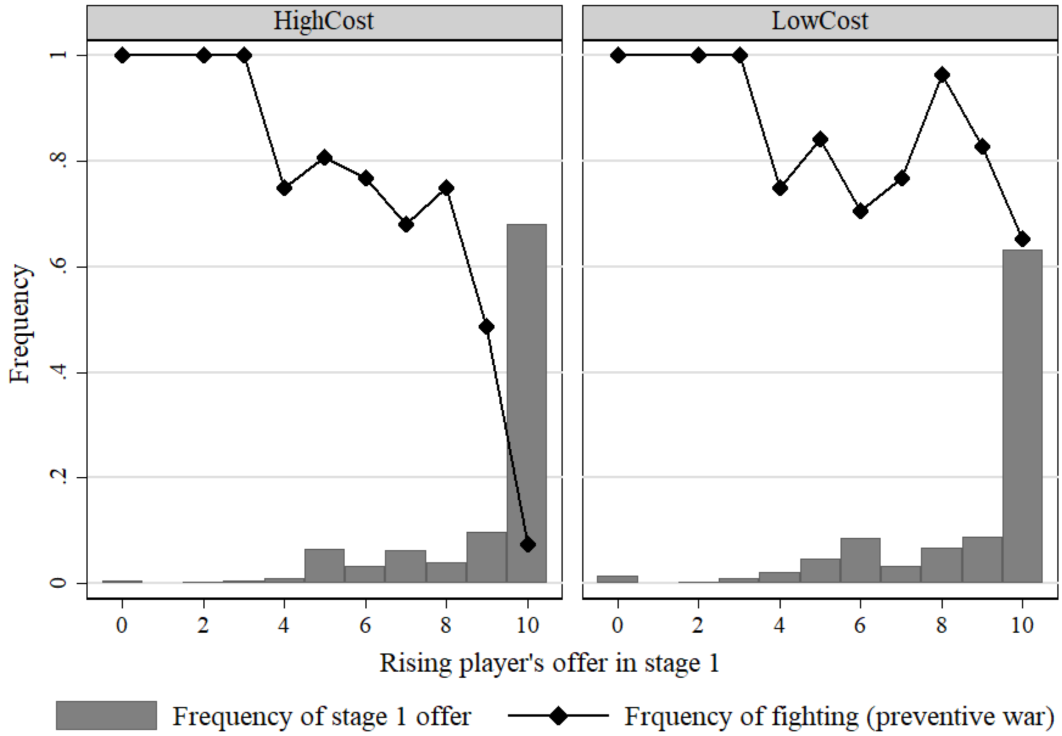

Figure B10: The frequency of preventive wars conditional on RP’s stage 1 offer (HighCost and LowCost treatments)

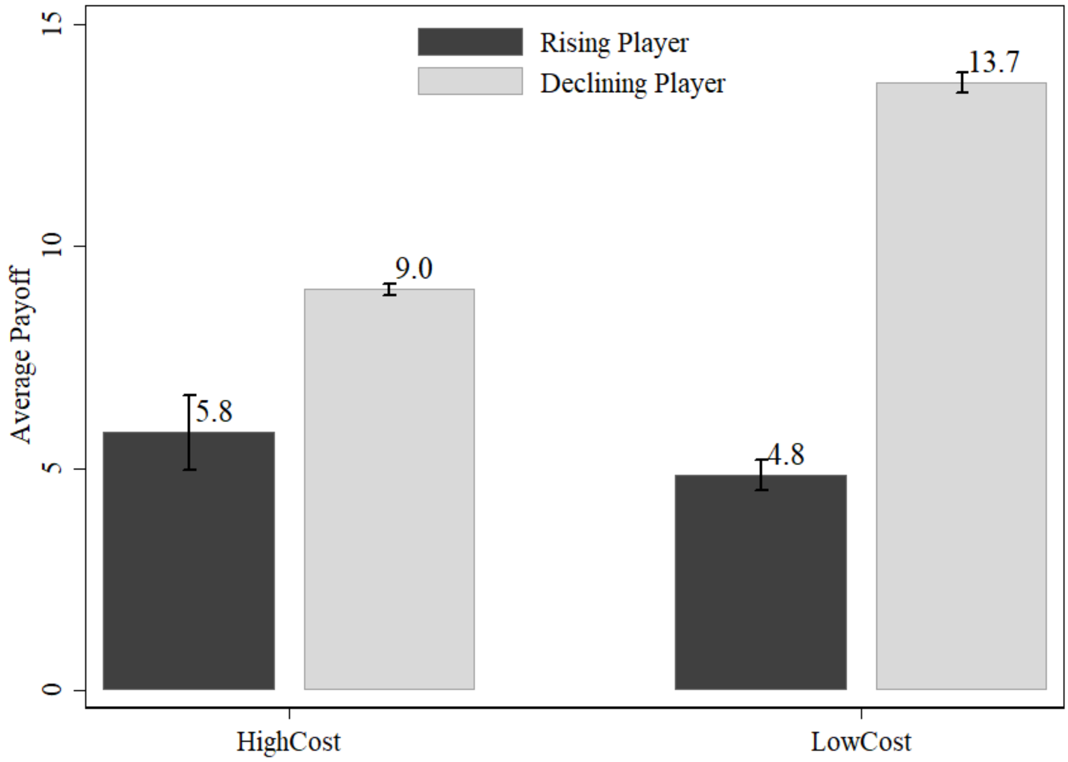

Figure B11: RP's and DP's average payoffs (excluding 5-token endowment; HighCost and LowCost treatments)

Note: Error bars represent one standard error of the mean clustered at the session level.