Discussion Paper Series – CRC TR 224

Discussion Paper No. 559
Project A 01

Bargaining Under the Threat of a Nuclear Option

Franziska Heinicke[1]
Wladislaw Mill[2]
Henrik Orzen[3]

June 2024

[1]University of Mannheim, Email: Franziheinicke@gmail.com
[2]University of Mannheim, Email: mill@uni-mannheim.de
[3] University of Mannheim, Email: henrik.orzen@uni-mannheim.de

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Collaborative Research Center Transregio 224 - www.crctr224.de
Rheinische Friedrich-Wilhelms-Universität Bonn - Universität Mannheim

Bargaining under the threat of a nuclear option*

Franziska Heinicke[†] Wladislaw Mill[‡] Henrik Orzen[§]

This paper addresses bargaining with a nuclear option. Players with access to such an option have the power to cause enormous damage to their negotiation partners. Figurative nuclear options are available in many important real-world settings and, being the ultimate threat, are often seen as effective in putting maximal pressure on the other party and as possibly efficiency-improving. On the other hand, since going nuclear is typically also very costly to the nuclear-option holder herself, the credibility of a nuclear threat may be questionable. We report the results from unstructured one-shot bargaining experiments and examine to what extent a nuclear option increases bargaining power, makes agreements more likely, and affects efficiency. We find that nuclear-option holders do not generally benefit while the other party is worse off compared to a baseline setting, particularly when the other party is intrinsically—i.e., save for the nuclear threat itself—in a strong position. Furthermore, the nuclear option increases the number of negotiations that end in agreements that are not efficiency-improving. Thus, the presence of a nuclear option in our bargaining setting is overall detrimental.

Keywords: Nuclear option; Power asymmetry; Bargaining; Experiment
JEL-Code: C91, D91, D01, C78, J52.

*We thank Sven Höppner, Timo Promann, Ali Seyhun Saral, Cornelius Schneider, Duk Kim, Olexander Nikolaychuk, Christian König, Marius Alt, Joerg Oechsler, Christoph Vanberg, Peter Doersch, and Emin Karagözoğlu for valuable feedback. We appreciate comments from the participants of the Annual Meeting of the Standing Field Committee of the German Economic Association 2021, the Meeting of the Economic Science Association 2022, the 6th BEET Workshop, Bargaining: Experiments, Empirics, and Theory 2022, the Meeting of the German Economic Association 2022, the Mannheim-ZEW-Seminar and the HeiKaMax. The experiment is preregistered at https://aspredicted.org #41190 and #89754. IRB approval at https://gfew.de/ethik/mfptmAoo. Support by the German Research Foundation (DFG) through CRC TR 224 (Project A01) is gratefully acknowledged.

[†]University of Mannheim, Department of Economics. Franziheinicke@gmail.com
[‡]University of Mannheim, Department of Economics. mill@uni-mannheim.de
[§]University of Mannheim, Department of Economics. henrik.orzen@uni-mannheim.de

1 Introduction

This paper addresses bargaining with a nuclear option. The possibility of 'going nuclear' is relevant in a range of important settings. Obviously, the cold war era was defined by the principle of mutually assured destruction via the nuclear option in a literal sense. In many other contexts, figurative nuclear options are debated. Commentators refer to China as having nuclear options in trade negotiations with the United States, as it could, for example, stop exporting rare earth minerals to the U.S. or sell much of its U.S. government debt, measures that would cause considerable damage to the American economy. Hardliners in the Brexit negotiations repeatedly brought up the possibility of a 'nuclear' no-deal scenario as a threat against the European Union. In the U.S. senate, the majority leader's power to trigger a process that effectively replaces the 60-vote threshold required to end a debate by simple majority has traditionally been seen as an almost unthinkable very last resort to force through a particular decision, a move so drastic that the 'nuclear' label stuck with it.

In this paper, we consider a general notion of a nuclear option as an action a negotiator might take following negotiation failure. Executing this action makes a change for the worse compared to the 'normal' disagreement point for both parties. For example, consider political bargaining between two coalition partners who both want to change a particular policy but have different views on what to replace it with. If they find themselves unable to come up with a feasible compromise, the result will be no new regulation, an unsatisfactory status quo. However, there is also the possibility of a much more drastic move: ending the coalition altogether. This would be the nuclear option. Or consider intra-household bargaining, with the spouses negotiating over how to share the burdens of parenthood. In this case, a non-agreement might mean not having children, whereas the nuclear option is a divorce. Other examples for nuclear options in bargaining include whistleblowing after a disagreement in an illegal cartel, the threat of dismissal in a wage negotiation, or a war of roses after an unhappy marriage.

The motive for bringing up the possibility of a nuclear option in a situation characterized by conflicting interests is typically to put pressure on the opposing party and/or to deter that party from taking particular actions. It is the ultimate threat. However, it is typically also a double-edged sword. If China really were to attack the U.S. economy in the way described above, it would hurt its own interests immensely as well. Similarly, a political party implementing fundamental change in the Senate rules to pass legislation runs the risk of experiencing severe regret in the future after a majority change. A no-deal outcome in the Brexit negotiations would have been extremely damaging to businesses and people on both sides of the channel; coalition partners who are unable to continue to govern together risk losing public approval; and a spouse is likely to experience serious emotional and financial costs in case of a divorce.

Given that going nuclear has severe consequences not only for the intended target but also for the party that pulls the trigger, basic questions about the credibility and efficacy of a nuclear threat arise. Consider a generic two-party bargaining situation in which one side has access to a nuclear option while the other does not. Does this give the nuclear-option holder any additional bargaining power? Is the looming nuclear threat efficiency-enhancing by making the negotiation partners more inclined to come to a mutually beneficial agreement? But what if the situation is one in which the status quo is in fact more efficient than any new deal could be? Can the nuclear-option holder then effectively force the other party into an unfavorable and inefficient 'agreement'?

In this paper, we empirically investigate the effects of the presence of a nuclear option in such negotiations. Given that observational data would likely come with caveats of reversed causality, omitted variable bias, and the inherent challenge of identifying nuclear options in naturally-occurring bargaining scenarios, we employ experimental methods that allow us to cleanly vary the relevant situational parameters in order to establish causality.

In our Baseline treatment, two players negotiate about the division of a fixed pie in an unstructured bargaining setting (Lozano et al., 2024; Tremewan and Vanberg, 2016; Anbarci and Feltovich, 2013; Karagözoğlu, 2019). They do so by sending proposals to each other in real time, and at any point they can overwrite a previous proposal by submitting a new one. No other form of communication is permitted. As soon as one side accepts their partner's current proposal, the negotiation phase ends and the agreed proposal is implemented. If no proposal is accepted within the given time frame, the negotiation ends as well, but in this case the players receive predefined and fixed disagreement payoffs that are publicly announced prior to the negotiation.

We compare our Baseline treatment to two treatments that introduce a nuclear option for one of the two negotiators (player B). The Nuke treatment is identical to the baseline in every way, except that player B can choose to trigger a nuclear option when a negotiation ends in disagreement. The Threat treatment offers the same additional choice, but player B must first *activate* the nuclear option in an additional step. Importantly, player B can activate the nuclear option at any time during or after the negotiation. Because the activation is visible to player A, player B might use this step strategically as a signal.

If player B does go nuclear, the regular disagreement payoffs are replaced by the nuclear outcome, which causes devastating damage to player A's payoff but is also very harmful to B herself. The interaction between the two parties in our experiment is one-shot to capture the credibility problem inherent to nuclear options. Hence, a selfish rational payoff maximizer would never go nuclear and, insofar as her negotiating partner assumes her to be selfish and rational, the existence of the nuclear option would be completely inconsequential. Nevertheless, we suspect that the salience and the credibility of the nuclear threat will in fact depend on the situational parameters under which the negotiation takes place. In our experiment, we therefore implement—in all treatments—variations in the setting by considering a range of scenarios with symmetric and asymmetric (regular) disagreement points. Thus, our design explores the effects of differences in relative bargaining power in the negotiation in terms of how much each party has to lose from a disagreement or unfavorable agreement or from the nuclear outcome itself. Furthermore, some of the scenarios vary in terms of the relative efficiency of outcomes. That is, in addition to the conventional bargaining situation in which coming to an agreement leads to an overall welfare improvement (i.e., the sum of disagreement payoffs is smaller than the pie), we consider settings in which an agreement is efficiency-neutral or even *reduces* overall welfare such that the players should 'agree to disagree'. Such scenarios are rarely studied in the literature but they do present an interesting case to investigate the broader consequences of a nuclear option. Examples for settings like this include dysfunctional relationships (where an amicable parting of ways may be better than a perpetual search for ways of making it work), business mergers (where the merging of two companies might lead to redundancies and a loss of total value, rather than creating synergies), or inefficient contracts (that should be dissolved when the seller's performance cost exceeds the buyer's value). If the nuclear option does create power asymmetry, the more powerful party may then be able to instill in the other party a sense of having "no choice" but to enter a particular agreement, even when this is inefficient. An extreme example can be seen in the film *The Godfather* where members of the Corleone family famously make offers that the other side "cannot refuse".

The negotiation procedure itself is rather neutral and symmetric, without the bargaining structure favoring or disfavoring either party, save for the nuclear option itself. Hence, we avoid interaction effects between the nuclear option and power asymmetries that would, for example, be created by an imposed sequence of moves.[1] By extending the action space of only one player, we keep the changes between treatments at a minimum and prevent possible interaction effects between two nuclear-option holders (such as revenge cycles), which would be hard to interpret. Furthermore, it can be argued that while in real life nuclear options might often in principle be available to both negotiating parties, typically only one party's nuclear option is at the center of the attention in concrete examples, and this asymmetry is reflected in our design.[2] Our experiment examines to what extent a nuclear-option holder feels encouraged to engage in tougher negotiation behavior and to what extent the other player engages in appeasement through concessions during the negotiation, and whether all of this has an effect on efficiency.

Our results show that the introduction of the nuclear option systematically decreases the joint payoffs of the two negotiation partners. The player without the nuclear option earns considerably less than in the baseline setting (where no nuclear option is present) and the player with the nuclear option does not generally benefit from it. Across a range of different disagreement points, player B earns slightly more in some situations but experiences overall no significant difference to the Baseline treatment. This can be explained by two changes in the negotiation outcomes. First, even though the nuclear option increases the rate of agreements, this mainly occurs in cases where agreements are undesirable from a social planner's point of view. Thus, players agree in situations where they should not, and this leaves little room for improvements in player B's payoff. Second, there are cases in which the nuclear option is not just a looming threat but is actually triggered, with detrimental effects for the payoffs of both players. These results are even more pronounced in the Threat treatment. Overall, our results suggest that access to a nuclear option does not pay off and lowers efficiency.

With our findings, we speak to two strands of literature. First, we contribute to the theoretical bargaining literature (Miller and Watson, 2013; Deneckere and Liang, 2006; Battaglini, 2021; Ravid, 2020; Larsen, 2020) by showing that a nuclear option, a clearly dominated outside alternative, can have an impact on bargaining behavior. We conceptualize this in a simple model based on spiteful preferences and impulsiveness. Second, and more importantly, we contribute to the empirical literature on negotiations (Doepke and Kindermann, 2019; Silveira, 2017; Ali et al., 2023; Exley et al., 2020; Kocher et al., 2017; Merkel and Vanberg, 2023; Miller and Vanberg, 2015). A priori, one might expect that nuclear options have the potential to be beneficial: Just like it should become easier to sustain cooperation in indefinitely-repeated social dilemma games when the non-cooperative stage-game equilibrium is unattractive, it seems plausible that the

[1] We also prefer an unstructured setting over structured bargaining because it is arguably a more natural representation of real-world negotiations and provides richer data to study the effects of a nuclear option.

[2] For instance, in the context of an unhappy marriage, one partner is often closer than the other to file for divorce, despite both having the option available. Similarly, the damage from dissolving a coalition will in most cases not be equally bad for all coalition partners and this leads to differences in their inclination to exit.

looming threat of a nuclear outcome might lead to more disciplined, efficiency-enhancing negotiations. However, our results point in the opposite direction. First, there is a danger that the nuclear option is actually triggered. Second, there is no increase in the number of efficiency-improving agreements. Third, inefficient agreements that are struck for the sake of avoiding a conflict with the nuclear-option holder do become more likely. Ultimately, not even the nuclear option holders are better off than in our baseline treatment. They can achieve detectable improvements only in a specific type of situation: When they encounter negotiation partners who employ a particularly tough bargaining strategy, then activating the nuclear option can make these partners more compromising.

The remainder of the paper is organized as follows. In Section 2, we discuss the related literature. Our experimental design is described in Section 3. Section 4 provides some theoretical considerations. We report our results in Section 5. Finally, Section 6 concludes.

2 Related literature

We are not aware of previous research that examines whether and how the presence of a nuclear option changes bargaining behavior and outcomes. However, the situation that we study is related to some settings that have been investigated in the existing bargaining literature.

In ultimatum game experiments, responders have the option to reject take-it-or-leave-it offers that they receive from a proposer (for reviews, see Camerer and Thaler, 1995; Oosterbeek et al., 2004; Güth, 1995). Note, however, that rejecting an offer in the ultimatum game does not constitute a nuclear option as studied here. While the rejection of an offer means foregoing a monetary payoff, it only implements the status quo—it is the refusal to accept a Pareto-improving transaction. This option is available even in our Baseline treatment. In contrast, triggering the nuclear option worsens the situation for both agents beyond the status quo. The ultimatum game also differs from our setting in that it is characterized by an ex-ante inherently asymmetric allocation of bargaining power. The findings from ultimatum bargaining experiments are nevertheless informative by showing that many people are willing to destroy resources, including their own payoff, if they are unhappy about the bargaining process or the distribution of outcomes.

Our study also relates to power-to-take games (Bosman and Van Winden, 2002; Bosman et al., 2005; Galeotti, 2015) and dictator games with punishment (Andreoni et al., 2003). In both of these, a dictator makes a decision that affects the payoff of a receiver who can subsequently punish the dictator at a cost. Similar to the ultimatum game, empirical evidence from these games show that receivers are willing to engage in costly punishment if treated unfairly (Bosman and Van Winden, 2002; Bosman et al., 2005; Galeotti, 2015; Andreoni et al., 2003). However, unlike in our setting, receivers in these extremely asymmetric situations have no influence on whether the dictator's decision is carried out—they can only make small adjustments to the outcome imposed by the dictator. In our experiment, instead, there is a negotiation in which both sides play a part in determining the initial outcome, which then might be annulled by the nuclear option.

Furthermore, our paper is related to the literature on money burning and bargaining, where one agent has the ability to destroy the potential surplus from the interaction between both agents (Manzini, 1999; Avery and Zemsky, 1994). The difference to our setting is that this literature focuses on repeated or sequential interactions where money burning can be used to enforce more generous offers from the other party in the future. In contrast, we are interested in one-shot situations where no future gains can be secured by using the nuclear option. Thus, in our setting, there is no strategic reason to trigger the nuclear option.

We study the impact of a nuclear option in an unstructured bargaining setting. A central insight of the literature on unstructured bargaining behavior is that outcomes depend on the relative bargaining positions of the agents involved.[3] The importance of bargaining power was already pointed out by Edgeworth (1881) and later formalized in axiomatic approaches to bargaining solutions (e.g., Nash Jr, 1950; Kalai and Smorodinsky, 1975). Bargaining outcomes often depend on power asymmetries in the specific bargaining environment. Existing experimental evidence on unstructured bargaining shows that outcomes are sensitive to various factors. For example, agents may enter negotiations with conflicting beliefs over how much they are entitled to in a specific setting. An agent who is, or appears to be, in a favored position may treat this position as a right to a larger share of the pie (Hoffman and Spitzer, 1985; Karagözoğlu and Riedl, 2014; Feltovich, 2019; Embrey et al., 2021). The findings in this literature suggest that bargaining power resulting from asymmetric beliefs over entitlements indeed allows agents with higher entitlements to secure higher payoffs. This holds for asymmetries that effectively change the game, such as different disagreement points (Hoffman and Spitzer, 1985; Anbarci and Feltovich, 2013, 2018) or information asymmetries (Camerer et al., 2018), but even apparently payoff-irrelevant asymmetries can affect bargaining outcomes. For example, relative or absolute performance in a preceding task that has no bearing on the players' options or their action space in the bargaining game has been shown to increase the bargaining power of high performers (Banerjee, 2020; Karagözoğlu and Riedl, 2014; Gächter and Riedl, 2005). Similarly, unrelated reference points that favor one player have been shown to shift bargaining outcomes (Bolton and Karagözoğlu, 2016), highlighting the sensitivity of bargaining outcomes to small changes in (perceived) bargaining power.

In summary, this literature demonstrates that power asymmetries—due to payoff-relevant or payoff-irrelevant factors—affect players' perceived entitlements, which in turn changes bargaining power and the corresponding outcomes. In line with this literature, we conjecture that the nuclear option may create a sense of power asymmetry between the negotiators. However, our setting differs from the existing research in that the nuclear option is self-harming, leaving the resulting effect ambiguous.

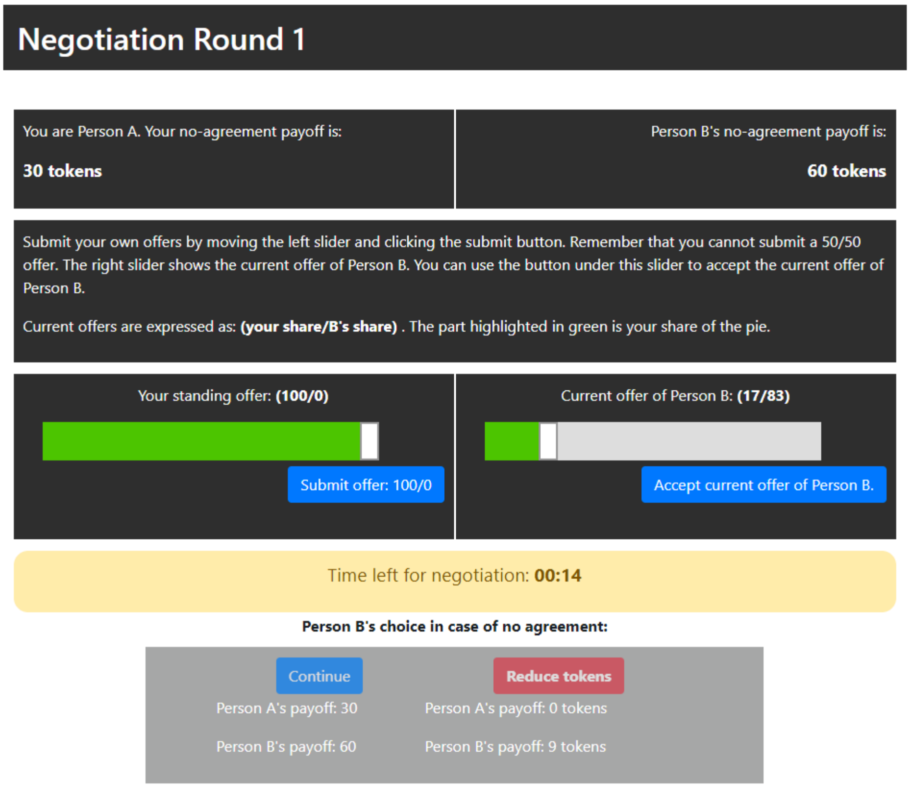
To the best of our knowledge, our study is the first to examine the effect of a nuclear option on bargaining behavior and outcomes and potential mechanisms behind this effect. We contribute to the debate on real-world figurative nuclear options and their consequences, and extend the literature on bargaining behavior. In particular, we examine power asymmetries in unstructured bargaining, which thus far has focused on the effect of unequal entitlements but has not allowed for an additional dominated choice.

[3] The economic literature on bargaining is vast and a comprehensive review is beyond the scope of this paper. We focus on contributions closely related to our study. For reviews see, e.g., Camerer (2011, pp.151-198), Roth (1995) or Thompson et al. (2010).

3 Experimental design

The goals of this paper are to establish whether having command over a nuclear option increases one's bargaining power in a negotiation, whether the presence of a nuclear option makes agreements more likely, and whether and how a nuclear option impacts efficiency. To achieve this, we compare experimental treatments with and without a nuclear option. Furthermore, we systematically vary the non-nuclear disagreement payoffs to study both symmetric and asymmetric bargaining situations and to examine scenarios that vary in terms of whether reaching an agreement is, relative to the disagreement point, efficiency-enhancing, efficiency-deteriorating, or efficiency-neutral.[4]

Figure 1: Decision screen (Nuke treatment)



3.1 Baseline treatment

Our setting is as follows. Two players, A and B, bargain over how to split a 100-token pie. In case of an agreement, the pie is split according to the suggested division. In case of no agreement, player A obtains a payoff $d_A > 0$ and player B obtains a payoff $d_B > 0$. These disagreement payoffs are public knowledge.

Players negotiate anonymously and in real time over the Internet, without any possibility of communicating beyond sending proposals about how to split the pie to the other party. To submit such a proposal, a player uses a slider and a button on the left-hand side of the computer interface, as depicted in Figure 1. By moving the slider's

[4] This experiment is pre-registered at https://aspredicted.org#41190. The Threat treatment was added later and is pre-registered at #89754. The deviations from the preregistration are discussed in Online Appendix E.

thumb to the left or to the right, the player specifies her proposal, which is shown on the screen as an integer value between 0 and 100. This value, also visually represented as a green area on the slider itself, indicates what share of the pie the player wishes to keep for herself. To actually send the proposal to her negotiation partner, the player must then click on the 'Submit' button. Proposals received from the negotiation partner show up in the form of a second, non-adjustable slider on the right-hand side of the screen. At the same time, an 'Accept' button appears underneath.

New proposals can be submitted at any time and as often as desired. Any new proposal a player makes supersedes her current one. This feature can be used to make concessions to the negotiation partner, but players can also choose to withdraw previous accommodating proposals by submitting new ones that are less generous. Because this arrangement allows considerable flexibility in players' negotiation strategies, our design produces data on quite nuanced measures of the bargaining process, such as the timings of offers or subtle differences in each agent's willingness to compromise. There are no restrictions in the permissible proposals, except that we do not allow a 50-50 split. Thus, if they want to come to an agreement, players must decide who gets a greater share of the pie, even though the difference may be minuscule (e.g., 49-51). The main reason for excluding the single 50-50 option is to ensure that participants actively engage in bargaining instead of directly agreeing on the strong focal point. Furthermore, this aspect of our design reflects that in many naturally occurring settings it can be difficult or even impossible to achieve a perfect compromise because of indivisibilities of particular rights or objects. In these situations, one side must give in at least a bit more than the other so that a feasible agreement can be achieved. In all other respects, our design ensures symmetry between the parties in the actual bargaining process.

Players are informed that a negotiation round can last up to 60 seconds and the decision screen displays a countdown timer. If one of the two players clicks on the 'Accept' button during the negotiation, an agreement is reached and the pie is split according to the accepted proposal. Otherwise, if neither player can bring herself to accept a partner proposal within the time limit, the round ends with the implementation of the disagreement point with payoffs d_A and d_B.

Participants play this game for a total of 11 rounds with disagreement points that vary from round to round so that we can examine bargaining in qualitatively different situations. We consider symmetric settings ($d_A = d_B$), asymmetric ones ($d_A \neq d_B$), and settings in which coming to an agreement in the negotiation phase is efficient ($d_A + d_B < 100$), is inefficient ($d_A + d_B > 100$), or is efficiency-neutral ($d_A + d_B = 100$). Table 1 provides an overview. To limit the effects of reciprocity across rounds and to alleviate other similar strategic concerns, we employ a stranger-matching protocol. Participants are assigned to twelve-person matching groups from which pairs are formed in each round. Each matching group constitutes a statistically independent observation. Our matching protocol is designed such that each participant experiences exactly 11 different disagreement points in random order. Given the large matching groups, each interaction in our experiment can be considered a one-shot game.

3.2 Nuke treatment

The Nuke treatment is identical to the Baseline treatment except that player B now has the power to trigger a nuclear option in case of no agreement (and only then). To make the looming threat from the nuclear option salient to both parties, we present it as a

Table 1: Disagreement points used in the experiment

	Agreement efficiency		
	Improving	Neutral	Deteriorating
$d_A = d_B$	30-30	50-50	60-60
$d_A > d_B$	40-30, 60-30	60-40	60-50
$d_A < d_B$	30-40, 30-60	40-60	50-60

red button that is visible throughout the negotiation in the lower part of the screen. At this point, the button is greyed out and is not responsive to mouse clicks (see Figure 1 again). It becomes enabled and fully visible (for player B) only if the negotiation ends without agreement. In this case, player B faces the choice between implementing the regular disagreement payoffs d_A and d_B (blue 'Continue' button) or activating the nuclear option. If B chooses the latter, player A's payoff is wiped out completely, and B's own payoff is reduced to 9 tokens.[5]

Importantly, player B's decision is **not** communicated to A until the very end of the experiment to avoid possible spill-over effects onto new partners and to keep information and, thus, participants' beliefs regarding the likelihood of nuclear-option usage constant across groups and rounds. Note that participants are randomly allocated to their role (A or B) at the beginning of a session and remain in that role for the entire duration of the experiment.

3.3 Threat treatment

The Threat treatment is identical to the Nuke treatment except that player B now has to *activate* the nuclear option before being able to use it. Player B can activate the option at any point during the negotiation. If she does, player A is informed about this by a flashing red frame around the grey box in Figure 1. Importantly, even if the nuclear option is activated, it is not triggered automatically, making it a pure signaling device. If no agreement is reached, player B faces the exact same screen as after a failed negotiation in the Nuke treatment and also has to decide whether to trigger the nuclear option or not. If the option is not activated during the negotiation and no agreement is reached, player B can still activate the nuclear option afterwards and then decide about using it. Thus, all choices remain the same as in the Nuke treatment except that player B has to click one additional button to be able to use the nuclear option. Player B can use this button during the negotiation to alert player A to the fact that the nuclear option is active. As in the Nuke treatment, player A is not informed about whether player B actually uses the nuclear option until the very end of the experiment.

[5] We choose this number so that player B's payoff is low enough to ensure that the option is in all cases very unattractive (9 tokens are just 30% of the smallest outside option player B can get in any scenario) but is still noticeably better than player A's payoff.

3.4 Procedures

Online experiment. The experiment was conducted online via Amazon's Mechanical Turk (MTurk), which has—in addition to offering a more representative sample (see Buhrmester et al., 2011; Berinsky et al., 2012; Paolacci et al., 2010)—two advantages for us compared to a conventional laboratory experiment. First, there is greater anonymity for participants (we only have their MTurk-IDs) which might result in more reliable results regarding behavior that may be deemed 'anti-social'.[6] Second, there are minimized reciprocity concerns, as participants have no way of meeting other participants or figuring out who was assigned as their partner in any round. To ensure a high-qualitative sample (i.e., participants understanding the task and paying attention), we restricted recruitment to (a) US-based individuals with an MTurk approval rate of 97% or higher and a history of more than 500 approved HITs, (b) individuals not using a mobile phone or VPN clients for the study, and (c) individuals passing a CAPTCHA test. Since our experiment involved real-time interactions over a series of rounds, it was necessary that all members of a matching group were logged in at the same time. We split up the experiments over two days to address this logistical problem. On the first day, volunteers participated in an introductory, individual-tasks session that they could start at any time and in which they learned about the setting. The main interactive experiment took place on the following day.

Introductory session. The first part of the experiment was publicized as an MTurk HIT with a fixed payment of $1. The HIT's description explained the two-part structure of the study and the time and date of the second part. The introductory session then started with an elicitation of basic demographic information, followed by an attention check. Failing the attention check resulted in the immediate exclusion from the experiment.

Participants then received written instructions for the next day's main experiment (see Online Appendix C). We tested participants' understanding with six control questions and excluded everyone with two or more mistakes. The understanding was further increased by five unincentivized practice rounds against the computer. The first of these was a tutorial in which all parts of the bargaining screen were explained, and participants were instructed to submit a particular proposal. Starting from round 3, we introduced a timer on all pages such that participants had to advance within a prescribed time limit. This procedure was introduced to prevent delays by inactive participants in the main task.

In the Baseline treatment, the introductory session ended after the last practice round. In the Nuke treatment, we went on to elicit participants' beliefs about how often the nuclear option would be used in each of the 11 different disagreement-point scenarios. This was done after individuals had been informed about their role (A or B) in the main experiment. For incentivization, one of the 11 scenarios and one of the participants were randomly determined following the main experiment. That participant received $100 if her stated belief was within a 1% margin of the actual usage of the nuclear option in that situation, $90 if her stated belief was within a 2% margin, and so on. In addition, we asked only the nuclear-option holders (people in role B) how often they think they themselves would use the nuclear option in each scenario, but this was not incentivized.

Main session. We sent two reminders of the main task, the first one 60 minutes and the second one 5 minutes before the start of each main-task session. On arrival, participants

[6] In the laboratory, participants might be worried that their use of the nuclear option is observable (at least during the payment process).

were provided with the instructions again and then redirected to a waiting page.[7] As soon as six A players and six B players had arrived, the computer created a matching group of 12, and the main task commenced. Timers on all stages ensured that participants moved on at similar speeds to reduce waiting times. Our policy for the case of a person dropping out was to remove the dropout's current interaction partner from the experiment as well and to pay her for a randomly chosen earlier round. Other members of the matching group were allowed to continue.

After the main task was complete, participants filled in a questionnaire about their experiences during the experiment. In the Nuke and Threat treatments, the questionnaire also contained questions about their thoughts on how the nuclear option had been used and what effects it might have had (see Online Appendix D.2 for details).

Payment. Participants were paid via the MTurk platform after the main session and according to their token balance at the end of one randomly selected round. On average, participants spent 21 minutes in the introductory session and 18 minutes in the main session. Average earnings amounted to $7.01, which therefore exceeded the target payment of $9 per hour for typical US-based MTurk workers (Berg, 2015).

3.5 Participants

Sample size. Because of the two-stage nature of the study and our screening procedures, many participants did not finish the experiment. Of the 5,559 participants who started, 1,555 failed the attention check, and 1,990 failed the control-question stage. A further 56 participants did not finish the introductory session and were dropped from the sample as well. Of the remaining 1,958 participants who were invited to the main session, 936 showed up, and 742 were successfully assigned to matching groups. Over the course of the main session, 21 participants dropped out, which caused their current partner to be forwarded to the questionnaire without finishing the remaining rounds. In our analysis, we rely on the recorded data of all participants, even the ones who dropped out. In the most conservative approach focusing only on completely independent observations, we have 22 matching groups in the Baseline treatment, and 20 matching groups in the Nuke and Threat treatment each.

Demographics. 50% of participants who took part in the main session are female, and their age range is 18 to 78 years with a median of 37.

4 Theoretical considerations

If we assume players to be rational and selfish payoff maximizers and this to be common knowledge, the nuclear option is irrelevant. There are no circumstances under which a player would choose the strictly dominated nuclear outcome over the regular disagreement point and, this being the case, its presence cannot have an effect on the negotiation phase.

In this section, we present a basic model in which player B *may* trigger the nuclear option, either because of a 'heat of the moment' decision or for more deliberate, social-preference-based reasons. Our ambition is not to develop a fully fledged theory of human behavior in the context of bargaining games. We merely seek to construct a framework within which we can think about the players' incentives and trade-offs, and which allows us to discuss plausible comparative-static effects.

[7] The instructions remained accessible at the bottom of the screen throughout the session.

4.1 Setup

Our model has three key ingredients.

(1) Own-payoff and relative-payoff motives. Players in our model are interested both in their own *absolute payoffs* and in the *payoff shares* that they receive (as in Bolton and Ockenfels, 2000). Player i's utility function is

$$u_i(\pi_i, \pi_j) = \left(\frac{\pi_i}{\pi_i + \pi_j}\right)^{\theta_i} (\pi_i)^{1-\theta_i} \tag{1}$$

where $\pi_i,\ \pi_j > 0$ are the players' monetary payoffs and where $\theta_i \in [0,\ 1]$ can be thought of as a 'spite' parameter.[8] $\theta_i = 0$ represents the standard case where player i is purely motivated by her own payoff, and $\theta_i = 1$ is the other extreme in which player i blocks out all thought about her absolute well-being and only cares about maximizing her share of the pie irrespective of its size.

(2) Noisy decision-making. We assume players to be boundedly rational in that their choices are probabilistic. Similar to Chen et al. (1997), when player i is confronted with two options S and T, she selects option S with probability

$$p_i(S,T) = \frac{\left(u_i^S\right)^{\lambda_i}}{\left(u_i^S\right)^{\lambda_i} + \left(u_i^T\right)^{\lambda_i}} \tag{2}$$

where u_i^S and u_i^T are the utility levels that the player obtains from choosing either S or T and where $\lambda_i \geq 0$ is a rationality parameter. At $\lambda_i = 0$ both options have a 50-50 chance of being selected. As λ_i increases, it becomes more and more likely that player i takes the option associated with the higher utility level, $\max\left\{u_i^S,\ u_i^T\right\}$. In the limit, when λ_i approaches infinity, player i makes the utility-maximizing choice for sure. For intermediate values of λ_i, the probability $p_i(S,T)$ is falling in the ratio u_i^T/u_i^S and equals 0.5 when $u_i^T = u_i^S$. Thus, the more clearly player i is positioned in her evaluation of alternatives S and T, the less likely she is to choose the lower-utility option. This captures the idea of an agent whose actual preferences may be briefly overruled by an emotional impulse but with a probability that is decreasing in the cost of committing such an 'error'.

(3) Nash bargaining. As the final ingredient of our model, we assume that *if* it comes to an agreement, it follows the Nash bargaining solution (Nash, 1953). Let x be the number of tokens that go to player B in an agreement, leaving $100 - x$ tokens for player A ($0 \leq x \leq 100$). In the Nash bargaining solution, $x = x^*$ solves the following optimization problem:

$$\max_x \big(\left(u_A(100-x, x) - \omega_A\right)\left(u_B(x,\ 100-x) - \omega_B\right)\big) \tag{3}$$

where ω_i is player i's expected utility from the outside option (i.e., no agreement outcome).

For our model, we assume that the x^*-solution is effectively on the table during the negotiation. Players individually decide whether to agree to this outcome. If at least one

[8] For $\theta_i > 0$ player i takes pleasure in any payoff reduction for player j. See Mill and Morgan (2022); Kirchkamp and Mill (2021); Mill and Morgan (2021); Mill and Stäbler (2022); Varma (2002); Klor and Shayo (2010) for similar approaches.

of them refuses to accept, the negotiation phase ends in a disagreement. Given the noisy decision-making element in our model, there is a positive probability for the x^*-solution (not) to be implemented even when this violates individual rationality.

4.2 Baseline treatment

In the Baseline treatment, there is no nuclear option. Players negotiate in the knowledge that if they do not come to an agreement, the outcome is defined by the regular disagreement payoffs d_A and d_B, implying utility

$$u_i(d_i, d_j) = \frac{d_i}{(d_i + d_j)^{\theta_i}} =: \omega_i$$

for player $i \in \{A, B\}$. The Nash product from (3) is maximized for:

$$x^*_{\text{Base}} = 50 + \frac{1}{2}\left(100^{\theta_B}\omega_B - 100^{\theta_A}\omega_A\right).$$

We refer to $x_A = 100 - x^*_{\text{Base}}$ and $x_B = x^*_{\text{Base}}$ as player A's and player B's "agreement payoffs". Player i's agreement payoff x_i has very natural comparative-static properties. For example, it is increasing in d_i and falling in d_j. Thus, a more favorable disagreement payoff implies a better outcome from the negotiation. Conversely, a higher disagreement payoff for the opponent hurts the player. Furthermore, in scenarios in which agreements are efficient ($d_i + d_j < 100$), being (known to be) spiteful improves a player's bargaining outcome and the more spiteful she is, the better she performs.[9] In contrast, when an agreement would be inefficient, x_i is falling in the spite parameter θ_i. For efficiency-neutral scenarios the model predicts $x_A = d_A$ and $x_B = d_B$. In this case, players are indifferent between accepting and rejecting the x^*-solution.

What is the probability of reaching an agreement for other values of d_A and d_B? Since an agreement will be struck only with both A's and B's approval, we must consider the effects of a change of one of the disagreement payoffs on both parties. First, while a player is predicted to achieve a more favorable agreement when her disagreement payoff d_i increases, any such improvement in the agreement does not keep up with the utility increase from the disagreement payoff itself. This means that a higher value of d_i *reduces* the relative attractiveness of the x^*-solution for player i, and she becomes less likely to accept.

Second, if there is an increase in the *rival's* disagreement payoff d_j, player i's bargaining position is harmed. While this clearly lowers i's utility from an agreement, the alternative is *also* negatively affected for any $\theta_i > 0$ because player i's relative position in the disagreement point becomes worse. Thus, a priori, the net effect is not clear. However, in the scenarios that we consider for our experiment, our model predicts a player to agree with a *lower* probability when the other player's disagreement payoff is increased.[10]

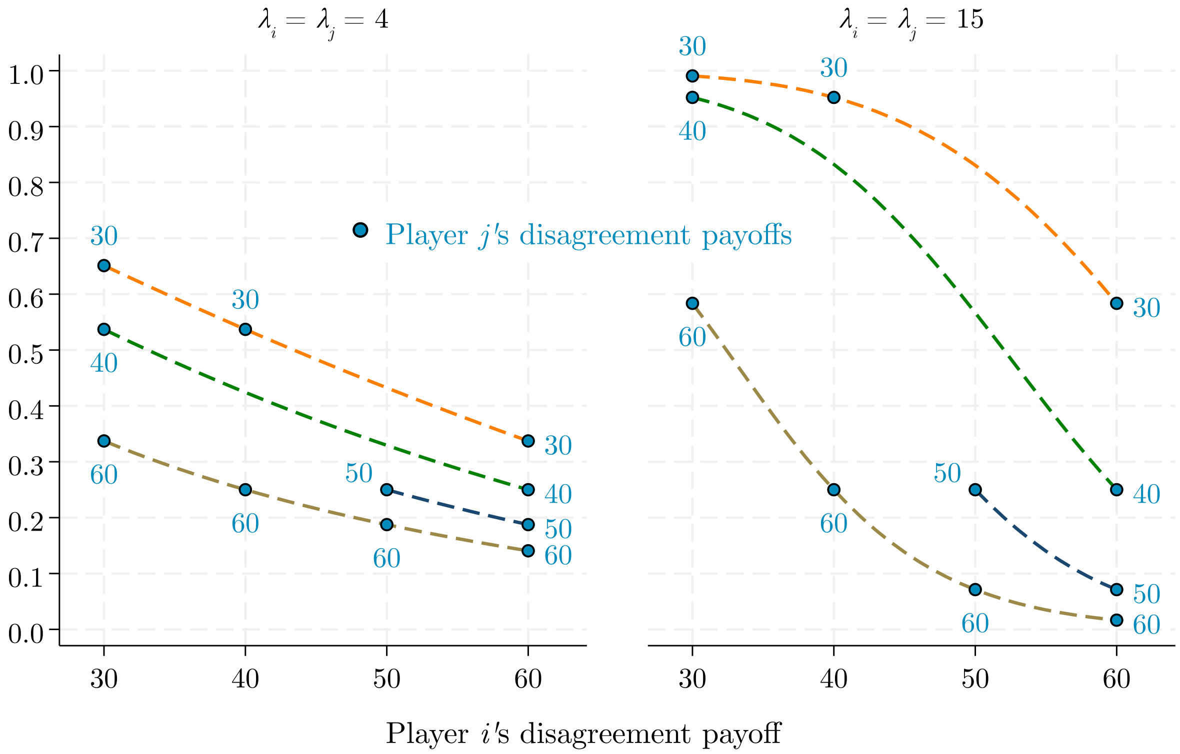
Taken together, the model therefore suggests that players' propensities to accept the x^*-solution decline whenever there is a rise in disagreement payoffs, regardless of which player is made better off. Consequently, as agreements become less efficient, they are predicted to occur with lower probabilities. Figure 2 illustrates how the agreement

[9] For details on our claims see Online Appendix A.

[10] There is one extreme case in which player i's accept probability remains constant when d_j is increased. This requires that $\theta_i = 1$, $\theta_j = 0$, and $d_i = 30$, and occurs when d_j is increased from 30 to 40 (see Online Appendix A).

probabilities depend on the disagreement points for the case of a low rationality parameter ($\lambda = 4$) and for the case of a relatively high one ($\lambda = 15$), assuming a spite parameter of $\theta = 0.3$.[11]

Figure 2: Agreement probabilities with $\theta_i = \theta_j = 0.3$.



The figure illustrates how the theoretical agreement probabilities depend on the players' disagreement payoffs and on the rationality parameter λ when $\theta = 0.3$. The scenarios relevant for the experiment are highlighted.

As the figure shows, agreements are predicted to be more likely when they lead to efficiency gains relative to the disagreement point—e.g., case 30-30 or case 30-40—than when they are efficiency-deteriorating—e.g., case 50-60 or case 60-60. In the 30-60/60-30 scenarios, there is a payoff advantage from striking an agreement relative to the disagreement points, but it is small. Hence, noise in the decision to accept is more pronounced and makes agreements more fragile. Efficiency-neutral scenarios present a special case since players are indifferent between the agreement payoffs and the outside options. Individually, they will agree with probability 0.5, implying that an agreement is struck with probability 0.25 regardless of the degree of noise in the decision to accept. We summarize our theoretical results for the Baseline treatment in

Hypothesis 1a. The frequency of agreements is inversely related to $d_i + d_j$. Agreements are rare in cases in which $d_i + d_j > 100$ but are common when agreements are efficient. In scenarios where $d_i + d_j = 100$, exactly 25% of negotiations lead to agreements—and when they do, payoffs replicate the disagreement point.

Hypothesis 1b. Player i's agreement payoff $x_i(d_i, d_j)$ increases in d_i and decreases in d_j such that $x_i(30, 60) < x_i(30, 40) < x_i(30, 30) < x_i(40, 30) < x_i(60, 30)$ for efficiency-

[11] These parameters are chosen for illustration, but they are in the range of estimates found in the literature (see for example Levine, 1998; Abbink and Sadrieh, 2009; Abbink and Herrmann, 2011; Mill and Morgan, 2022; Chen et al., 2012).

improving agreements, and $x_i(40,60) < x_i(50,60) < x_i(50,50) = x_i(60,60) < x_i(60,50) < x_i(60,40)$ for inefficient and efficiency-neutral agreements.

4.3 Nuclear treatments

In the Nuke and Threat treatments, player B may trigger the nuclear option following a disagreement. If he does, he earns 9 tokens, while player A receives 0. Player B's utility is thus $u_B(9,\ 0) = 9^{1-\theta_B} =: \eta_B$. This is preferable to the regular disagreement point if

$$\eta_B \geq \omega_B \qquad \Leftrightarrow \qquad \theta_B \geq \frac{\ln(d_B) - \ln(9)}{\ln(d_A + d_B) - \ln(9)} =: \widehat{\theta}_B.$$

If player B is completely selfish, he has no motive to trigger the nuclear option ($\theta_B = 0 < \widehat{\theta}_B$). If relative payoff concerns play a role ($\theta_B > 0$), the severity of the nuclear threat is directly linked to how disadvantageous the regular disagreement point is for player B, as $\widehat{\theta}_B$ is decreasing in d_A and increasing in d_B. For example, if d_A is high, then the critical spite parameter value $\widehat{\theta}_B$ is low and there is greater scope for player B to find the nuclear option appealing.

A threat may also arise from a player B who is perhaps not exceedingly spiteful but makes a hasty, emotional decision after a disagreement has occurred. In the model, this would be captured by a low λ_B parameter. The effects are similar to the case of spite: Since the probability of going against one's actual preference is inversely related to the cost of doing so, player B is the less likely to trigger the nuclear option due to a "heat of the moment" impulse the more attractive the regular disagreement point is. Hence, the prediction is that a nuclear outcome is most likely to occur in scenario 60-30 and least likely to occur in scenario 30-60 since out of the eleven cases we consider, these two differ most from each other in terms of absolute and relative payoff for player B.

Formally, the probability of player B choosing the nuclear option (N) over the regular disagreement point (D) is

$$\sigma_B^N = \frac{\eta_B^{\lambda_B}}{\eta_B^{\lambda_B} + \omega_B^{\lambda_B}} = \frac{9^{\lambda_B(1-\theta_B)}\,(d_A + d_B)^{\lambda_B\theta_B}}{9^{\lambda_B(1-\theta_B)}\,(d_A + d_B)^{\lambda_B\theta_B} + d_B^{\lambda_B}}$$

with $\partial\sigma_B^N/\partial d_B < 0$ and $\partial\sigma_B^N/\partial d_A \geq 0$.

Hypothesis 2. In treatments Nuke and Threat some B-players trigger the nuclear option. The relative frequency of this increases in d_A and decreases in d_B.

The central question of our study is whether, and if so how, the introduction of a nuclear option affects the bargaining-stage outcomes. Given the expected utility from a disagreement, $\sigma_B^N\eta_i + \left(1 - \sigma_B^N\right)\omega_i$, the new Nash bargaining solution is

$$x^*_{\text{Nuke}} = \left(1 - \sigma_B^N\right) x^*_{\text{Base}} + \sigma_B^N \left(50 + 100^{\theta_B}\frac{\eta_B}{2}\right). \tag{4}$$

and hence $x^*_{\text{Nuke}} \geq x^*_{\text{Base}} \;\Leftrightarrow$

$$100^{\theta_A}\omega_A \geq 100^{\theta_B}\left(\omega_B - \eta_B\right). \tag{5}$$

Note that player B actually *preferring* the nuclear outcome ($\omega_B < \eta_B$) is a sufficient condition for increased bargaining power (as it turns the right-hand side of inequality

(5) negative), but not a necessary one. Moreover, it follows directly from (5) that the Nuke treatment gives player B an advantage in any symmetric setting with $d_A = d_B$ and $\theta_A = \theta_B$ regardless of player B's preferences over the nuclear vs. non-nuclear disagreement outcomes. Any further increase in d_A relative to d_B makes the introduction of the nuclear option even more effective for player B's bargaining position.

As we show in the Online Appendix A, player B's agreement payoff is predicted to (weakly) improve relative to the Baseline treatment for most of the scenarios we consider. A reduction in player B's bargaining power due to the nuclear option is possible in our model only when d_B is relatively high, θ_B is very low (no spite), and λ_B is close to zero (high impulsivity). In this case, player B does not truly wish to trigger the nuclear option but is essentially afraid of 'losing it' if it came to a disagreement, which makes him more inclined to seek a compromise in the negotiation. In the most extreme case—scenario (30, 60) with $\lambda_B = 0$ and $\theta_A = \theta_B = 0$—this would shift the bargaining outcome by about 5 tokens in favor of player A. For other parameter values and other cases, larger shifts in favor of B are predicted. Thus, while we acknowledge a 'backfiring' nuclear option as a theoretical possibility, our directional prediction is that player B's negotiation power increases when the nuclear option is available.

Hypothesis 3. B-players obtain higher agreement payoffs in the Nuke and Threat treatments than in Baseline. The difference is largest in the 60-30 scenario and smallest in 30-60.

Next, we consider how the presence of the nuclear option may affect the probability of A and B coming to an agreement. Intuitively, if player A perceives the nuclear threat as a non-negligible risk, she may be more willing to accept a given proposal than in the Baseline treatment. On the other hand, if player B makes less generous proposals because of increased bargaining power, player A should be *less* inclined to accept. For B, there are similar trade-offs depending on his preference parameter.

Overall, our model predicts that the nuclear option generally increases the chances of an agreement.[12]

Hypothesis 4. Agreement rates in the Nuke treatment are higher than in the Baseline treatment in all scenarios.

A final variable of major interest is efficiency. On the one hand, the role of a potential nuclear outcome is not merely one of deterrence according to our model—instead, the nuclear option is actually triggered with positive probability. This is detrimental to efficiency. On the other hand, the nuclear threat can have a disciplining effect on players, making agreements more likely and thereby improving efficiency. Unfortunately, however, the greater tendency to accept proposals in the negotiation stage includes cases in which agreeing is in fact socially undesirable.

From our analysis, we obtain clear-cut predictions when $d_A + d_B \geq 100$. In these cases, the predicted increase in agreement rates is either efficiency-neutral or efficiency-deteriorating, and at the same time, there is a risk for nuclear outcomes to occur.

[12] The only possibility for an effect in the opposite direction arises when player B is practically solely interested in maximizing his share of the pie (θ_B close to 1) while A is more balanced in her motivation. In such a scenario, player A makes desperate concessions to B in the negotiation, but in spite of this, player B rejects the x^*-solution with a roughly 50-50 chance leading to an overall lower agreement probability than in the Baseline treatment. Thus, our model's predictions are again not entirely unambiguous when considering the entire parameter range and we again focus on 'moderate' preference parameter values.

Hypothesis 5. In scenarios 40-60, 50-50, 50-60, 60-60, 60-50, and 60-40, efficiency is lower in Nuke and Threat than in Baseline.

For the remaining scenarios, in which $d_A + d_B < 100$, the model's predictions are highly parameter-dependent. The net effect depends on the frequency with which the nuclear option is triggered and on how agreement rates increase relative to the Baseline treatment.

4.4 Nuke versus Threat

Our theoretical approach does not produce a specific testable hypothesis regarding differences between the Nuke and the Threat treatment. This treatment variation is more exploratory and addresses research questions that naturally arise in the context of bargaining with a nuclear option. To the extent that the nuclear option has any effect at all, is it enough when the threat is implicit (as in the Nuke treatment), lurking in the background of the negotiation? Or is it essential that player B actively chooses to put the nuclear option on the table (as she can do in the Threat treatment), signaling a willingness to create havoc unless she gets what she wants? Under what circumstances, if at all, is player B inclined to activate the nuclear option? How does player A respond to such an explicit threat during the negotiation? Do activations in the Threat treatment lead to even higher agreement rates than in the Nuke treatment? Do such "strategic" activations lead to better outcomes for player B?

5 Results

5.1 Bargaining behavior in the Baseline treatment

As a first step, we briefly describe how a "typical" negotiation proceeds to get a rough idea of the general dynamics of the bargaining task. Based on median values from the Baseline treatment, we see a first proposal being submitted after 7 seconds, asking for 60 percent of the 100-token pie. If the players come to an agreement (which they do in 62% of cases), they reach it 21 seconds later. The median number of submitted proposals in these cases is only 3. Across all scenarios, an agreement yields a median payoff of 49 for the individual who accepts. In negotiations that do not end up in an agreement, a larger number of proposals is exchanged (median = 5). The difference is statistically significant ($p < 0.001$).[13] In these cases, the last proposal is submitted 8 seconds before the end of the one-minute time limit and contains a demand of 60 points.[14]

According to Hypothesis 1a, negotiation outcomes should be systematically affected by the relevant disagreement payoffs. We examine this in Figures 3 and 4.[15] Figure 3

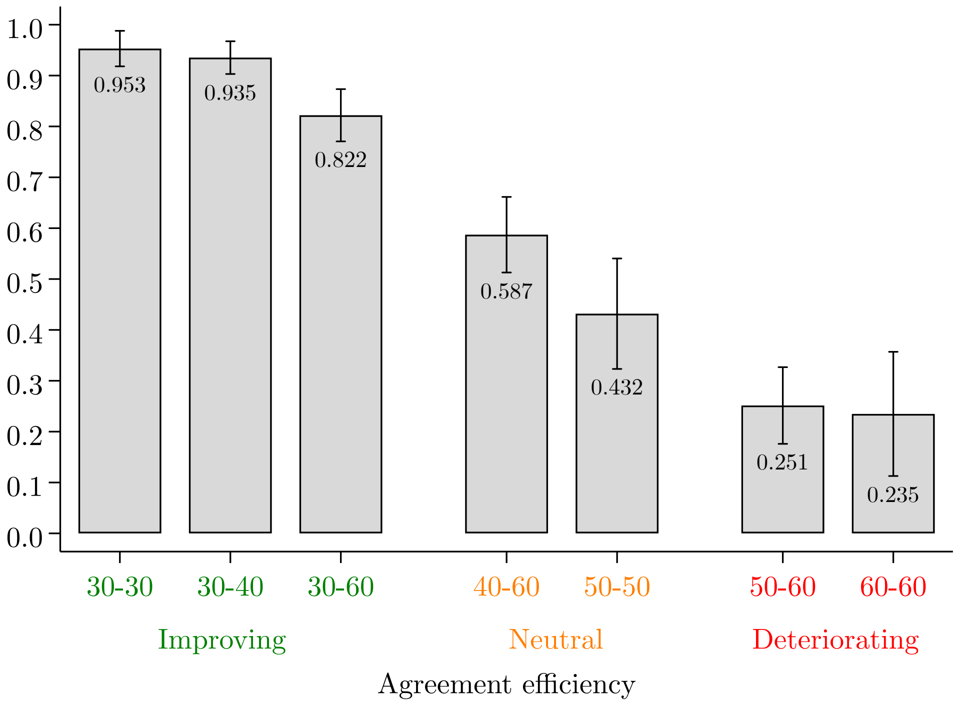
[13] Unless indicated otherwise, reported p-values in the results section stem from non-parametric two-sided Fisher-Pitman permutation tests conducted at the level of statistically independent matching groups.

[14] One can think of various questions about individual negotiation strategies (e.g., the timing, the frequency, the nature of offers) and how they affect outcomes. While we have done some analyses of this kind, we found them to be more of a distraction from the main focus of our paper than to provide relevant additional insights. Therefore, we do not report them here. However, we are happy to share our analyses with interested parties upon request.

[15] The player designations ("A", "B") should be irrelevant in the Baseline treatment, and we indeed find no significant differences for any of our measures between these labels. Thus, except where stated

shows, across the various disagreement points, how often negotiators come to an agreement. Consistent with Hypothesis 1a, the frequency of agreements decreases in the sum of disagreement payoffs. This result is also corroborated by a regression analysis reported in Table 2 in the appendix. Overall, the bargaining outcomes thus appear to reflect the given incentives across the different scenarios. However, Figure 3 also shows that agreements are not always optimal: Not all negotiations are successful when an agreement would be efficiency-enhancing (scenarios highlighted in green) and, conversely, sometimes negotiators do agree to share the pie when this is detrimental to efficiency (red cases). For cases in which the disagreement payoffs sum up to exactly 100 (orange), we find that agreements are much more common than our point prediction of 25% ($p < 0.001$).

Figure 3: Agreement rates in Baseline.



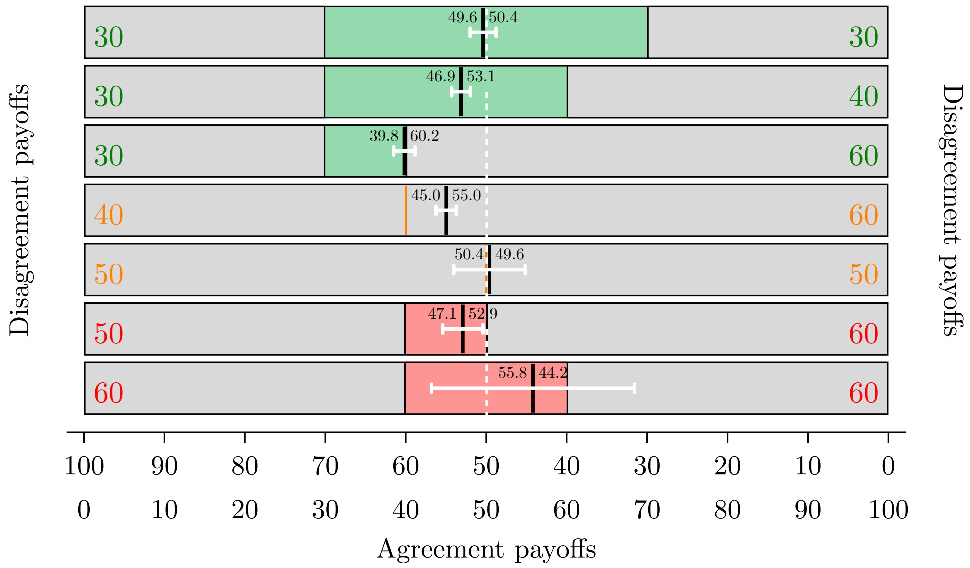
Notes: Relative frequencies of agreements for all disagreement points in Baseline. 95% confidence intervals based on the means of the independent matching groups.

Figure 4 shows what agreements look like in each case. When the disagreement point is symmetric, the median agreement is statistically indistinguishable from an even split. In cases with asymmetric disagreement points, the player with the higher disagreement payoff receives a larger share of the pie but benefits *less* from the agreement than her negotiation partner when outcomes are compared to the disagreement point. This is most visible in the situation where one party has a disagreement payoff of 30 while the other would obtain 60: The player with the smaller disagreement payoff is able to secure almost the entire surplus from the agreement! This result is not predicted by our model—a possible explanation might be that inequality aversion among our participants creates a bias towards the 50-50 split. Nevertheless, in line with Hypothesis 1b, regression

otherwise, we pool these data in this subsection. For example, when we examine the 30-40 scenario, our analysis is based on all encounters in which one player's fallback position is 30 and the other one's is 40—irrespective of the player designations. Thus, we consider 7 distinct scenarios in the Baseline treatment, whereas we have 11 different scenarios in the other treatments where the situation for players A and B is no longer symmetric.

analysis (reported in Table 3 in the appendix) suggests that a negotiator's payoff from an agreement increases in their own disagreement payoff and decreases in their partner's disagreement payoff.

Figure 4: Agreement payoffs in Baseline.



Notes: The figure shows how the 100-token pie is divided in each of the 7 experimental scenarios. Averages and 95% confidence intervals are based on the means obtained for the independent matching groups. The thin white vertical dashed line represents a 50-50 split. In the scenarios where agreements are efficiency-improving (30-30, 30-40, 30-60), the green-shaded areas represent the Pareto-superior (relative to the disagreement point) agreement range. When agreements are efficiency-deteriorating (50-60, 60-60), the red-shaded areas represent outcomes in which both parties lose. For symmetric disagreement points (30-30, 50-50, 60-60), the outcomes shown refer to the agreement payoffs by player designation (left: "Player A", right: "Player B").

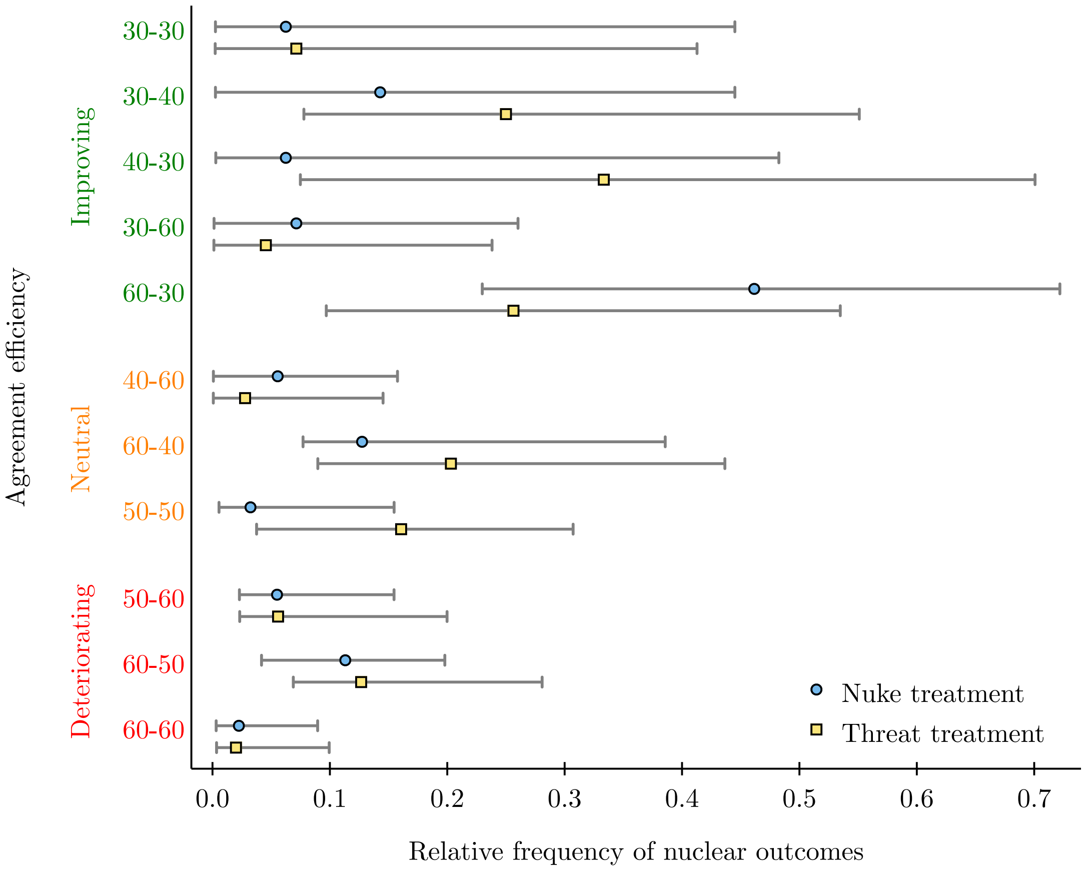
5.2 Nuclear treatments: Expected and actual nuclear use

Before we address the effects of introducing the nuclear option on bargaining behavior, we first provide an overview of its *actual*, as well as its *anticipated*, usage to assess whether it is (perceived to be) a threat despite being a strictly dominated choice for rational and own-payoff-maximizing decision makers.

Actual nuclear use. Overall, B-players do not appear to be very "trigger-happy": 85% of them *never* use the nuclear option and fewer than 3% use it more than twice. As a result, the nuclear outcome in our two nuclear treatments is avoided in more than 97% of encounters. Nevertheless, there are instances in which B-players do go nuclear in spite of the very high cost. Figure 5 displays, for each of our scenarios and for both the Nuke and the Threat treatment, how often this occurs conditional on a negotiation ending in disagreement. The treatment variation turns out not to be a major factor: The overall probability of player B choosing the nuclear outcome instead of the regular disagreement point is statistically the same in both treatments ($p = 0.869$) and, as the figure shows, trigger rates are similar between treatments in most scenarios, with a few exceptions. At the same time, there is considerable variation *across* disagreement points—the risk of a nuclear outcome can be very low or can be quite substantial. Because agreement rates are high in many cases (see further below for a detailed analysis), the data basis for our analysis is limited and mean trigger rates often have wide confidence intervals, making pairwise statistical comparisons difficult.[16] However, our regression analysis (Table 4 in the appendix) indicates that trigger rates in both nuclear treatments systematically increase in player A's disagreement payoff and decrease in player B's disagreement payoff. In Figure 5, this pattern can be illustrated, for example, for the Nuke treatment in the 30-60 scenario vis-à-vis the 60-30 scenario. B-players are quite likely to go nuclear in the latter case when the nuclear damage is most severe for their negotiation partners and least severe for themselves. Yet, in the former case, when the situation is reversed, nuclear outcomes are rare. This type of behavior lends support to Hypothesis 2.

In summary, it appears that the nuclear threat is real in our setting. Under specific and predictable circumstances, holders of a clearly dominated nuclear option may be prepared to actually use it. The question is whether A-players anticipate this and, if so, how this affects bargaining behavior.

Figure 5: Nuclear option trigger rates after no agreement.



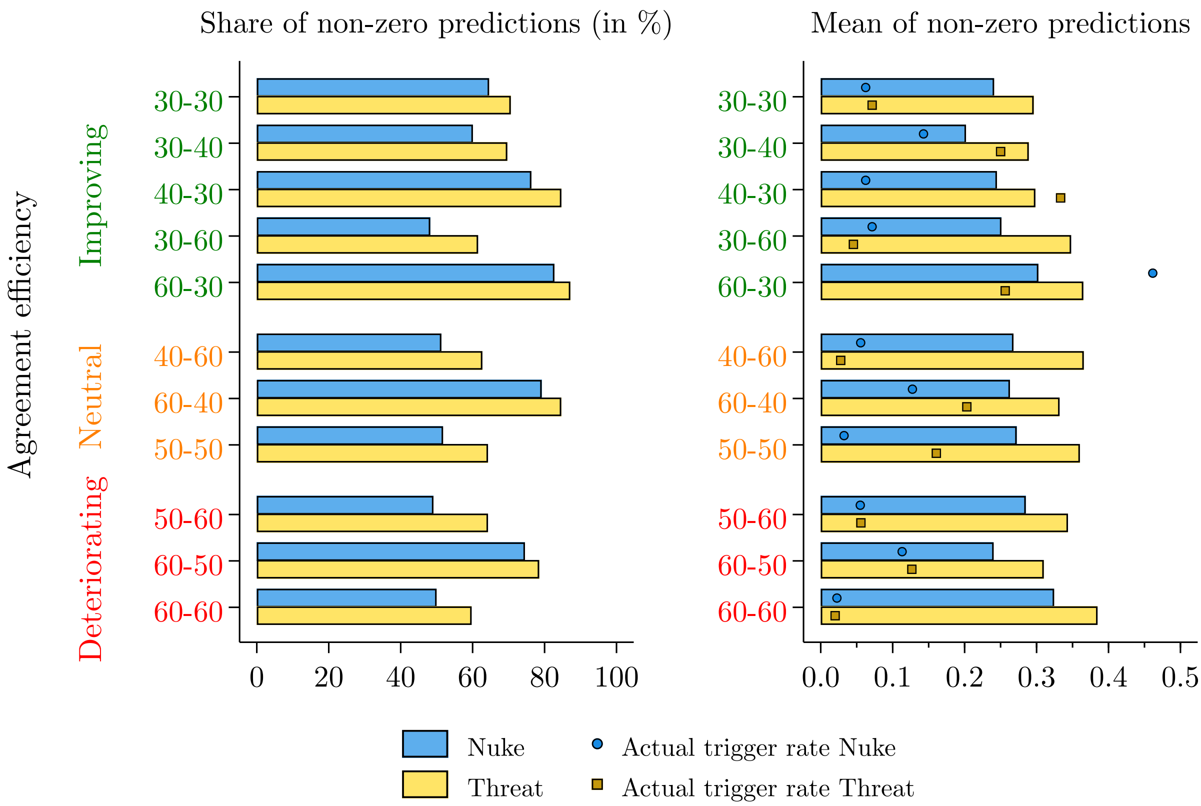
Notes: Error bars show 95% Clopper-Pearson confidence intervals based on individual choices.

Nuclear use beliefs. Recall that we prevented A-players from experiencing nuclear outcomes during the experiment: The individual negotiations were one-shot and A-player subjects received no feedback at the end of a round whether or not the nuclear option had been triggered. To examine to what extent the nuclear threat may have been present during the negotiations, we now turn to players' beliefs which we elicited prior to the main experiment. Figure 6 shows how likely participants think it is that the nuclear

[16] Note also that because most B-players go nuclear at most once, the variation across disagreement points does not so much reflect changes in individual propensities to trigger but is driven by different rare users of the nuclear option.

option will be triggered if no agreement is reached.[17] We find that in every scenario there is a substantial fraction of participants who expect that the nuclear option will not be triggered *at all*. However, the frequency of these 'zero predictions' is not constant across disagreement points. This can be seen in the left panel of the figure (which displays the share of non-zero predictions). The panel on the right shows, for each case, the mean prediction of those who do think that there is a chance that the nuclear outcome might occur.

Figure 6: Participants' nuclear trigger rate predictions.



Notes: The figure depicts, for each disagreement point, participants' stated beliefs over the trigger rates. The left panel shows the proportion of predictions that do not mark the nuclear outcome as a zero-probability event. The right panel shows the average non-zero prediction.

By and large, the shares of participants' non-zero predictions on the left-hand side of the figure follow the theoretical predictions summarized in Hypothesis 2. Participants become more inclined to attach a positive probability to a nuclear outcome when player A's disagreement payoff increases, and less inclined to do so when there is an increase in player B's disagreement payoff. This pattern corresponds to the results of our regression analysis of actual trigger rates in Table 4. Again, a striking difference can be observed when we compare the 60-30 and the 30-60 cases.

The *means* of the non-zero predictions (on the right-hand side of the figure) vary as well across disagreement points, but there are a few more deviations from the pattern

[17] As it turns out, there are no statistically significant differences between A-player predictions and B-player predictions, neither overall nor conditional on the individual disagreement points (see Table 6 in the appendix). Thus, the mere assignment to roles does not appear to affect predictions, and we therefore pool the A-player and B-player data here. Furthermore, it turns out that B-players do not predict for themselves to be any less or more willing to trigger the nuclear option than other B-players (see Table 7). Thus, our data does not support the notion that using the nuclear option is generally seen as an outlandish overreaction that one would mainly ascribe to others.

based on how damaging the nuclear outcome is for each player. For example, the mean non-zero predictions for 60-60 (Nuke: 0.32; Threat: 0.38) are even slightly *larger* than those for 60-30 (0.30 and 0.37, respectively) even though the nuclear option is costlier for player B in the 60-60 scenario.

When we combine zero and non-zero predictions to determine the overall means, the picture we obtain is again consistent with our observations of actual trigger rates. Table 5 reports the findings of a regression analysis of participants' predictions on treatment and disagreement points. The results qualitatively replicate those from the analogous analysis of the relative frequencies of actual nuclear outcomes (Table 4).

However, we also find that subjects are on average unable to correctly predict the overall *scale* of the nuclear threat. As illustrated in Figure 6, the mean non-zero prediction is typically far above the actual trigger rate. This is confirmed even when we include all those predictions that characterize the nuclear outcome as a zero-probability event. We then find that the observed trigger rates in the Nuke and in the Threat treatment are significantly lower than predicted by subjects in all but four cases.[18]

It is also remarkable to see that the nuclear option is considered a much bigger issue when it has to be activated first. Players expect higher trigger frequencies across the board in the Threat treatment relative to the Nuke treatment ($p < 0.001$). In actual fact, of course, this turns out not to be the case, as our analysis above has shown.

5.3 Bargaining under the shadow of the nuclear option

We have seen that the nuclear option is used ex post and predicted ex ante. This provides reason to expect that it will also affect negotiations. In this subsection, our main focus is on the Nuke treatment where the availability of the nuclear option is fully exogenous. We consider this to be the most conservative test since player B cannot, unlike in the Threat treatment, use the option strategically as a signaling device in his interaction with player A.

Total payoffs across treatments. Our first main result concerns the *efficiency* of introducing a nuclear option. On the one hand, it is conceivable that the nuclear threat improves outcomes overall by having a disciplining effect on players. Player A in particular may be inclined to negotiate more carefully and seriously, seeking to avoid falling out with player B. On the other hand, it is also possible that the presence of the nuclear option fuels any discord by making player B more obstinate or by creating a degree of defiance in player A. If both parties perceive a real shift in bargaining power due to the nuclear option, player B may even come up with the idea of enforcing a favorable agreement in situations in which an amicable disagreement would be the preferred outcome from a social planner's point of view. Furthermore, whenever the nuclear option is not merely used as a bargaining chip but is actually triggered—whether intentionally or by accident—this is obviously very destructive.

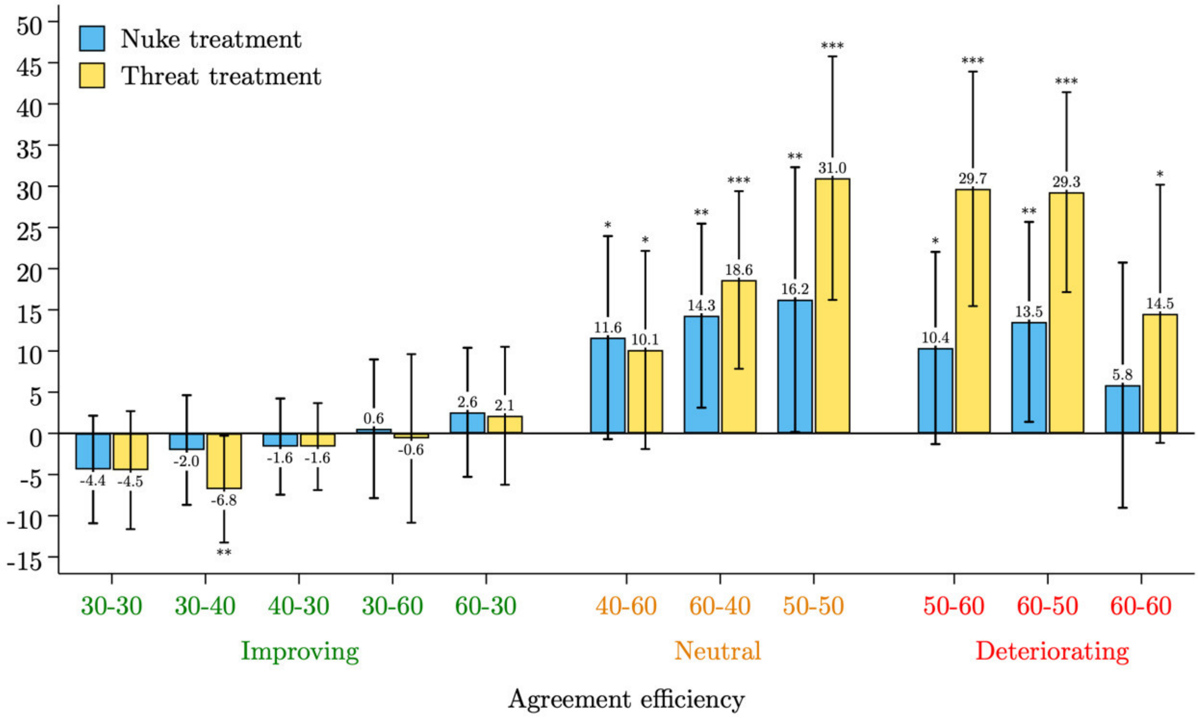
In our data, we find that the introduction of the nuclear option has a systematic *negative* effect on efficiency. Overall, payoffs fall by more than 3% in the Nuke treatment relative to the Baseline treatment. The difference is highly significant ($p < 0.001$). When we measure the attained level of efficiency by relating observed total payoffs to the difference between efficient and inefficient negotiation outcomes, the severity of the impact

[18] The exceptions in the Nuke treatment are 30-40 (no significant difference) and 30-60 (significant *under*-estimation of the nuclear threat). The exceptions in the Threat treatment are 30-40 and 60-30, where the actual trigger rates are significantly higher than the corresponding predictions.

of the nuclear option becomes even more palpable.[19] In Baseline, the average attained efficiency level in this sense is 84%, while the corresponding figure for the Nuke treatment is only 57%.

Changes in efficiency can come about in two ways: directly, by B-players deciding to trigger the nuclear option, and indirectly, by affecting negotiation behavior which then leads to changes in agreement rates. We will first consider the latter and then extend the analysis to include the former.

Figure 7: Percentage point changes in agreement rates relative to Baseline.



Notes: 95% confidence intervals are based on the outcomes of the independent matching groups. Stars indicate significance (using non-parametric two-sided two-sample Fisher-Pitman permutation tests) at the 10% (*), 5% (**), and 1% (***) level.

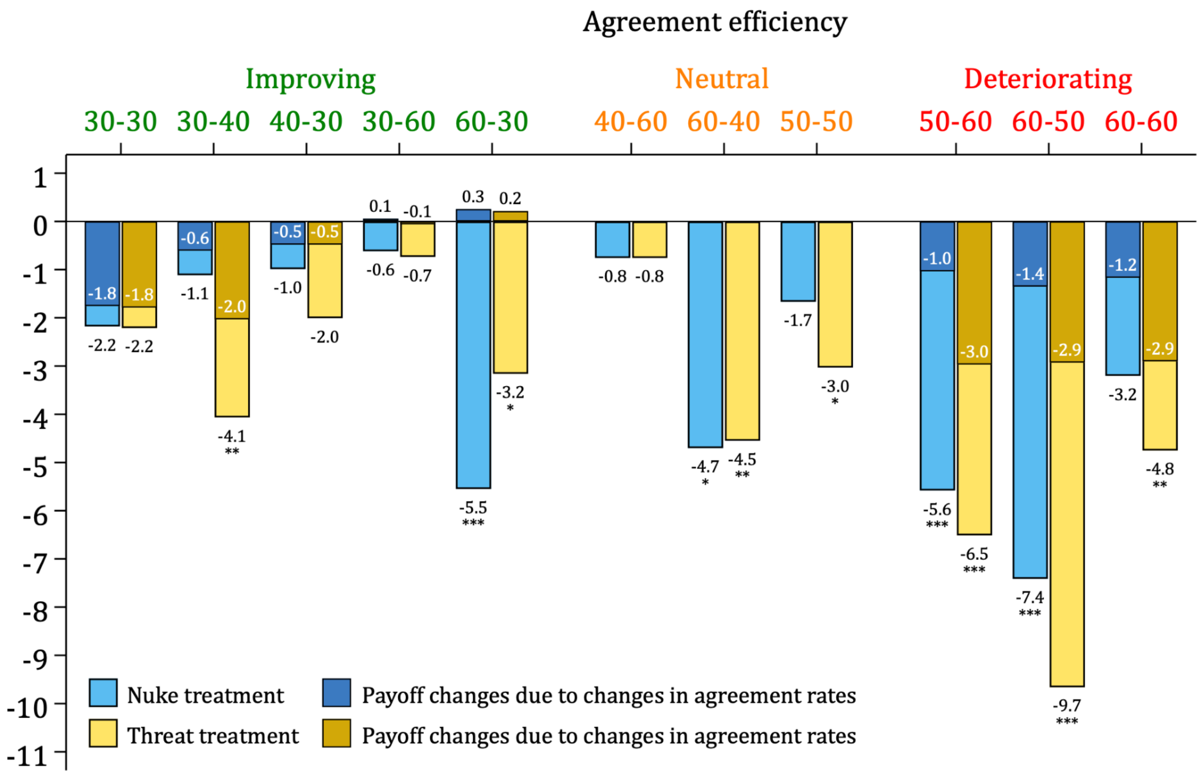
Efficiency losses due to changes in agreement rates. When we aggregate the Nuke data across all scenarios, we find that the nuclear option substantially improves the chance of coming to an agreement relative to the Baseline treatment. On average, the difference exceeds 10 percentage points ($p = 0.049$). And yet, efficiency is *lower*. The reason can be found in Figure 7 which displays the changes in agreement rates for each disagreement point. This reveals that the introduction of the nuclear option fails to make agreements more likely precisely when this would be desirable from an efficiency perspective (scenarios highlighted in green): On average, the agreement rate in these cases falls, compared to Baseline, by about 1 percentage point in Nuke and by about 2 percentage points in Threat ($p = 0.606$ and $p = 0.328$, respectively). Instead, the higher agreement rates in Nuke and Threat are driven by negotiation rounds in which the sum of the disagreement payoffs is equal to (scenarios highlighted in orange) or greater than (red scenarios) the 100 tokens that the players negotiate over. The increased chance of an agreement compared to Baseline amounts to, on average, 14 percentage points in Nuke

[19] For this exercise, we focus on scenarios in which the total disagreement payoff ($d_A + d_B$) diverges from 100. Let x be the observed sum of payoffs. The attained efficiency is then $(x - min\{d_A + d_B, 100\})/(max\{d_A + d_B, 100\} - min\{d_A + d_B, 100\})$.

($p = 0.013$) and 20 percentage points in Threat ($p < 0.001$) for scenarios characterized by efficiency-neutral agreements, and the corresponding figures for efficiency-deteriorating agreement scenarios are 10 percentage points (Nuke; $p = 0.063$) and 24 percentage points (Threat; $p < 0.001$). Thus, with the nuclear option being present, negotiators are more inclined to agree when this is not helpful from a social planner's perspective. For these cases, the data supports Hypothesis 4 whereas we find no significant treatment effects in scenarios in which agreements are efficiency-improving.

Efficiency losses due to nuclear destruction. Besides the indirect harm caused by the adverse effects on bargaining behavior, there is the direct damage to welfare arising from actual nuclear outcomes. To quantify the relative importance of both effects, consider Figure 8. The blue and yellow bars indicate the overall changes in total payoffs across treatments for each disagreement point (blue: Nuke vs. Baseline; yellow: Threat vs. Baseline). As the figure shows, the introduction of the nuclear option lowers efficiency in all cases. Additionally, the figure highlights how payoffs have changed due to different agreement rates alone (dark blue and brown bars). The direct nuclear damage on payoffs is therefore equivalent to the difference between the total and the payoff change due to a change in the propensity to come to an agreement.

Figure 8: Changes in joint payoffs relative to Baseline.



Notes: Stars indicate significance (using non-parametric two-sided two-sample Fisher-Pitman permutation tests) at the 10% (*), 5% (**), and 1% (***) level.

By design there is no indirect effect in negotiation rounds in which the sum of the disagreement payoffs *equals* 100 (neutral agreement efficiency) as no loss or gain can be accumulated due to agreements being struck or avoided. Here, the payoff reduction is exclusively caused by B-players triggering the nuclear option. In cases where an agreement might be beneficial (green scenarios) we find that both the direct effect (nuclear outcomes) and the indirect effect (lower agreement rates) contribute to the small overall reduction in payoffs discussed above. For example, in the Nuke 30-30 scenario payoffs are 2.2% lower than in Baseline and this is largely due to lower agreement rates. Only 0.4 percentage points of the reduction are caused by player B triggering the nuclear option. In contrast, in the 30-60 and the 60-30 scenarios, the actual use of the nuclear option is the sole culprit for the reduced efficiency. Even though the agreement rates are a little higher than in the Baseline treatment (see Figure 7), the resulting efficiency improvement is negligible compared to the direct effect. When agreements are detrimental (cases highlighted in red), we find that the increased agreement propensities we diagnosed earlier reduce the joint payoffs, and this reduction is amplified further—in some cases dramatically—by player B triggering the nuclear option relatively often.

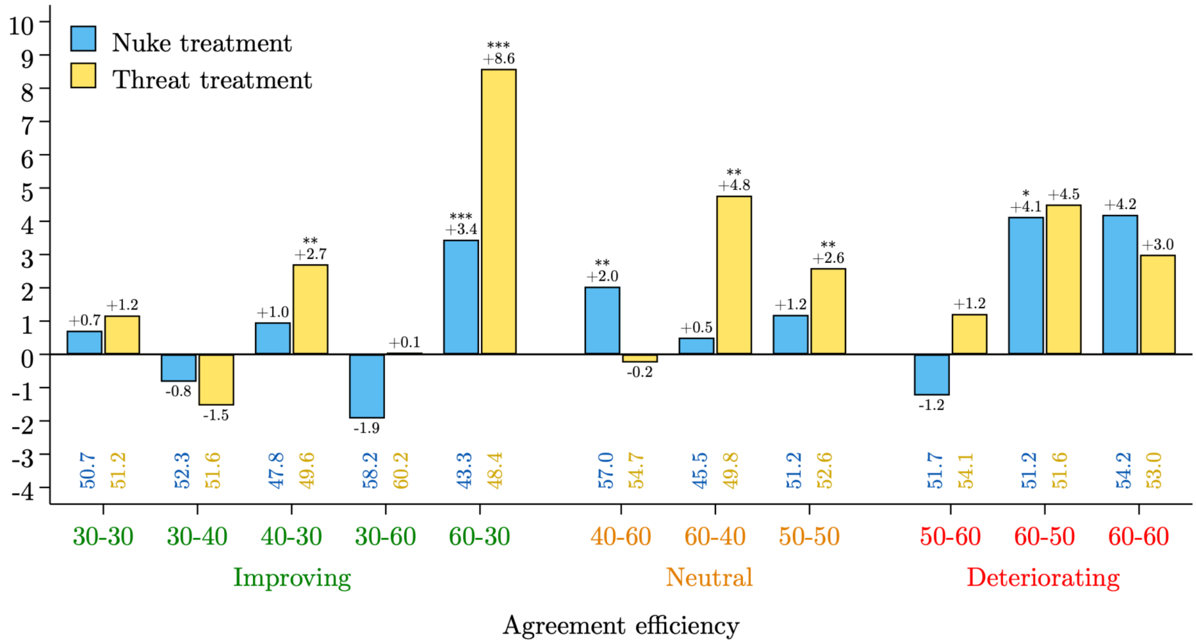
Thus, in summary, our data suggests that players react to the looming danger of the nuclear option and arrive at different bargaining outcomes. However, the nuclear threat does not encourage the negotiation partners to strike mutually beneficial deals but rather makes them more prone to enter into bad agreements where no mutual gains can be obtained. To make matters worse, sometimes the threat becomes reality, as some B-players actually trigger the nuclear option. The upshot is that the nuclear option systematically lowers efficiency, in line with Hypothesis 5.

Negotiation agreements in the presence of a nuclear threat. Our gloomy results on overall efficiency leave the question whether at least player B gains an advantage from holding the nuclear option. Figure 9 illustrates how agreements, when they are reached, change when the nuclear option is available and when it is not. As it turns out, player B can often achieve better, sometimes much better, agreements than in the Baseline treatment. While there are situations in which B-players perform on average worse compared to the Baseline, we fail to reject the null hypotheses in these cases. Aggregating over all scenarios, the improvement relative to Baseline is statistically significant for both the Nuke treatment ($p = 0.037$) and the Threat treatment ($p = 0.019$). Thus, it appears that, in line with Hypothesis 3, B-players can to some extent leverage their new power, at least when the negotiation ends in an agreement.

Intuitively and also in line with Hypothesis 3, the effectiveness of the nuclear option depends on player A's opportunity cost. This is visible in Figure 9 if we fix B's disagreement payoff and allow the disagreement payoff for player A to grow. For example, when player B's disagreement payoff is 30, an increase in A's disagreement payoff yields increasing returns to B. Table 9 exhibits the same pattern in all such cases. Holding the disagreement payoff for player B constant and raising that for player A, two counteracting forces are at play: On the one hand, as previously observed in the Baseline treatment, A's bargaining power should increase due to her improving outside option from a (non-nuclear) disagreement. But on the other hand, A's position in the negotiation also becomes more fragile because, as the stakes grow for player A, she has more and more to lose if player B did decide to go for the nuclear outcome. Although the second effect cannot fully absorb the first, it substantially takes away from what player A could have achieved in the absence of the nuclear option.

Thus, conditional on negotiations leading to agreements, B-players do appear to get some leverage out of being given the power to trigger a nuclear option. But does this also translate into an increase in B's *overall* payoff? Or do B-players obliterate their payoff advantage from their new bargaining power by (over-)using the nuclear option or by shifting their attention to striking agreements in the more inefficient settings? Figure 10 shows the *overall* payoff changes for player B—and player A—for all disagreement points. This figure provides more detailed insights into what the nuclear option does for

Figure 9: Changes in B-players' agreement payoffs relative to Baseline.



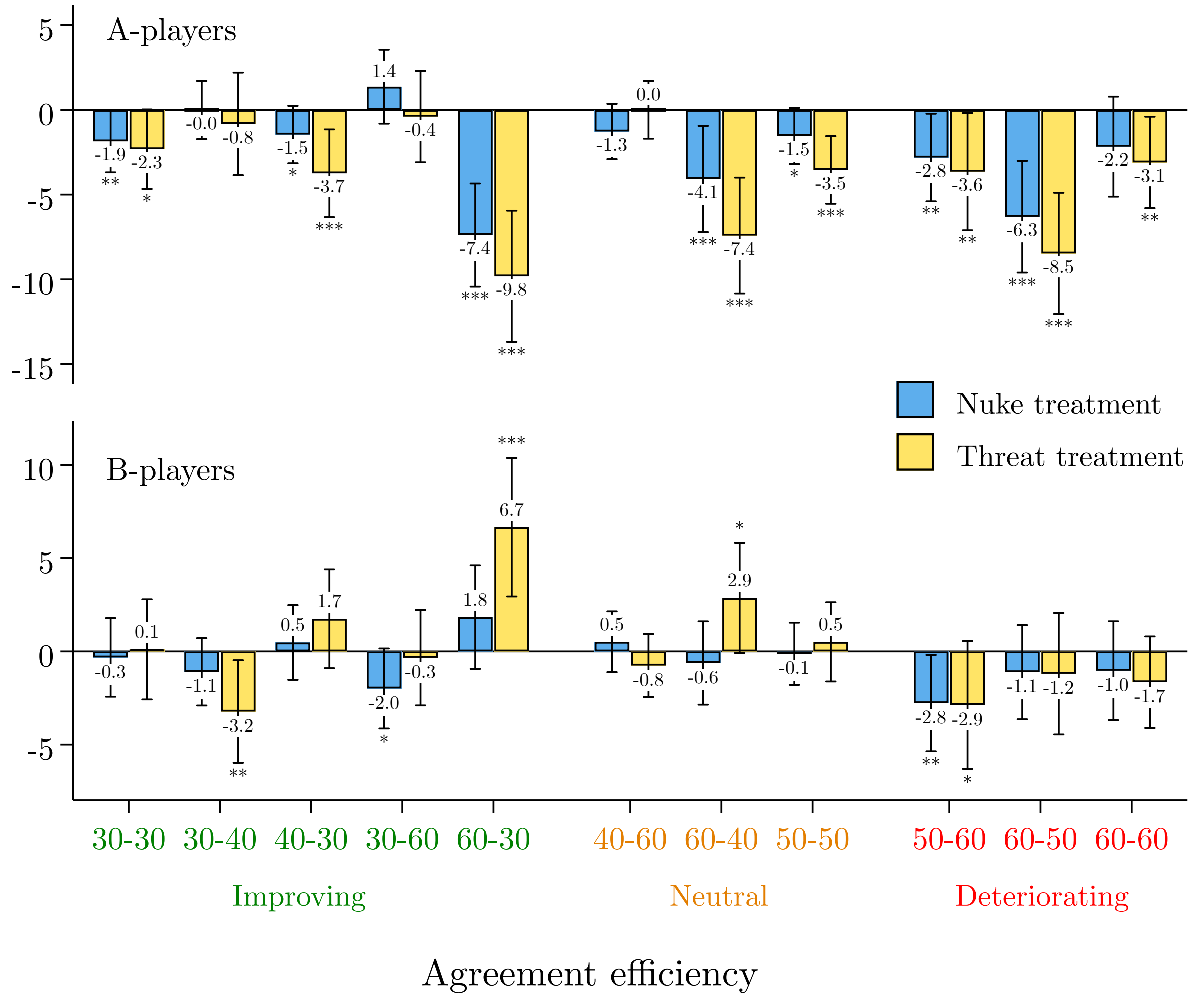
Notes: The figure shows the changes in B-players' payoffs conditional on agreement for each disagreement point relative to the Baseline treatment. The numbers at the bottom denote B-players' absolute mean payoffs in the Nuke and Threat treatments. Stars indicate significance (using non-parametric two-sided two-sample Fisher-Pitman permutation tests) at the 10% (*), 5% (**), and 1% (***) level.

each party, and how the joint payoff reductions shown above are distributed between A and B.

The willingness of B-players to use the nuclear option, the higher chance of agreements in situations where they are not beneficial, and the increase in B's relative bargaining power all paint a bleak picture for B's negotiation partner. As a result, A-players make a significant loss in their overall payoffs almost across the board. When aggregating over all disagreement points, we find a large and significant decrease in player A's payoff compared to Baseline (-2.5 points, $p < 0.001$). Once again, the two scenarios of 60-30 and 30-60 show a stark difference between situations in which B is likely to be reluctant to use the nuclear option and situations in which B might be more willing to go nuclear. In the 30-60 scenario, where the nuclear option is particularly costly for B and has the least worst impact on his partner, A-players can even improve their payoffs slightly (Nuke vs. Baseline) even though the difference is not statistically significant ($p = 0.217$). In contrast, in the opposite scenario, 60-30, A-players have to endure substantial losses in their overall payoffs.

Perhaps somewhat surprisingly, the situation is not much better for player B. Even though B-players enjoy increased bargaining power due to the nuclear option, the willingness of B-players to (occasionally) actually trigger the nuclear option leads to considerable payoff losses and thus, B-player payoffs do mostly not improve and in the few cases where they do, they are not significantly different from zero (with two exceptions in the Threat treatment). When aggregating over all disagreement points again, we find a (non-significant) decrease in B-players' average payoff compared to the Baseline treatment (0.6 points, $p = 0.204$). All in all, we cannot confirm that equipping B-players with a nuclear option improves their overall payoff. Conversely, the overall *reduction in joint payoffs* we discussed above is mostly borne by player A.

Figure 10: Who gains, who loses?—Overall payoff changes relative to Baseline.



Notes: The figure shows changes in overall payoffs (with and without agreements) for A-players (top panel) and B-players (bottom panel) for each disagreement point relative to the Baseline treatment. The 95% confidence intervals are based on independent matching groups. Stars indicate significance (using non-parametric two-sided two-sample Fisher-Pitman permutation tests) at the 10% (*), 5% (**), and 1% (***) level.

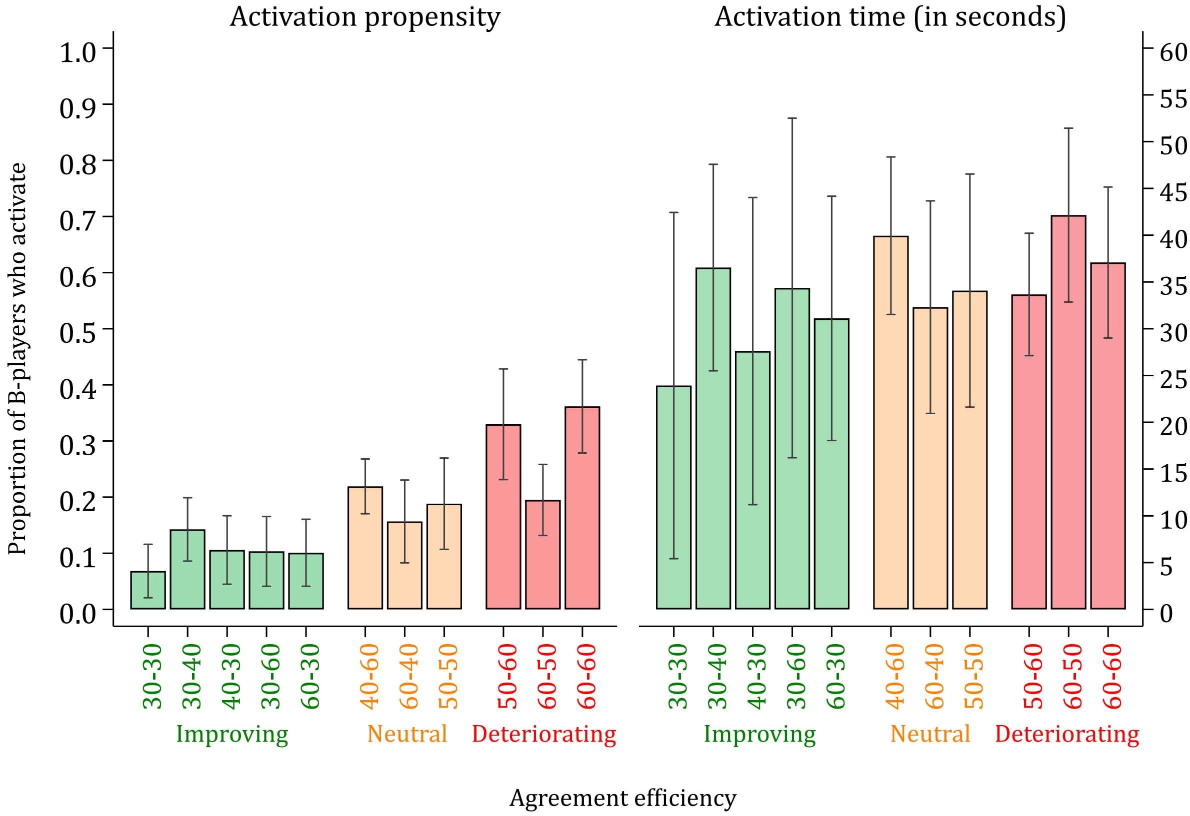
To summarize, the results from the Nuke treatment are in line with Hypothesis 4 (higher agreement rates than in Baseline)—at least for cases that are not characterized by efficiency-improving agreements—and with Hypothesis 5 (lower efficiency than in Baseline). Hypothesis 3, however, is not confirmed: Player B is unable to significantly improve his bargaining outcome with the help of the nuclear option. We provide further support for these findings by regression analyses in the appendix (Tables 8, 10, and 12 report analyses of Hypotheses 3, 4, and 5).

5.4 Threat treatment: Activating the nuclear option

Our Threat treatment introduces an additional step that player B must execute before he can trigger the nuclear option: He first has to *activate* the option, either ex post when no agreement has been achieved, or earlier while the negotiation is still ongoing. Since an early activation can be interpreted as a signal to player A, there is the potential for player B to use the nuclear option more strategically than in the Nuke treatment. In this section, we will examine whether endogenizing the availability of the nuclear option in this way is important for how negotiations proceed.

Over the course of the experiment, 57% of B-players activate the nuclear option at least once, and on average A-players see the nuclear option being activated by their partner in 17% of negotiation rounds. In another 10% of cases, B-players activate it ex post.[20] The median activation time is 39 seconds into the negotiation. Figure 11 depicts the propensity to activate the nuclear option during the negotiation and the time of activation across the different scenarios. While there is relatively little variation in the timing of the nuclear option, the propensity to activate does seem to depend on the nature of the disagreement point. Table 13 in the appendix reports the results of a regression of activation propensity and activation time on the disagreement payoffs. We find that the disagreement payoffs do not appear to have an influence on the *activation time* but do affect the *activation propensity*. In a nutshell, we find that the higher the disagreement payoffs, the more likely it is that the nuclear option is activated. The probability of an activation is lowest when agreements are efficiency-improving, and highest when the sum of disagreement payoffs exceeds the 100-token pie. This suggests that B-players use the activation decision strategically: In the scenarios in which—according to the Baseline treatment data—agreements can be found easily, they rarely activate the nuclear option; but in cases in which agreements are not the natural outcome they seem to be more inclined to send a message by putting the nuclear option on the negotiation table.

Figure 11: Propensity and timing of the activation of the nuclear option.

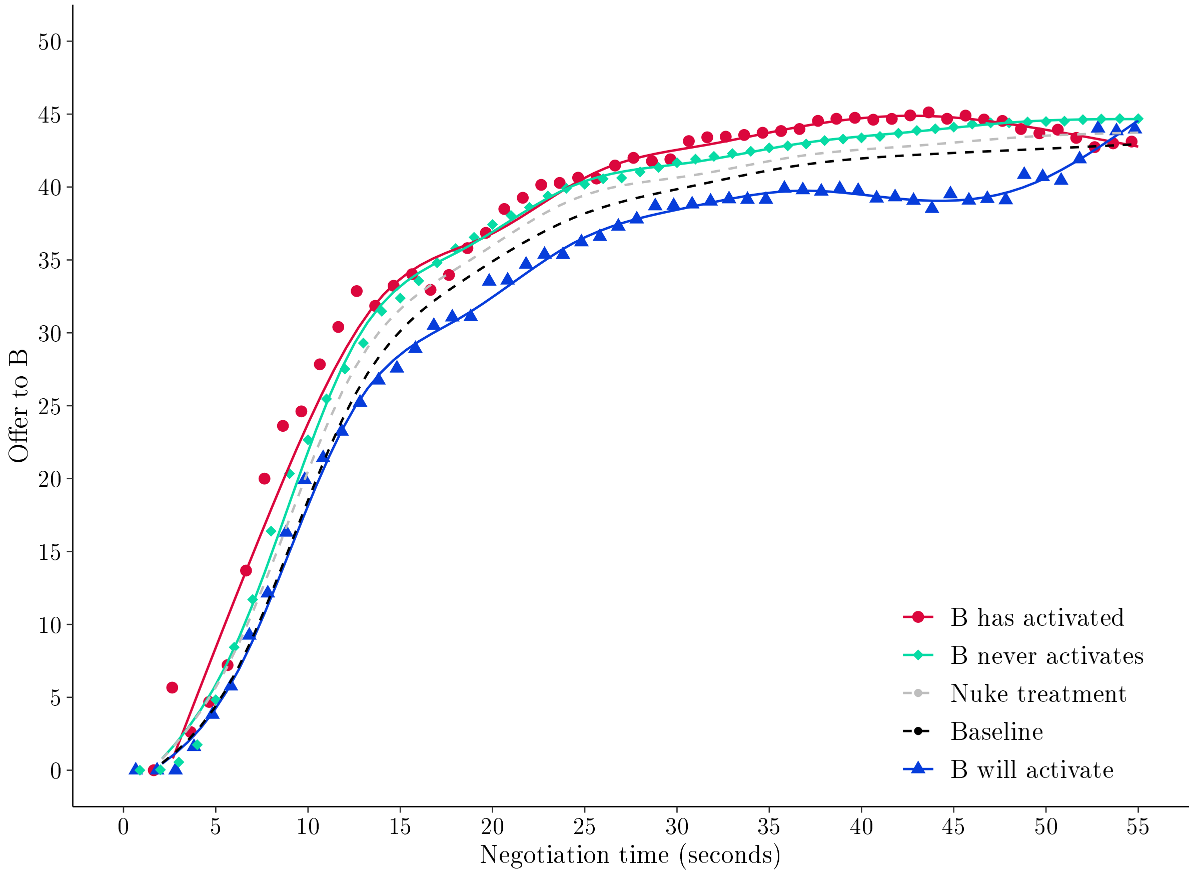


Notes: The figure depicts the activation behavior in the Threat treatment. The left panel displays the propensity to activate the nuclear option during the negotiation (i.e., the extensive margin). The right panel shows the time of activation conditional on activation (i.e., the intensive margin). Error bars denote 95% confidence intervals based on independent matching groups.

[20] As A-players are not informed about ex-post activations, this type of activation has no signalling value.

Does activating the nuclear option affect the course of the negotiation and its outcome? To a casual observer, it might seem as if activations backfire: Agreement rates and B-players' earnings relative to their disagreement payoffs are significantly *lower* when the option is activated than when it is not ($p < 0.001$ for both comparisons). However, closer inspection of the data reveals a different picture. As we have already seen, B-players do not activate the option at random, and as a consequence the low agreement rates and payoffs say more about already existing tensions in the negotiation than what the effects of activating the nuclear option are. It turns out that offers from A-players that *precede* an activation by player B are lower than average ($p < 0.001$) but improve substantially, by about 8 tokens, post-activation ($p < 0.001$). While this already gives some indication that activations are effective, it is important to control for time here to compare like with like. Consider therefore Figure 12 which plots offers made by Player A over time for encounters in which the nuclear option is or is not activated during the negotiation. In cases where it is, we differentiate between offers before and after the activation.

Figure 12: Offers from player A to player B by time and activation.



Notes: The figure depicts A-players' offers over time for the Threat treatment. Dots represent means at each point in time. Lines denote GAM-estimated regression lines through the dots. The dashed lines in gray and black show the offers for the Nuke and Baseline treatments for comparison.

As one might expect, negotiators (A-players in this case) begin by trying out proposals that, if accepted, would give themselves the lion's share of the pie. Following that initial phase, offers quickly rise to more reasonable levels. As the negotiation proceeds, they increase further but at a much slower pace. While this describes how a negotiation proceeds on average, there are notable differences between the different types of negotiations shown in Figure 12. When the nuclear option has not been activated *yet*, offers from player A are considerably lower than in negotiations in which the activation has either already occurred or is never going to occur. This suggests that B-players activate the nuclear option in response to receiving low offers and that A-players improve their offers because of this. When A-players behave in a more compromising way, it appears that B-players do not see the need to threaten them by activating the nuclear option.

Are B-players in the Threat treatment able to utilize the power of the nuclear option more effectively than their counterparts in the Nuke treatment? To examine this in more detail, we return to our analysis from Section 5.3. The overall impression is that the effects of the nuclear option are largely confirmed and sometimes even magnified. Average *total payoffs* decrease even more than in the Nuke treatment, with a fall of 3.7% relative to Baseline compared to the 3% drop we found for Nuke.[21] Likewise, *agreement rates* are even higher in the Threat treatment (+11 percentage points for Threat-vs.-Baseline compared to +6 percentage points for Nuke-vs.-Baseline). The differences between Threat and Baseline for these aggregate measures are highly significant ($p < 0.001$) although the differences between Threat and Nuke are not ($p = 0.562$ for total payoffs; $p = 0.134$ for agreement rates). Going back to Figure 7, we see that agreement rates in Threat are particularly high in the Pareto-deteriorating scenarios, and indeed this is where we do find a statistically significant treatment effect between Threat and Nuke ($p = 0.018$). The consequences of this are depicted in Figure 8. While payoffs are lower than in Baseline in both nuclear treatments, inefficient agreements (as opposed to direct nuclear damage) play a greater role in the Threat treatment than in the Nuke treatment in the efficiency-deteriorating situations. In these scenarios the joint payoffs are significantly lower in the Threat treatment than in the Nuke treatment ($p = 0.018$).

Finally, compared to Baseline, B-players are again unable to utilize the nuclear option for greater earnings ($p = 0.714$) and A-players again suffer a substantial loss in their overall earning ($p < 0.001$), but it is not worse for them than in the Nuke treatment ($p = 0.192$).

In summary, we find that enabling player B to use the nuclear option as an explicit threat slightly aggravates the welfare reducing effects we found earlier. Nuclear outcomes are just as frequent as in the Nuke treatment and the negotiators are steered even more towards inefficient agreements, resulting in payoff losses for player A without yielding payoff gains for player B. Thus, introducing the possibility of a more strategic utilization of the nuclear option does not help to mitigate the efficiency losses, let alone lead to better outcomes. Instead, the results from the Threat treatment just corroborate the notion that the availability of a nuclear option in bargaining is unambiguously detrimental.

5.5 Motives

To get a better understanding of what drives our results, we will now examine results from the post-experimental questionnaire. We administered the questionnaire to get a sense of why the nuclear option might play a role in the negotiations.[22] Specifically, we asked participants to tell us what they considered the main reason for using the nuclear option (see Online Appendix D.2 for the exact wording). This provides us with some suggestive evidence on the motives underlying the (expected) usage of the nuclear

[21] The attained efficiency level relative to (in)efficient negotiation outcomes (as defined in Section 5.3) is 53% in the Threat treatment, compared to 84% in Baseline and 57% in Nuke.

[22] In the following, we pool the data from the Nuke and Threat treatments. Figures 15-17 showing the questionnaire outcomes are included in the Online Appendix.

option. Irrespective of their assigned role, participants think that others would use the nuclear option mainly due to anger (44%). The two other most common indicated reasons for using the nuclear option are spite (ensuring that player B earns more than player A, 21%), or retaliation (teaching player A a lesson, 18%). Very few participants believe that the main motive for using the nuclear option is to try it out (11%) or by accident (6%).

The actual motives, i.e., the motives B-players state for their own use of the nuclear option, differ substantially both from A-players' beliefs and from what B-players think might be the motives for *others* in the role of player B ($\chi^2(4) = 52.9$, $p < 0.001$). Specifically, we see that the most common motives for player B are retaliation (29%), curiousness (26%), and anger (22%). Spite is the least reported motive (12%).

Most importantly, we see that A-players believe anger and retaliation to be important drivers for the nuclear option usage. This might explain the increased chance of reaching an agreement during the negotiations as A-players may not have wanted to trigger a negative emotional state in player B that might result in the use of the nuclear option.

Table 11 in the appendix shows that agreement rates strongly correlate with participants' predictions of the trigger frequencies. Furthermore, as we have seen, agreements are generally less favorable for A-players in the nuclear treatments than in the baseline. Taken together, this suggests that, on average, the fear of a nuclear outcome tends to make A-players more inclined to compromise during the negotiation, which naturally increases the chances of coming to an agreement. The nuclear threat is effective.

6 Conclusion

This paper uses a controlled experiment to study the strategic value of a nuclear option as well as its impact on efficiency in a bargaining setting. We design this option as an extremely destructive and clearly dominated additional action that one negotiation partner can take after a negotiation has ended without agreement. Our experiment aims at isolating the effects of the nuclear option in a simple situation that can be seen as an abstract representation of many bargaining settings in which nuclear options are explicitly or implicitly present.

The central insight of our study is that the introduction of a nuclear option systematically reduces overall efficiency. Unsurprisingly, the loss is most severe for the negotiation party not in control of the nuclear option. Interestingly, however, the nuclear option holder does not benefit from this. Thus, the threat of the nuclear option does not overall generate better payoffs for the party with the power to go nuclear. This challenges the belief that making 'nuclear' threats leads to a strategic advantage in negotiations. The reason for this result is twofold. First, the nuclear threat does not encourage the negotiation partners to strike mutually beneficial deals but rather makes them prone to enter into bad agreements where no mutual gains can be obtained. Second, the threat of a nuclear outcome can become reality as nuclear option holders sometimes appear to enter a mindset that makes them feel compelled to actually trigger the option, which obviously harms both negotiation partners. Our Threat treatment reveals that matters do not improve if the nuclear option holder is given a way of communicating a willingness to trigger. In fact, the detrimental effect of the nuclear option tends to be even more pronounced, highlighting that the nuclear option systematically lowers efficiency.

This paper is a first attempt to study the effect of a nuclear option. While our goal is to conceptually capture the nature of such an option in a wide range of settings, there are some limitations to our design. The scale of the nuclear option is necessarily different in the experiment than in many real-world situations. We represent a nuclear option by giving one negotiation partner the choice to drastically reduce payoffs below the original disagreement point, but we cannot fully replicate the dimensions that nuclear options might take in the real world. There, consequences of a nuclear option could be life-altering for individuals or have long-lasting effects for whole countries. We would expect the larger scale of consequences to reduce the willingness of bargaining partners to trigger the nuclear option. However, on the other hand, triggering would then also increase the harm done by the nuclear option. Thus, it is not clear in how far real-world situations would actually differ from the experiment in expected terms. Given the impossibility of mirroring real nuclear options in scale, we believe that our study provides valuable insights into general behavioral patterns in bargaining settings with a dominated additional choice that is highly destructive. Moreover, designing the nuclear option to be detrimental but not life-altering enables us to speak to all the situations where a figurative nuclear, with negative, but not life-altering, consequences are present.

Our study is a conservative first test of the effect of making nuclear options explicit in bargaining situations. In the Nuke treatment, the nuclear option holder has no way of communicating their intentions or using the nuclear option as an explicit threat. Thus, this treatment captures the effect of a nuclear option merely being available. The Threat treatment does offer a first step in using the nuclear option to put pressure on the negotiation partner. However, in both treatments, participants are well informed about the nuclear option, and it is prominently presented on screen. In many real-world situations, it is plausible that negotiation partners are not aware of the nuclear option or do not fully consider it until it is mentioned as a threat. An arguing couple is not paying any attention to the possibility of divorce until one of them brings it up, and a political coalition will negotiate without the thought of ending a coalition until one member brings it to the table. The possible unexpectedness of the nuclear option in those situations might intensify the threat of the option being brought up. In an experiment without deception it is very difficult to make a nuclear option available without drawing immediate attention to it. Nonetheless, our data shows that the option to actively use the nuclear option as a threat leads to outcomes that are qualitatively identical to and sometimes even more drastic than those emerging in a setting in which the nuclear option is merely present. Bringing up a nuclear option unexpectedly is conceptually a more extreme version of this treatment, and it would seem plausible that our results would escalate further in a surprise-threat setting. In addition, the nuclear option in our design is randomly assigned to one player and essentially not earned. Such a design choice was inevitable in order to cleanly study the effects of the nuclear option without issues of self-selection. However, recent evidence indicates that earning the position of power might shift the balance from the weaker to the stronger player even further (Anbarci and Feltovich, 2013, 2018; Feltovich, 2019).

This paper contributes to theoretical and experimental research on bargaining behavior by showing the importance of considering all possible choices that may be available to negotiators, even when they are clearly dominated. We also generate insights for practitioners who are interested in successful negotiation strategies. Popular belief holds that 'tough' negotiators who do not shy away from making (figurative) nuclear threats are more successful in a bargaining context. Our results caution against this strategy. While the party not in control of the nuclear option seems to be threatened to some degree and adjusts bargaining behavior, this does not generate the desired effect of higher payoffs. Admittedly, our participants are not professional negotiators and are unlikely to pre-plan negotiations to a degree that political representatives or similarly experienced negotiators would. It is possible that more experienced participants would be able to use the nuclear option more effectively. However, our MTurk sample represents a heterogeneous group of people and allows us to observe general behavior in a general setting. Many bargaining settings do not—or do not exclusively—consist of highly experienced negotiators. In a company setting, managers regularly are involved in negotiations and often interact with less experienced negotiators. For example, employees need to bargain over their contracts and wages, and small suppliers are also not used to bargaining at a large scale. Our results show that threatening the other party with a nuclear option, e.g., termination of contract, might not yield the expected result of better outcomes. Instead, the threatened person is more likely to give in even if agreement does not lead to an overall efficiency enhancing situation.

In conclusion, our research underscores the detrimental impact of a (figurative) nuclear option. The mere presence of such an option does not confer any distinct advantage to its holder—instead, it significantly compromises overall efficiency, compelling parties to strike an agreement when it would be more optimal not to. While we deliberately focused only on one-shot interactions, future research should delve into the enduring dynamics of similar settings and investigate, for example, how reputation effects impact on negotiation outcomes over time.

References

Abbink, Klaus and Abdolkarim Sadrieh, “The pleasure of being nasty,” *Economics Letters*, 2009, *105* (3), 306 – 308.

_ **and Benedikt Herrmann**, “The moral costs of nastiness,” *Economic Inquiry*, 2011, *49* (2), 631–633.

Ali, S. Nageeb, Navin Kartik, and Andreas Kleiner, “Sequential Veto Bargaining With Incomplete Information,” *Econometrica*, 2023, *91* (4), 15271562.

Anbarci, Nejat and Nick Feltovich, “How sensitive are bargaining outcomes to changes in disagreement payoffs?,” *Experimental Economics*, December 2013, *16* (4), 560–596.

_ **and** _ , “How fully do people exploit their bargaining position? The effects of bargaining institution and the 5050 norm,” *Journal of Economic Behavior & Organization*, January 2018, *145*, 320–334.

Andreoni, James, William Harbaugh, and Lise Vesterlund, “The carrot or the stick: Rewards, punishments, and cooperation,” *American Economic Review*, 2003, *93* (3), 893–902.

Avery, Christopher and Peter B. Zemsky, “Money Burning and Multiple Equilibria in Bargaining,” *Games and Economic Behavior*, September 1994, *7* (2), 154–168.

Banerjee, Subrato, “Effect of reduced opportunities on bargaining outcomes: an experiment with status asymmetries,” *Theory and Decision*, 2020, *89* (3), 313–346.

Battaglini, Marco, “Coalition Formation in Legislative Bargaining,” *Journal of Political Economy*, November 2021, *129* (11), 32063258.

Berg, Janine, “Income security in the on-demand economy: Findings and policy lessons from a survey of crowdworkers,” *Comp. Lab. L. & Pol'y J.*, 2015, *37*, 543.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz, “Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk,” *Political Analysis*, 2012, *20* (3), 351.

Bolton, Gary E and Axel Ockenfels, “ERC: A Theory of Equity, Reciprocity, and Competition,” *The American Economic Review*, 2000, *90* (1), 166–193.

_ **and Emin Karagözoğlu**, “On the influence of hard leverage in a soft leverage bargaining game: The importance of credible claims,” *Games and Economic Behavior*, 2016, *99*, 164–179.

Bosman, Ronald and Frans Van Winden, “Emotional hazard in a power-to-take experiment,” *The Economic Journal*, 2002, *112* (476), 147–169.

_ **, Matthias Sutter, and Frans van Winden**, “The impact of real effort and emotions in the power-to-take game,” *Journal of Economic Psychology*, 2005, *26* (3), 407–429.

Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling, "Amazon's Mechanical Turk," *Perspectives on Psychological Science*, 2011, *6* (1), 3–5.

Camerer, Colin F, *Behavioral game theory: Experiments in strategic interaction*, Princeton University Press, 2011.

_ **and Richard H Thaler**, "Anomalies: Ultimatums, dictators and manners," *Journal of Economic perspectives*, 1995, *9* (2), 209–219.

Camerer, Colin F., Gideon Nave, and Alec Smith, "Dynamic Unstructured Bargaining with Private Information: Theory, Experiment, and Outcome Prediction via Machine Learning," *Management Science*, May 2018, *65* (4), 1867–1890.

Chen, Hsiao-Chi, James W Friedman, and Jacques-François Thisse, "Boundedly rational Nash equilibrium: A probabilistic choice approach," *Games and Economic Behavior*, January 1997, *18* (1), 32–54.

Chen, Yefen, Xuanming Su, and Xiaobo Zhao, "Modeling Bounded Rationality in Capacity Allocation Games with the Quantal Response Equilibrium," *Management Science*, October 2012, *58* (10), 19521962.

das Varma, Gopal, "Standard Auctions with Identity-Dependent Externalities," *The RAND Journal of Economics*, 2002, *33* (4), 689.

Deneckere, Raymond and Meng-Yu Liang, "Bargaining with Interdependent Values," *Econometrica*, September 2006, *74* (5), 13091364.

Doepke, Matthias and Fabian Kindermann, "Bargaining over Babies: Theory, Evidence, and Policy Implications," *American Economic Review*, September 2019, *109* (9), 3264–3306.

Edgeworth, Francis Ysidro, *Mathematical psychics: An essay on the application of mathematics to the moral sciences*, Vol. 10, Kegan Paul, 1881.

Embrey, Matthew, Kyle Hyndman, and Arno Riedl, "Bargaining with a residual claimant: An experimental study," *Games and Economic Behavior*, mar 2021, *126*, 335–354.

Exley, Christine L., Muriel Niederle, and Lise Vesterlund, "Knowing When to Ask: The Cost of Leaning In," *Journal of Political Economy*, March 2020, *128* (3), 816854.

Feltovich, Nick, "Is earned bargaining power more fully exploited?," *Journal of Economic Behavior & Organization*, nov 2019, *167*, 152–180.

Galeotti, Fabio, "Do negative emotions explain punishment in power-to-take game experiments?," *Journal of Economic Psychology*, 2015, *49*, 1–14.

Güth, Werner, "On ultimatum bargaining experimentsA personal review," *Journal of Economic Behavior & Organization*, 1995, *27* (3), 329–344.

Gächter, Simon and Arno Riedl, "Moral Property Rights in Bargaining with Infeasible Claims," *Management Science*, February 2005, *51* (2), 249–263.

Hoffman, Elizabeth and Matthew L. Spitzer, "Entitlements, Rights, and Fairness: An Experimental Examination of Subjects' Concepts of Distributive Justice," *Journal of Legal Studies*, 1985, *14* (2), 259–298.

Jr, John F Nash, "The bargaining problem," *Econometrica: Journal of the Econometric Society*, 1950, *18* (2), 155–162.

Kalai, Ehud and Meir Smorodinsky, "Other solutions to Nashs bargaining problem," *Econometrica*, 1975, *43* (3), 513–518.

Karagözoğlu, Emin, "On "Going Unstructured" in Bargaining Experiments," *Studies in Economic Design*, 2019, pp. 295 – 304.

_ **and Arno Riedl**, "Performance Information, Production Uncertainty, and Subjective Entitlements in Bargaining," *Management Science*, December 2014, *61* (11), 2611–2626.

Kirchkamp, Oliver and Wladislaw Mill, "Spite vs. risk: Explaining overbidding in the second-price all-pay auction," *Games and Economic Behavior*, November 2021, *130*, 616–635.

Klor, Esteban F. and Moses Shayo, "Social identity and preferences over redistribution," *Journal of Public Economics*, 2010, *94* (3), 269 – 278.

Kocher, Martin G., Odile Poulsen, and Daniel J. Zizzo, "Social preferences, accountability, and wage bargaining," *Social Choice and Welfare*, February 2017, *48* (3), 659678.

Larsen, Bradley J, "The Efficiency of Real-World Bargaining: Evidence from Wholesale Used-Auto Auctions," *The Review of Economic Studies*, February 2020, *88* (2), 851882.

Levine, David K., "Modeling Altruism and Spitefulness in Experiments," *Review of Economic Dynamics*, 1998, *1* (3), 593 – 622.

Lozano, Lina, Arno M. Riedl, and Christina Rott, "The Impact of the Menstrual Cycle on Bargaining Behavior," *SSRN Electronic Journal*, 2024.

Manzini, Paola, "Strategic bargaining with destructive power," *Economics Letters*, December 1999, *65* (3), 315–322.

Merkel, Anna and Christoph Vanberg, "Multilateral bargaining with subjective claims under majority vs. unanimity rule: An experiment," *Journal of Economic Psychology*, mar 2023, *95*, 102601.

Mill, Wladislaw and John Morgan, "The cost of a divided America: an experimental study into destructive behavior," *Experimental Economics*, November 2021, *25* (3), 974–1001.

_ **and** _ , "Competition between friends and foes," *European Economic Review*, August 2022, *147*, 104171.

_ **and Jonathan Stäbler**, "Spite in litigation," 2022. Mimeo.

Miller, A. David and Joel Watson, “A Theory of Disagreement in Repeated Games With Bargaining,” *Econometrica*, 2013, *81* (6), 2303–2350.

Miller, Luis and Christoph Vanberg, “Group size and decision rules in legislative bargaining,” *European Journal of Political Economy*, March 2015, *37*, 288302.

Nash, John, “Two-Person Cooperative Games,” *Econometrica*, 1953, *21* (1), 128–140.

Oosterbeek, Hessel, Randolph Sloof, and Gijs Van De Kuilen, “Cultural differences in ultimatum game experiments: Evidence from a meta-analysis,” *Experimental Economics*, 2004, *7* (2), 171–188.

Paolacci, Gabriele, Jesse Chandler, and Panagiotis G. Ipeirotis, “Running experiments on Amazon Mechanical Turk,” *Judgment and Decision Making*, 2010, *5* (5), 411–419.

Ravid, Doron, “Ultimatum Bargaining with Rational Inattention,” *American Economic Review*, September 2020, *110* (9), 29482963.

Roth, Alvin E, “Bargaining Experiments chapter 5 in John H. Kagel and Alvin E. Roth, editors, The Handbook of Experimental Economics,” 1995.

Silveira, Bernardo S., “Bargaining With Asymmetric Information: An Empirical Study of Plea Negotiations,” *Econometrica*, 2017, *85* (2), 419452.

Thompson, Leigh L, Jiunwen Wang, and Brian C Gunia, “Negotiation,” *Annual review of psychology*, 2010, *61*, 491–515.

Tremewan, James and Christoph Vanberg, “The dynamics of coalition formation – A multilateral bargaining experiment with free timing of moves,” *Journal of Economic Behavior & Organization*, oct 2016, *130*, 33–46.

Appendix

Table 2: Regression analysis — Agreement rates (Hyp. 1a).

	Agreement rate				
	All cases		Improving	Neutral	Deteriorating
	(1)	(2)	(3)	(4)	(5)
Constant	0.894***	0.621***	0.813***	0.535***	0.286***
	(0.019)	(0.015)	(0.021)	(0.027)	(0.085)
$d_i + d_j$		−0.258***	−0.088***		−0.034
		(0.010)	(0.017)		(0.070)
$d_i + d_j = 100$	−0.359***				
	(0.024)				
$d_i + d_j > 100$	−0.647***				
	(0.024)				
Sbj.Spec.Effects	✓	✓	✓	✓	✓
Group.Spec.Effects	✓	✓	✓	✓	✓
Observations	1,448	1,448	658	396	394
Log Likelihood	−716.852	−751.737	−151.456	−287.350	−216.938
Bayesian Inf. Crit.	1,477.372	1,539.864	335.358	598.626	463.758
Note:	$^{+}$p<0.10;*p<0.05;**p<0.01;***p<0.001;				

This table presents the results of regressions testing for Hypothesis 1a: the agreement rate in the baseline treatment. The first two columns list all the combinations of disagreement points. The third, fourth, and fifth columns show the combinations of disagreement points where the sum is below, equal to, or above 100 (improving, neutral, and deteriorating, respectively). $d_i + d_j$ denotes the normalized sum of disagreement payoffs (i.e., a one-unit increase denotes an one-standard error increase in the sum of disagreement payoffs). *Sbj.Spec.Effects* refers to subject-specific random effects, while *Group.Spec.Effects* denotes matching-group specific random effects.

Table 3: Regression analysis — Agreement payoff (Hyp. 1b).

	Agreement payoff			
	All cases	Improving	Neutral	Deteriorating
	(1)	(2)	(3)	(4)
Constant	48.438***	49.618***	26.096***	56.724*
	(1.288)	(1.700)	(3.127)	(28.150)
d_i	0.337***	0.341***	0.484***	0.226
	(0.020)	(0.026)	(0.061)	(0.284)
d_j	−0.292***	−0.330***		−0.322
	(0.020)	(0.026)		(0.286)
Sbj.Spec.Effects	✓	✓	✓	✓
Group.Spec.Effects	✓	✓	✓	✓
Observations	1,794	1,176	424	194
Log Likelihood	−6,860.431	−4,345.281	−1,661.129	−833.848
Bayesian Inf. Crit.	13,765.820	8,732.982	3,352.508	1,699.304
Note:	$^{+}$p<0.10;*p<0.05;**p<0.01;***p<0.001;			

This table presents the results of regressions testing for Hypothesis 1b: the payoff conditional on agreement in the baseline. The first two columns list all the combinations of disagreement points. The third, fourth, and fifth columns show the combinations of disagreement points where the sum is below, equal to, or above 100 (improving, neutral, and deteriorating, respectively). d_i, and d_j refer to the own and the others disagreement payoff, respectively. *Sbj.Spec.Effects* refers to subject-specific random effects, while *Group.Spec.Effects* denotes matching-group specific random effects.

Table 4: Regression analysis — Trigger rates (Hyp. 2).

	Nuclear trigger rate							
	Both nuclear treatments				Nuke only		Threat only	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.029***	0.111***	0.028**	0.107***	0.028***	0.106***	0.030**	0.117***
	(0.006)	(0.017)	(0.009)	(0.023)	(0.008)	(0.023)	(0.010)	(0.025)
d_B	−0.005+	−0.076***	−0.005	−0.079***	−0.005	−0.078***	−0.005	−0.074***
	(0.003)	(0.011)	(0.004)	(0.015)	(0.004)	(0.015)	(0.004)	(0.017)
d_A	0.016***	0.030**	0.019***	0.032*	0.019***	0.032*	0.013**	0.027+
	(0.003)	(0.011)	(0.004)	(0.015)	(0.004)	(0.014)	(0.004)	(0.016)
Threat			0.003	0.009				
			(0.013)	(0.033)				
d_B x Threat			0.0003	0.005				
			(0.006)	(0.022)				
d_A x Threat			−0.006	−0.006				
			(0.006)	(0.021)				
Sbj.Spec.Effects	✓	✓	✓	✓	✓	✓	✓	✓
Group.Spec.Effects	✓	✓	✓	✓	✓	✓	✓	✓
Cond. on disagreement	No	Yes	No	Yes	No	Yes	No	Yes
Observations	2,478	727	2,478	727	1,239	391	1,239	336
Log Likelihood	1,077.938	−59.234	1,066.579	−67.533	532.784	−15.263	536.811	−50.012
Bayesian Inf. Crit.	−2,108.985	158.001	−2,062.821	194.366	−1,022.836	66.339	−1,030.890	134.928

$^{+}$p<0.10;*p<0.05;**p<0.01;***p<0.001;

This table presents the results of regressions testing for Hypothesis 2 (the trigger rate of the nuclear option). The first four columns include both nuclear treatments. Columns (5) and (6) [columns (7) and (8)] restrict the sample to the Nuke [Threat] treatment. Columns (2), (4), (6), and (8) focus on failed negotiations. d_A, and d_B refer to the normalized disagreement payoff of player A, and player B, respectively. Thus, a one-unit increase of d_A denotes an one-standard error increase in the disagreement payoff of player A. *Sbj.Spec.Effects* refers to subject-specific random effects, while *Group.Spec.Effects* denotes matching-group specific random effects.

Table 5: Regression of trigger rate predictions.

	Nuclear trigger predictions											
	Overall trigger predictions (in %)				Making a nonzero prediction (extensive margin)				Non-zero trigger rate prediction (intensive margin)			
	Both nuclear treatments		Nuke only	Threat only	Both nuclear treatments		Nuke only	Threat only	Both nuclear treatments		Nuke only	Threat only
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Constant	20.448***	16.681***	16.671***	24.194***	0.678***	0.632***	0.632***	0.723***	27.882***	23.908***	23.952***	31.441***
	(1.038)	(1.453)	(1.326)	(1.561)	(0.017)	(0.023)	(0.024)	(0.023)	(1.189)	(1.705)	(1.646)	(1.685)
d_B	−2.021***	−2.474***	−2.474***	−1.568***	−0.094***	−0.107***	−0.107***	−0.081***	−0.583^{+}	−1.421**	−1.392**	0.085
	(0.235)	(0.332)	(0.330)	(0.335)	(0.004)	(0.006)	(0.006)	(0.006)	(0.308)	(0.456)	(0.466)	(0.408)
d_A	2.506***	2.728***	2.727***	2.281***	0.049***	0.057***	0.057***	0.041***	2.188***	2.656***	2.642***	1.808***
	(0.234)	(0.332)	(0.329)	(0.333)	(0.004)	(0.006)	(0.006)	(0.006)	(0.296)	(0.434)	(0.445)	(0.395)
Threat		7.501***				0.092**				7.548**		
		(2.050)				(0.033)				(2.361)		
d_B x Threat		0.906^{+}				0.026**				1.519*		
		(0.470)				(0.008)				(0.618)		
d_A x Threat		−0.445				−0.016^{+}				−0.851		
		(0.469)				(0.008)				(0.593)		
Sbj.Spec.Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Group.Spec.Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	4,955	4,955	2,477	2,478	4,955	4,955	2,477	2,478	3,325	3,325	1,550	1,775
Log Likelihood	−21,595.100	−21,584.360	−10,750.060	−10,830.690	−1,654.289	−1,654.630	−903.749	−745.723	−14,640.990	−14,630.160	−6,850.463	−7,777.678
Bayesian Inf. Crit.	43,241.250	43,245.290	21,547.010	21,708.270	3,359.627	3,385.833	1,854.386	1,538.338	29,330.630	29,333.290	13,745.000	15,600.250

Note: $^{+}$p<0.10;*p<0.05;**p<0.01;***p<0.001;

This table presents the results of regressions testing the predictions of trigger rates of the nuclear option. The first four columns focus on the predictions of the overall trigger rate. Columns (5)-(8) focus on the extensive margin, i.e. whether a triggering of the nuclear option was predicted at all. Columns (9)-(12) focus on the intensive margin, i.e. conditional on predicting a positive probability of trigger, the predictions of the trigger rate. Columns (1),(2),(5),(6),(9),(10) include both nuclear treatments. Columns (3), (7) and (11) [columns (4), (8) and (12)] restrict the sample to the Nuke [Threat] treatment. d_A, and d_B refer to the normalized disagreement payoff of player A, and player B, respectively. Thus, a one-unit increase of d_A denotes an one-standard error increase in the disagreement payoff of player A. *Sbj.Spec.Effects* refers to subject-specific random effects, while *Group.Spec.Effects* denotes matching-group specific random effects.

Table 6: Regression of differences in trigger rate predictions between A and B-players.

	Nuclear trigger predictions								
	Overall trigger predictions (in %)			Making a nonzero prediction (extensive margin)			Non-zero trigger rate prediction (intensive margin)		
	Both nuclear treatments	Nuke only	Threat only	Both nuclear treatments	Nuke only	Threat only	Both nuclear treatments	Nuke only	Threat only
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	20.518***	17.421***	23.588***	0.673***	0.626***	0.719***	28.433***	25.783***	30.719***
	(1.471)	(1.880)	(2.212)	(0.024)	(0.034)	(0.032)	(1.687)	(2.330)	(2.395)
d_B	−2.056***	−3.038***	−1.079*	−0.102***	−0.127***	−0.076***	−0.364	−1.546*	0.554
	(0.332)	(0.466)	(0.473)	(0.006)	(0.009)	(0.008)	(0.439)	(0.675)	(0.574)
d_A	2.610***	3.047***	2.177***	0.057***	0.065***	0.049***	2.082***	2.615***	1.678**
	(0.332)	(0.466)	(0.472)	(0.006)	(0.009)	(0.008)	(0.420)	(0.633)	(0.560)
A-player	−0.141	−1.500	1.213	0.010	0.012	0.008	−1.084	−3.675	1.439
	(2.079)	(2.656)	(3.128)	(0.033)	(0.048)	(0.045)	(2.380)	(3.291)	(3.376)
d_B x A-player	0.071	1.127^{+}	−0.977	0.015^{+}	0.041***	−0.010	−0.428	0.311	−0.950
	(0.470)	(0.659)	(0.669)	(0.008)	(0.012)	(0.012)	(0.616)	(0.935)	(0.817)
d_A x A-player	−0.208	−0.640	0.208	−0.015^{+}	−0.016	−0.015	0.207	0.059	0.265
	(0.469)	(0.659)	(0.667)	(0.008)	(0.012)	(0.012)	(0.592)	(0.891)	(0.790)
Sbj.Spec.Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Group.Spec.Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	4,955	2,477	2,478	4,955	2,477	2,478	3,325	1,550	1,775
Log Likelihood	−21,593.020	−10,745.210	−10,826.440	−1,661.419	−906.831	−753.660	−14,638.020	−6,845.994	−7,773.327
Bayesian Inf. Crit.	43,262.610	21,560.760	21,723.220	3,399.412	1,883.996	1,577.657	29,349.020	13,758.100	15,613.990

$^{+}$p<0.10;*p<0.05;**p<0.01;***p<0.001;

This table presents the results of regressions testing the difference in predictions of trigger rates of the nuclear option between A and B-players. The first three columns focus on the predictions of the overall trigger rate. Columns (4)-(6) focus on the extensive margin, i.e., whether a triggering of the nuclear option was predicted at all. Columns (7)-(9) focus on the intensive margin, i.e., conditional on predicting the nuclear option to be used with positive probability, the predicted trigger rate. Columns (1), (4), and (7) include both nuclear treatments. Columns (2), (5), and (8) [columns (3), (6), and (9)] restrict the sample to the Nuke [Threat] treatment. d_A and d_B refer to the normalized disagreement payoff of player A and player B, respectively. Thus, a one-unit increase of d_A denotes an one-standard error increase in the disagreement payoff of player A. *Sbj.Spec.Effects* refers to subject-specific random effects, while *Group.Spec.Effects* denotes matching-group specific random effects.

Table 7: Regression of the difference in predictions between own and other trigger rates (B-players only).

	Nuclear trigger predictions			
	Difference in predictions between own and other trigger rate (in %)			
	Both nuclear treatments		Nuke only	Threat only
	(1)	(2)	(3)	(4)
Constant	0.197	−1.442	−1.457	1.821*
	(0.762)	(1.071)	(1.241)	(0.866)
d_B	−1.885***	−1.380**	−1.378**	−2.388***
	(0.336)	(0.474)	(0.510)	(0.436)
d_A	0.376	0.135	0.134	0.612
	(0.335)	(0.474)	(0.510)	(0.435)
Threat		3.265*		
		(1.513)		
d_B x Threat		−1.008		
		(0.671)		
d_A x Threat		0.472		
		(0.670)		
Sbj.Spec.Effects	✓	✓	✓	✓
Group.Spec.Effects	✓	✓	✓	✓
Observations	2,478	2,478	1,239	1,239
Log Likelihood	−10,668.070	−10,662.090	−5,428.456	−5,212.821
Bayesian Inf. Crit.	21,383.030	21,394.530	10,899.650	10,468.370
Note:			$^{+}$p<0.10;*p<0.05;**p<0.01;***p<0.001;	

This table presents the results of regressions testing the difference in predictions between own and other trigger rates of the nuclear option for B-players. The dependent variable is the difference between the predicted trigger rate of other B-players and the own predicted trigger rate (i.e. positive values indicate a belief that for a given disagreement point the own trigger rate is higher than the trigger rate of other B-players). Columns (1),(2) include both nuclear treatments. Column (3) [column (4)] restrict the sample to the Nuke [Threat] treatment. d_A, and d_B refer to the normalized disagreement payoff of player A, and player B, respectively. Thus, a one-unit increase of d_A denotes an one-standard error increase in the disagreement payoff of player A. *Sbj.Spec.Effects* refers to subject-specific random effects, while *Group.Spec.Effects* denotes matching-group specific random effects.

Table 8: Regression analysis — Player B's overall payoff (Hyp. 3).

	Payoff player B							
	All cases		30-60		60-30		All cases	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant (Baseline)	50.689***	49.751***	60.568***	60.729***	38.758***	40.380***	50.689***	50.066***
	(0.543)	(0.606)	(0.906)	(1.075)	(1.308)	(1.253)	(0.550)	(0.628)
Threat treatment	0.174	2.691**	−0.951	−0.643	6.004**	8.203***	0.154	2.672**
	(0.796)	(0.877)	(1.323)	(1.565)	(1.918)	(1.837)	(0.805)	(0.897)
Nuclear treatment	−0.427	1.223	−2.475+	−2.496	1.177	3.013	−0.473	1.144
	(0.796)	(0.882)	(1.330)	(1.564)	(1.918)	(1.834)	(0.806)	(0.906)
d_B							5.719***	3.927***
							(0.269)	(0.336)
d_A							−3.157***	−3.855***
							(0.270)	(0.329)
d_B x Threat treatment							−1.543***	−1.189*
							(0.397)	(0.472)
d_A x Threat treatment							−0.592	−0.340
							(0.397)	(0.479)
d_B x Nuclear treatment							1.252**	2.400***
							(0.397)	(0.465)
d_A x Nuclear treatment							0.349	1.132*
							(0.397)	(0.473)
Sbj.Spec.Effects	✓	✓	×	×	×	×	✓	✓
Group.Spec.Effects	✓	✓	✓	✓	✓	✓	✓	✓
Conditional on agreement	×	✓	×	✓	×	✓	×	✓
Observations	3,926	2,648	362	295	358	303	3,926	2,648
Log Likelihood	−15,444.610	−10,276.970	−1,324.903	−1,083.440	−1,417.727	−1,152.721	−14,946.800	−10,028.290
Bayesian Inf. Crit.	30,938.870	20,601.230	2,679.264	2,195.315	2,864.857	2,334.011	29,992.900	20,151.150

Note: +p<0.10;*p<0.05;**p<0.01;***p<0.001;

This table presents the results of regressions testing for Hypothesis 3 (overall payoff of player B). The first two columns list all the combinations of disagreement points. The third, and fourth columns show the situations where player A has a disagreement payoff of 30 and player B has a disagreement payoff of 60. The fifth, and sixth columns show the situations where player A has a disagreement payoff of 60 and player B has a disagreement payoff of 30. Columns (2), (4), (6) and (8) restrict the sample to all successful negotiations only. *Sbj.Spec.Effects* refers to subject-specific random effects, while *Group.Spec.Effects* denotes matching-group specific random effects.

Table 9: Agreement payoff of player B as a function of player A's disagreement payoff.

	Agreement Payoff player B				
	All cases	d_B=30	d_B=40	d_B=50	d_B=60
	(1)	(2)	(3)	(4)	(5)
Constant (Baseline)	48.512***	44.669***	48.071***	51.420***	53.448***
	(0.604)	(0.589)	(0.856)	(2.279)	(1.118)
d_A	−4.102***	−4.054***	−3.133***	−4.751	−5.458***
	(0.351)	(0.492)	(0.542)	(2.947)	(0.866)
Threat treatment	3.723***	4.988***	2.515*	1.176	1.884
	(0.868)	(0.873)	(1.235)	(2.941)	(1.509)
Nuclear treatment	1.791*	1.962*	0.606	−0.553	1.870
	(0.876)	(0.870)	(1.239)	(3.100)	(1.554)
d_A x Threat treatment	2.674***	3.033***	2.604***	3.777	2.281*
	(0.496)	(0.728)	(0.779)	(3.682)	(1.162)
d_A x Nuclear treatment	1.277*	1.147	0.256	5.183	2.976*
	(0.504)	(0.727)	(0.778)	(3.978)	(1.216)
Sbj.Spec.Effects	✓	✓	✓	✓	✓
Group.Spec.Effects	✓	✓	✓	✓	✓
Conditional on agreement	✓	✓	✓	✓	✓
Observations	2,648	964	578	345	761
Log Likelihood	−10,174.390	−3,565.434	−2,179.213	−1,355.144	−2,979.254
Bayesian Inf. Crit.	20,419.710	7,192.709	4,415.662	2,762.880	6,018.220
Note:	$^{+}$p<0.10;*p<0.05;**p<0.01;***p<0.001;				

This table presents the results of regressions testing the effect of A's disagreement payoff on B's agreement payoff (payoff conditional on agreement). The first column list all the combinations of disagreement points. The columns 2-5 show the situations where player B has a disagreement payoff of 30, 40, 50 and 60, respectively. d_A denotes the scaled disagreement payoff of A, which ranges from 0 to 1. *Sbj.Spec.Effects* refers to subject-specific random effects, while *Group.Spec.Effects* denotes matching-group specific random effects.

Table 10: Regression analysis — agreement rates (Hyp 4).

	Agreement rate			
	All cases	Improving	Neutral	Deteriorating
	(1)	(2)	(3)	(4)
Constant (Baseline)	0.620***	0.894***	0.535***	0.246***
	(0.011)	(0.010)	(0.019)	(0.021)
Threat treatment	0.109***	−0.025	0.193***	0.246***
	(0.017)	(0.015)	(0.028)	(0.030)
Nuclear treatment	0.064***	−0.009	0.148***	0.101***
	(0.017)	(0.015)	(0.028)	(0.030)
Sbj.Spec.Effects	✓	✓	✓	✓
Group.Spec.Effects	✓	✓	✓	✓
Observations	7,852	3,580	2,148	2,124
Log Likelihood	−5,096.228	−993.096	−1,420.033	−1,339.984
Bayesian Inf. Crit.	10,246.270	2,035.291	2,886.100	2,725.935
Note:	$^{+}$p<0.10;*p<0.05;**p<0.01;***p<0.001;			

This table presents the results of regressions testing for Hypothesis 4 (agreement rates). The first column lists all the combinations of disagreement points. The second, third, and fourth columns show the combinations of disagreement points where the sum is below, equal to, or above 100. *Sbj.Spec.Effects* refers to subject-specific random effects, while *Group.Spec.Effects* denotes matching-group specific random effects.

Table 11: Correlating predicted trigger frequencies with agreement rates.

	Agreement rate							
	All cases		Improving		Neutral		Deteriorating	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Correlating predictions and agreement rates — Aplayers								
Constant (Nuclear)	0.676***	0.616***	0.901***	0.876***	0.662***	0.621***	0.327***	0.283***
	(0.019)	(0.022)	(0.018)	(0.022)	(0.031)	(0.037)	(0.035)	(0.040)
Threat treatment	0.042+	0.035	−0.011	−0.017	0.037	0.035	0.136**	0.130**
	(0.025)	(0.024)	(0.023)	(0.023)	(0.041)	(0.040)	(0.046)	(0.046)
Predicted Trigger Frequency	0.0005		−0.001*		0.001+		0.001+	
	(0.0004)		(0.0004)		(0.001)		(0.001)	
Non-zero Predicted Trigger Frequency		0.109***		0.014		0.102**		0.116**
		(0.022)		(0.023)		(0.039)		(0.042)
Sbj.Spec.Effects	✓	✓	✓	✓	✓	✓	✓	✓
Group.Spec.Effects	✓	✓	✓	✓	✓	✓	✓	✓
Observations	2,477	2,477	1,131	1,131	678	678	668	668
Log Likelihood	−1,548.298	−1,532.876	−346.127	−344.491	−429.373	−423.554	−456.922	−450.670
Bayesian Inf. Crit.	3,143.485	3,112.640	734.439	731.167	897.861	886.223	952.871	940.367
Panel B: Correlating predictions and agreement rates — A and B-players								
Constant (Nuclear)	0.678***	0.628***	0.893***	0.884***	0.675***	0.649***	0.329***	0.290***
	(0.013)	(0.016)	(0.013)	(0.016)	(0.022)	(0.026)	(0.024)	(0.029)
Threat treatment	0.042*	0.037*	−0.012	−0.016	0.041	0.040	0.136***	0.134***
	(0.018)	(0.018)	(0.017)	(0.017)	(0.029)	(0.029)	(0.033)	(0.032)
Predicted Trigger Frequency	0.0004		−0.001+		0.001		0.001*	
	(0.0003)		(0.0003)		(0.0005)		(0.001)	
Non-zero Predicted Trigger Frequency		0.088***		0.0001		0.055*		0.100***
		(0.016)		(0.016)		(0.027)		(0.030)
Sbj.Spec.Effects	✓	✓	✓	✓	✓	✓	✓	✓
Group.Spec.Effects	✓	✓	✓	✓	✓	✓	✓	✓
Observations	4,955	4,955	2,263	2,263	1,356	1,356	1,336	1,336
Log Likelihood	−3,073.510	−3,054.759	−676.041	−673.830	−847.436	−841.952	−909.312	−902.131
Bayesian Inf. Crit.	6,198.069	6,160.567	1,398.428	1,394.006	1,738.146	1,727.177	1,861.808	1,847.447

Note: $^{+}p<0.10$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$;

This table presents the results of regressions testing who predicted trigger frequencies affect agreement rates. The first two columns lists all the combinations of disagreement points. The columns 3-8 show the combinations of disagreement points where the sum is below, equal to, or above 100. *Predicted Trigger Frequency* denotes player A's predicted probability of player B's triggering the nuclear option. *Non-zero Predicted Trigger Frequency* denotes a dummy with value one if player A's predicted probability of player B's triggering is non zero. *Sbj.Spec.Effects* refers to subject-specific random effects, while *Group.Spec.Effects* denotes matching-group specific random effects.

Table 12: Regression analysis — joint payoffs (Hyp. 5).

	Overall payoff			
	All cases	Improving	Neutral	Deteriorating
	(1)	(2)	(3)	(4)
Constant (Baseline)	101.907***	98.149***	100.000***	110.101***
	(0.393)	(0.471)	(0.476)	(0.706)
Threat treatment	−3.965***	−2.632***	−2.936***	−7.174***
	(0.576)	(0.692)	(0.700)	(1.042)
Nuclear treatment	−3.138***	−2.136**	−2.416***	−5.520***
	(0.576)	(0.692)	(0.701)	(1.041)
Sbj.Spec.Effects	✓	✓	✓	✓
Group.Spec.Effects	✓	✓	✓	✓
Observations	7,852	3,580	2,148	2,124
Log Likelihood	−32,041.240	−14,078.330	−8,416.238	−9,055.706
Bayesian Inf. Crit.	64,136.290	28,205.760	16,878.510	18,157.380
Note:				+p<0.10;*p<0.05;**p<0.01;***p<0.001;

This table presents the results of regressions testing for Hypothesis 5 (joint payoffs). The first column lists all the combinations of disagreement points. The second, third, and fourth columns show the combinations of disagreement points where the sum is below, equal to, or above 100. *Sbj.Spec.Effects* refers to subject-specific random effects, while *Group.Spec.Effects* denotes matching-group specific random effects.

Table 13: Regression of the activation behavior.

	Dependent variable:							
	Propensity to activate				The time of activation			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.158***	0.158***	0.158***	0.089***	35.065***	34.976***	35.049***	35.386***
	(0.020)	(0.020)	(0.020)	(0.022)	(2.115)	(2.066)	(2.101)	(2.612)
d_B	0.057***		0.060***		−0.219		−0.234	
	(0.009)		(0.009)		(1.082)		(1.067)	
d_A	0.034***	0.039***			−0.081	−0.113		
	(0.009)	(0.009)			(1.026)	(1.012)		
$d_A + d_B = 100$				0.074***				0.019
				(0.021)				(2.372)
$d_A + d_B > 100$				0.181***				−0.866
				(0.021)				(2.307)
Sbj.Spec.Effects	✓	✓	✓	✓	✓	✓	✓	✓
Group.Spec.Effects	✓	✓	✓	✓	✓	✓	✓	✓
Conditional on activation	×	×	×	×	✓	✓	✓	✓
Observations	1,239	1,239	1,239	1,239	199	199	199	199
Log Likelihood	−376.954	−395.125	−381.129	−370.356	−813.479	−814.496	−814.425	−811.987
Bayesian Inf. Crit.	796.641	825.859	797.868	783.444	1,658.717	1,655.458	1,655.316	1,655.734
Note:								+p<0.10;*p<0.05;**p<0.01;***p<0.001;

This table presents the results of regressions studying the activation behavior. The first four columns focuses on the propensity to activate the nuclear option. The last four columns depict the time (in seconds) of the activation during the negotiation. d_A, and d_B refer to the normalized disagreement payoff of player A, and player B, respectively. Thus, a one-unit increase of d_A denotes an one-standard error increase in the disagreement payoff of player A. $d_A + d_B = 100$ and $d_A + d_B > 100$ denote a dummy with value one if the scenarios is one where the sum of the disagreement points equal to 100, and above 100, respectively. *Sbj.Spec.Effects* refers to subject-specific random effects, while *Group.Spec.Effects* denotes matching-group specific random effects.

ONLINE APPENDIX

A Details on the theoretical analysis

A.1 Bargaining in Baseline

A.1.1 Nash bargaining solution

The Nash product is given by

$$\left(\frac{100-x}{100^{\theta_A}}-\omega_A\right)\left(\frac{x}{100^{\theta_B}}-\omega_B\right)$$

where $\omega_i = d_i/(d_i+d_j)^{\theta_i}$ is player i's expected utility from the regular disagreement point with payoffs d_i and d_j. Maximizing the Nash product yields

$$x^*_{\text{Base}} = 50 + \frac{1}{2}\left(100^{\theta_B}\omega_B - 100^{\theta_A}\omega_A\right).$$

Player i's agreement payoff is therefore given by

$$x_i = 50 + \frac{1}{2}\left(100^{\theta_i}\omega_i - 100^{\theta_j}\omega_j\right).$$

The effect of the spite parameter θ_i on x_i depends on whether an agreement would be efficient, since

$$\begin{aligned}\frac{\partial x_i}{\partial \theta_i} &= \frac{1}{2}\left(\frac{\partial 100^{\theta_i}}{\partial \theta_i}\omega_i + 100^{\theta_i}\frac{\partial \omega_i}{\partial \theta_i}\right)\\ &= \frac{d_i}{2}\left(\frac{100}{d_i+d_j}\right)^{\theta_i}\ln\left(\frac{100}{d_i+d_j}\right) \lesseqgtr 0 \text{ iff } d_A + d_B \gtreqless 100.\end{aligned}$$

For the special case $d_i + d_j = 100$, the spite parameter is irrelevant, and we have $x_i = d_i$. The partial derivatives of x_i with respect to d_i and d_j are

$$\frac{\partial x_i}{\partial d_i} = \frac{1}{2}\left(100^{\theta_i}\frac{\partial \omega_i}{\partial d_i} - 100^{\theta_j}\frac{\partial \omega_j}{\partial d_i}\right) = \frac{100^{\theta_i}\left((1-\theta_i)d_i + d_j\right)}{2\left(d_i+d_j\right)^{\theta_i+1}} + \frac{100^{\theta_j}d_j\theta_j}{2\left(d_i+d_j\right)^{\theta_j+1}} > 0$$

and

$$\frac{\partial x_i}{\partial d_j} = \frac{1}{2}\left(100^{\theta_i}\frac{\partial \omega_i}{\partial d_j} - 100^{\theta_j}\frac{\partial \omega_j}{\partial d_j}\right) = -\frac{100^{\theta_i}d_i\theta_i}{2\left(d_i+d_j\right)^{\theta_i+1}} - \frac{100^{\theta_j}\left(d_i + \left(1-\theta_j\right)d_j\right)}{2\left(d_i+d_j\right)^{\theta_j+1}} < 0.$$

Thus, any improvement in d_i—the outside option if no agreement can be found—strengthens player i's bargaining position and implies a higher agreement payoff x_i for her. However, since the disagreement option itself becomes more attractive as well, it remains to be seen whether an increase in d_i makes player i more inclined or less inclined to accept.

Similarly, while a higher value of d_j is bad news for player i's agreement payoff, the utility from the alternative, the disagreement point, is also negatively affected—at least if player i has relative-payoff concerns ($\theta_i > 0$). We address these issues in the following subsection.

A.1.2 Agreements versus disagreements and the propensity to accept

Player i accepts the x^*-solution with probability

$$p_i = \frac{u_i\left(x_i, 100 - x_i\right)^{\lambda_i}}{u_i\left(x_i, 100 - x_i\right)^{\lambda_i} + \omega_i^{\lambda_i}}.$$

Letting $k_i = u_i(x_i, 100 - x_i)/\omega_i$ we can rewrite this as

$$p_i = \frac{k_i^{\lambda_i}}{k_i^{\lambda_i} + 1.}$$

First, we consider the impact of a change in **player i's disagreement payoff d_i** on her willingness to accept the x^*-solution. The partial derivative of p_i with respect to d_i is

$$\frac{\partial p_i}{\partial d_i} = \frac{\partial k_i^{\lambda_i}/\partial d_i}{\left(k_i^{\lambda_i} + 1\right)^2} = \frac{\lambda_i k_i^{\lambda_i - 1}}{\left(k_i^{\lambda_i} + 1\right)^2} \frac{\partial k_i}{\partial d_i}$$

and therefore $\text{sgn}(\partial p_i/\partial d_i) = \text{sgn}(\partial k_i/\partial d_i)$. We get

$$\frac{\partial k_i}{\partial d_i} = \frac{100^{\theta_j - \theta_i} d_j}{2d_i^2\left(d_i + d_j\right)^{1-\theta_i}} \left(\frac{\left(1 - \theta_i + \theta_j\right) d_i + d_j}{\left(d_i + d_j\right)^{\theta_j}} - \frac{100\left(2 - \theta_i\right)}{100^{\theta_j}}\right)$$

and since the expression in the parentheses is zero for $\theta_i = \theta_j = 1$ and negative for all other values of θ_i and θ_j, we find that $\partial p_i/\partial d_i \leq 0$. An improvement in her disagreement payoff d_i makes player i less inclined to accept the x^*-solution.

Next, we examine how the **partner's disagreement payoff d_j** affects player i's accept probability. Using the same approach as before, we find that the sign of $\partial p_i/\partial d_j$ depends on the sign of

$$\frac{\partial k_i}{\partial d_j} = \frac{100\theta_i - \left(\dfrac{100}{d_i + d_j}\right)^{\theta_j}\left(d_i + \left(1 + \theta_i - \theta_j\right) d_j\right)}{2d_i 100^{\theta_i}\left(d_i + d_j\right)^{1-\theta_i}}.$$

Although $\text{sgn}(\partial k_i/\partial d_j)$ is generally ambiguous, for almost all parameter combinations relevant in our experimental setting $\partial k_i/\partial d_j < 0$. To see this, note that $\partial k_i/\partial d_j$ reaches a maximum at $\theta_i = 1$, $\theta_j = 0$, $d_j = 30$ with

$$\left.\frac{\partial k_i}{\partial d_j}\right|_{\theta_i=1,\ \theta_j=0,\ d_j=30} = \frac{40 - d_i}{200 d_i}.$$

Thus, among our experimental parameters, there is only one case—$\theta_i = 1$, $\theta_j = 0$, $d_i = d_j = 30$—in which a 10-point increase in d_j does not lead to an unambiguous decline in player i's propensity to accept the x^*-solution. As it turns out,

$$k_i\left(\theta_i = 1,\ \theta_j = 0,\ d_i = d_j = 30\right) = k_i\left(\theta_i = 1,\ \theta_j = 0,\ d_i = 30,\ d_j = 40\right) = 1.2.$$

Hence, when the value of d_j is increased from 30 to 40, k_i and therefore player i's accept probability p_i remain constant—but only if $\theta_i = 1$ and $\theta_j = 0$. In all other cases, an increase in d_j leads to a decline in p_i.

A.2 Bargaining under a nuclear threat

A.2.1 Probability of a nuclear outcome

In the Nuke and Threat treatments, the probability of player B triggering the nuclear option conditional on a disagreement is

$$\sigma_B^N = \frac{\eta_B^{\lambda_B}}{\eta_B^{\lambda_B} + \omega_B^{\lambda_B}}$$

with $\eta_B = 9^{1-\theta_B}$. σ_B^N is increasing in d_A since

$$\frac{\partial \sigma_B^N}{\partial d_A} = -\frac{\eta_B^{\lambda_B}}{\left(\eta_B^{\lambda_B} + \omega_B^{\lambda_B}\right)^2}\frac{\partial \omega_B^{\lambda_B}}{\partial d_A}$$

and $\partial\omega_B^{\lambda_B}/\partial d_A \leq 0$ for any $\theta_B \in [0,\ 1]$. At the same time, σ_B^N is decreasing in d_B since

$$\frac{\partial \sigma_B^N}{\partial d_B} = -\frac{\eta_B^{\lambda_B}}{\left(\eta_B^{\lambda_B} + \omega_B^{\lambda_B}\right)^2}\frac{\partial \omega_B^{\lambda_B}}{\partial d_B}$$

and $\partial\omega_B^{\lambda_B}/\partial d_B > 0$ for $\theta_B \in [0,\ 1]$. Finally, it is noteworthy that

$$\frac{\partial \sigma_B^N}{\partial \lambda_B} = \frac{\eta_B^{\lambda_B}\omega_B^{\lambda_B}}{\left(\eta_B^{\lambda_B} + \omega_B^{\lambda_B}\right)^2}\ln\left(\frac{\eta_B}{\omega_B}\right) \lesseqgtr 0 \text{ iff } \eta_B \lesseqgtr \omega_B.$$

A.2.2 Expected utility from a disagreement and the Nash bargaining solution

If no agreement is reached during the negotiation stage, either the regular disagreement point is implemented, or players face the nuclear outcome. Thus, player i's expected utility is $\sigma_B^N\eta_i + \left(1-\sigma_B^N\right)\omega_i$. The Nash product,

$$\left(\frac{100-x}{100^{\theta_A}} - \left(1-\sigma_B^N\right)\omega_A\right)\left(\frac{x}{100^{\theta_B}} - \left(\sigma_B^N\eta_B + \left(1-\sigma_B^N\right)\omega_B\right)\right),$$

is maximized at

$$\begin{aligned} x^*_{\text{Nuke}} &= \left(1-\sigma_B^N\right)\left(50 + \frac{1}{2}\left(100^{\theta_B}\omega_B - 100^{\theta_A}\omega_A\right)\right) + \sigma_B^N\left(50 + 100^{\theta_B}\frac{\eta_B}{2}\right) \\ &= \left(1-\sigma_B^N\right)x^*_{\text{Base}} + \sigma_B^N\left(50 + 100^{\theta_B}\frac{\eta_B}{2}\right). \end{aligned} \tag{6}$$

A.2.3 Comparing agreement payoffs between treatments

As equation (6) shows, player B's agreement payoff can be written in a format that puts a weight of $1-\sigma_B^N$ (the probability that player B *does not* trigger the nuclear option) on 'no change' relative to the Baseline treatment and a weight of σ_B^N (the probability that the nuclear outcome *does* occur) on the payoff in the parentheses. Accordingly, the nuclear option is predicted to have an impact if $x^*_{\text{Base}} \neq 50 + 100^{\theta_B}\eta_B/2$. This implies that player B's agreement payoff is (weakly) greater in the Nuke treatment ($x^*_{\text{Nuke}} \geq x^*_{\text{Base}}$) iff

$$100^{\theta_A}\omega_A \geq 100^{\theta_B}\left(\omega_B - \eta_B\right). \tag{7}$$

Because $\partial\omega_i/\partial d_i > 0$, player B's (dis-)advantage arising from the nuclear threat intuitively depends on how much each player has to lose in case of a nuclear outcome. For example, with purely selfish players ($\theta_A = \theta_B = 0$), condition (7) boils down to $d_A \geq d_B - 9$. This holds in all scenarios except 30-40, 30-60, 40-60, and 50-60. Thus, whenever there is more at stake for player B than for player A, having been given access to the nuclear option is *bad* news for B, provided that $\sigma_B^N > 0$. In such cases, B would not trigger the nuclear option out of spite ($\theta_B = 0$ in the example), but there is the possibility of a heat-of-the-moment decision to go nuclear (low λ_B-value). This is detrimental for player B's agreement payoff, effectively because B is worried about "losing it" if no agreement can be reached, which makes him more inclined to seek a compromise in the negotiation.

To quantify the range of possible treatment effects within our model, we consider

$$\Delta x^* = x^*_{\text{Nuke}} - x^*_{\text{Base}} = \frac{\sigma_B^N}{2}\left(100^{\theta_A}\omega_A - 100^{\theta_B}\left(\omega_B - \eta_B\right)\right).$$

For each scenario d_A-d_B the range of Δx^* depends on the three parameters θ_A, θ_B, and λ_B. Our analysis is based on

Proposition A.1.

$$\frac{\partial \Delta x^*}{\partial \theta_B} > 0 \tag{8}$$

which we will prove below.

Under condition 8, $\theta_B = 0$ [$\theta_B = 1$] will minimize [maximize] Δx^*. Furthermore, note that

$$\frac{\partial \Delta x^*}{\partial \theta_A} = \frac{\sigma_B^N}{2}\left(\frac{100}{d_A + d_B}\right)^{\theta_A}\ln\left(\frac{100}{d_A + d_B}\right) d_A \lesseqgtr 0 \quad \text{iff} \quad d_A + d_B \gtreqless 100.$$

Thus, to evaluate the minimum [maximum] of Δx^* we set $\theta_A = 0$ if $d_A + d_B < 100$ and $\theta_A = 1$ if $d_A + d_B > 100$ [$\theta_A = 1$ if $d_A + d_B < 100$ and $\theta_A = 0$ if $d_A + d_B > 100$]. For $d_A + d_B = 100$ the value of θ_A has no effect on Δx^*. This yields the following table:

Note that $\max\left(\Delta x^*\right) > 0$ in all cases, whereas $\min\left(\Delta x^*\right)$ can turn negative if d_B is sufficiently large for a given d_A. Note further that $\sigma_B^N = 9^{\lambda_B}/(9^{\lambda_B} + d_B^{\lambda_B})$ is strictly decreasing in λ_B. Thus, to determine the full range of values, we assume $\lambda_B = 0$ in $\min\left(\Delta x^*\right)$ when $\min\left(\Delta x^*\right) < 0$ and assume $\lambda_B \to \infty$ otherwise. The results of this exercise are displayed in Figure 13.

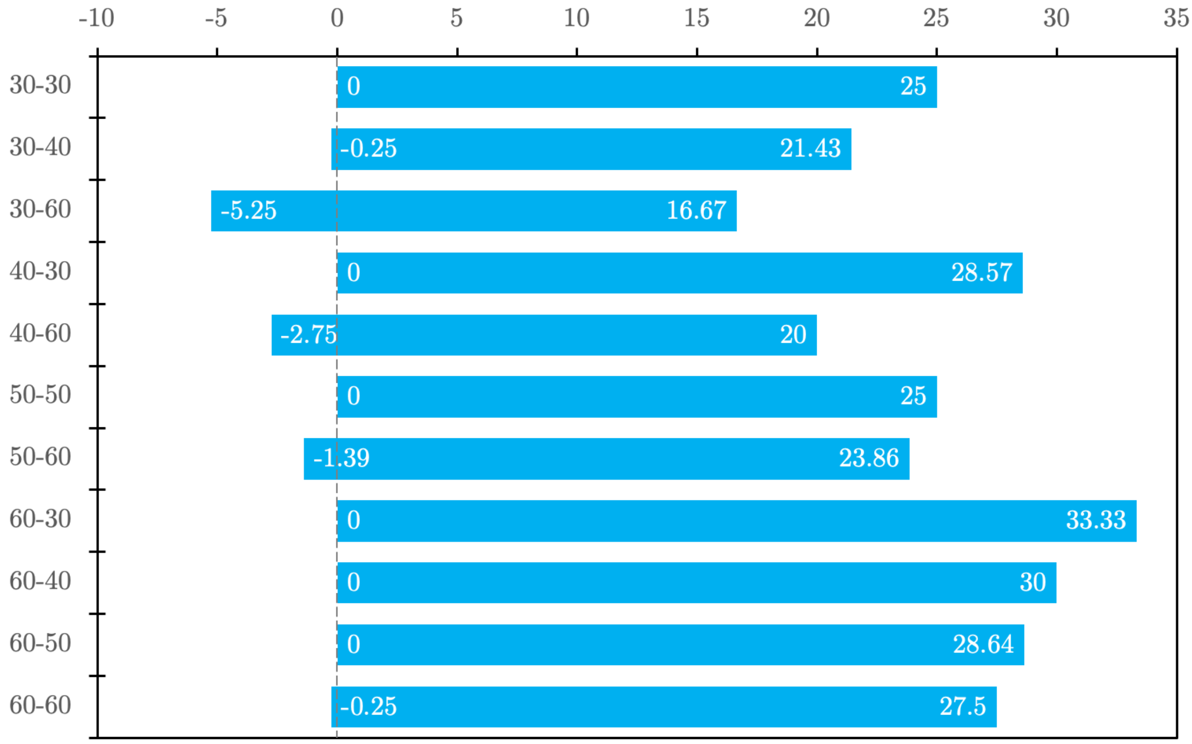
Finally, for a proof of proposition A.1, note first that

$$\frac{\partial \Delta x^*}{\partial \theta_B} = \frac{1}{2}\frac{\partial \sigma_B^N}{\partial \theta_B}\left(100^{\theta_A}\omega_A - 100^{\theta_B}\left(\omega_B - \eta_B\right)\right) - \frac{\sigma_B^N}{2}\frac{\partial}{\partial \theta_B}\left(100^{\theta_B}\left(\omega_B - \eta_B\right)\right)$$

Table 14: Minimum and maximum values of Δx^*

	Efficiency-improving or -neutral agreements	Efficiency-deteriorating agreements
$\min(\Delta x^*)$	$\frac{1}{2}\left(9 + d_A - d_B\right)\sigma_B^N$ $(\sigma_A = 0, \sigma_B = 0)$	$\frac{1}{2}\left(9 + \frac{100 d_A}{d_A + d_B} - d_B\right)\sigma_B^N$ $(\sigma_A = 1, \sigma_B = 0)$
$\max(\Delta x^*)$	$\frac{100 d_A}{d_A + d_B}\sigma_B^N$ $(\sigma_A = 1, \sigma_B = 1)$	$\frac{d_A}{2}\left(1 + \frac{100}{d_A + d_B}\right)\sigma_B^N$ $(\sigma_A = 0, \sigma_B = 1)$

Figure 13: Prediction range for the effect of the nuclear option on B's agreement payoff.



The figure displays how, according to the model, player B's agreement payoffs in the eleven scenarios may change in the Nuke and Threat treatments relative to Baseline (Δx^*, see Table 14). While the range of predictions always includes zero and sometimes even negative effects, the model mostly predicts player B to benefit from the nuclear option.

and that

$$\frac{\partial \sigma_B^N}{\partial \theta_B} = \lambda_B \sigma_B^N \left(1 - \sigma_B^N\right) \ln\left(\frac{d_A + d_B}{9}\right) \geq 0.$$

for all values $d_i \in \{30, 40, 50, 60\}$. Thus, to show that proposition 6 holds, it is sufficient to demonstrate that $\partial(100^{\theta_B}(\omega_B - \eta_B))/\partial\theta_B < 0$. We obtain

$$\frac{\partial}{\partial \theta_B}\left(100^{\theta_B}\left(\omega_B - \eta_B\right)\right) = 100^{\theta_B}\left(\omega_B \ln\left(\frac{100}{d_A + d_B}\right) - \eta_B \ln\left(\frac{100}{9}\right)\right).$$

Evaluated at $\theta_B = 0$ this expression simplifies to

$$d_B \ln\left(\frac{100}{d_A + d_B}\right) - 9 \ln\left(\frac{100}{9}\right)$$

which is negative for all values $d_i \in \{30, 40, 50, 60\}$. Moreover, it turns out that

$$\frac{\partial^2}{\partial \theta_B^2} \left(100^{\theta_B} (\omega_B - \eta_B)\right) < 0 \qquad \Leftrightarrow \qquad \frac{\omega_B}{\eta_B} < \left(\frac{\ln 100 - \ln 9}{\ln 100 - \ln (d_A + d_B)}\right)^2$$

which always holds within the relevant parameter range. Hence,

$$\frac{\partial (100^{\theta_B} (\omega_B - \eta_B))}{\partial \theta_B} < 0$$

throughout, completing the proof of the proposition.

A.2.4 Comparing agreement probabilities between treatments

Consider first player A who accepts the x^*-solution in Baseline with probability

$$p_A = \frac{k_A^{\lambda_A}}{k_A^{\lambda_A} + 1} \qquad \text{with } k_A = \frac{100 - x^*_{\text{Base}}}{100^{\theta_A} \omega_A}$$

and accepts the x^*-solution in Nuke with probability

$$q_A = \frac{h_A^{\lambda_A}}{h_A^{\lambda_A} + 1} \qquad \text{with } h_A = \frac{100 - x^*_{\text{Nuke}}}{(1 - \sigma_B^N)\, 100^{\theta_A} \omega_A}.$$

At $\lambda_A = 0$ player A accepts with probability 0.5 in both treatments. For $\lambda_A > 0$,

$$q_A \geq p_A \quad \Leftrightarrow \quad h_A \geq k_A \quad \Leftrightarrow \quad \frac{100 - x^*_{\text{Nuke}}}{100 - x^*_{\text{Base}}} \geq 1 - \sigma_B^N \quad \Leftrightarrow \quad 100 \geq 100^{\theta_B} \eta_B$$

which is true for any $\theta_B \in [0,\ 1]$ (the equality sign holds for $\theta_B = 1$).

Now consider player B for whom

$$k_B = \frac{x^*_{\text{Base}}}{100^{\theta_B} \omega_B} \qquad \text{and} \qquad h_B = \frac{x^*_{\text{Nuke}}}{100^{\theta_B} \left(\sigma_B^N \eta_B + (1 - \sigma_B^N)\, \omega_B\right)}$$

and $h_B \geq k_B \quad \Leftrightarrow$

$$100^{\theta_A} \omega_A \geq 100 \left(1 - \frac{\omega_B}{\eta_B}\right). \tag{9}$$

Player B weakly preferring the regular disagreement point to the nuclear outcome ($\omega_B \geq \eta_B$) is a sufficient condition for the inequality to hold, and in this case player B—like player A—is more likely to agree in the Nuke treatment than in the Baseline. Even if we assume that player B is maximally spiteful ($\theta_B = 1$) such that $\omega_B \ll \eta_B$, the condition still holds in all cases in which $d_A + d_B \geq 100$. For $d_A + d_B < 100$, player B is *less* likely to accept the x^*-solution in Nuke than in Baseline iff θ_B exceeds a critical value:

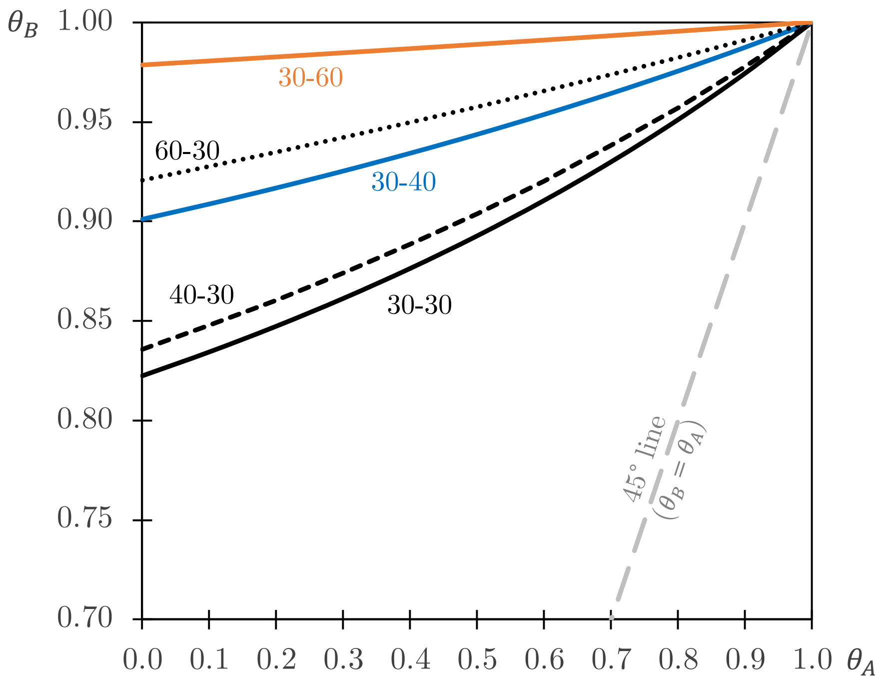
$$\theta_B > \hat{\theta}_B + \frac{\ln(100) - \ln\left(100 - 100^{\theta_A} \omega_A\right)}{\ln (d_A + d_B) - \ln(9)}$$

where

$$\hat{\theta}_B = \frac{\ln(d_B) - \ln(9)}{\ln(d_A + d_B) - \ln(9)}$$

is the value of θ_B at which player B is indifferent between the regular disagreement point and the nuclear outcome. Figure 14 displays a plot of the critical value for the relevant parameter specifications. As the figure shows, the model predicts that the introduction of the nuclear option lowers the accept probability for player B only if player B's spite parameter is extremely high—both in absolute terms and relative to player A's spite parameter.

Figure 14: Critical θ-values for negative treatment effect in B's accept probability.



The figure shows for which combinations of θ_A and θ_B player B is, according to the model, *less* inclined to agree to the Nash bargaining solution in the nuclear treatments than in the Baseline treatment. As indicated in the figure, this possibility arises only for very high values of θ_B in combination with relatively low values of θ_A. In the scenarios not shown in the figure (40-60, 60-40, 50-50, 50-60, 60-50), *no* combination of θ_A and θ_B leads to a lower inclination for B to to accept.

B Further analysis & Results

B.1 Demographics by treatment

Table 15 summarizes all demographic variables by the treatment and the role assigned. We can see that the treatments are balanced and subjects do not differ on observables.

		aA (N=144)	aB (N=142)	bA (N=151)	bB (N=164)	nA (N=168)	nB (N=165)	Total (N=934)	p value
Gender		0.49 (0.52)	0.55 (0.51)	0.50 (0.53)	0.59 (0.52)	0.49 (0.50)	0.56 (0.51)	0.53 (0.51)	0.391
Age		41.44 (12.13)	41.13 (11.23)	40.11 (11.30)	39.55 (12.00)	38.94 (11.45)	38.33 (10.39)	39.85 (11.44)	0.121
Ethnic									0.525
	African American	5 (3.5%)	11 (7.7%)	16 (10.6%)	15 (9.1%)	14 (8.3%)	16 (9.7%)	77 (8.2%)	
	Asian	15 (10.4%)	9 (6.3%)	13 (8.6%)	9 (5.5%)	13 (7.7%)	11 (6.7%)	70 (7.5%)	
	Hispanic	5 (3.5%)	9 (6.3%)	3 (2.0%)	12 (7.3%)	8 (4.8%)	12 (7.3%)	49 (5.2%)	
	Native	1 (0.7%)	3 (2.1%)	1 (0.7%)	2 (1.2%)	1 (0.6%)	1 (0.6%)	9 (1.0%)	
	White	118 (81.9%)	110 (77.5%)	118 (78.1%)	126 (76.8%)	132 (78.6%)	125 (75.8%)	729 (78.1%)	
Degree									0.772
	NoSchool	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.6%)	1 (0.6%)	0 (0.0%)	2 (0.2%)	
	GED	1 (0.7%)	4 (2.8%)	2 (1.3%)	1 (0.6%)	3 (1.8%)	3 (1.8%)	14 (1.5%)	
	HighSchool	31 (21.5%)	34 (23.9%)	40 (26.5%)	47 (28.7%)	42 (25.0%)	42 (25.5%)	236 (25.3%)	
	College	4 (2.8%)	4 (2.8%)	4 (2.6%)	6 (3.7%)	10 (6.0%)	16 (9.7%)	44 (4.7%)	
	BA	81 (56.2%)	74 (52.1%)	80 (53.0%)	82 (50.0%)	81 (48.2%)	82 (49.7%)	480 (51.4%)	
	MA	20 (13.9%)	20 (14.1%)	18 (11.9%)	19 (11.6%)	25 (14.9%)	14 (8.5%)	116 (12.4%)	
	Professional	4 (2.8%)	4 (2.8%)	3 (2.0%)	4 (2.4%)	4 (2.4%)	5 (3.0%)	24 (2.6%)	
	PhD	3 (2.1%)	2 (1.4%)	4 (2.6%)	4 (2.4%)	2 (1.2%)	3 (1.8%)	18 (1.9%)	
Job									0.452
	Employed full-time	94 (65.3%)	103 (72.5%)	88 (58.3%)	89 (54.3%)	97 (57.7%)	104 (63.0%)	575 (61.6%)	
	Employed part-time	13 (9.0%)	11 (7.7%)	13 (8.6%)	23 (14.0%)	15 (8.9%)	15 (9.1%)	90 (9.6%)	
	Out of work	10 (6.9%)	6 (4.2%)	20 (13.2%)	19 (11.6%)	13 (7.7%)	14 (8.5%)	82 (8.8%)	
	Self-employed	11 (7.6%)	10 (7.0%)	13 (8.6%)	13 (7.9%)	21 (12.5%)	15 (9.1%)	83 (8.9%)	
	student	2 (1.4%)	2 (1.4%)	2 (1.3%)	4 (2.4%)	5 (3.0%)	4 (2.4%)	19 (2.0%)	
	Unable	14 (9.7%)	10 (7.0%)	15 (9.9%)	16 (9.8%)	17 (10.1%)	13 (7.9%)	85 (9.1%)	
Income									0.464
	<10k	5 (3.5%)	2 (1.4%)	4 (2.6%)	7 (4.3%)	3 (1.8%)	7 (4.2%)	28 (3.0%)	
	<15k	9 (6.2%)	2 (1.4%)	3 (2.0%)	6 (3.7%)	1 (0.6%)	6 (3.6%)	27 (2.9%)	
	<20k	5 (3.5%)	2 (1.4%)	9 (6.0%)	6 (3.7%)	5 (3.0%)	4 (2.4%)	31 (3.3%)	
	<30k	16 (11.1%)	16 (11.3%)	19 (12.6%)	17 (10.4%)	8 (4.8%)	20 (12.1%)	96 (10.3%)	
	<40k	11 (7.6%)	17 (12.0%)	11 (7.3%)	13 (7.9%)	19 (11.3%)	16 (9.7%)	87 (9.3%)	
	<50k	12 (8.3%)	20 (14.1%)	16 (10.6%)	19 (11.6%)	26 (15.5%)	18 (10.9%)	111 (11.9%)	
	<75k	34 (23.6%)	34 (23.9%)	33 (21.9%)	36 (22.0%)	35 (20.8%)	34 (20.6%)	206 (22.1%)	
	<100k	23 (16.0%)	22 (15.5%)	27 (17.9%)	33 (20.1%)	34 (20.2%)	23 (13.9%)	162 (17.3%)	
	<125k	7 (4.9%)	12 (8.5%)	16 (10.6%)	12 (7.3%)	19 (11.3%)	18 (10.9%)	84 (9.0%)	
	<150k	7 (4.9%)	7 (4.9%)	4 (2.6%)	9 (5.5%)	9 (5.4%)	9 (5.5%)	45 (4.8%)	
	<200k	9 (6.2%)	6 (4.2%)	5 (3.3%)	5 (3.0%)	4 (2.4%)	8 (4.8%)	37 (4.0%)	
	>200k	6 (4.2%)	2 (1.4%)	4 (2.6%)	1 (0.6%)	5 (3.0%)	2 (1.2%)	20 (2.1%)	
OnlineWork		18.62 (14.59)	15.69 (11.02)	15.37 (12.23)	17.01 (12.35)	17.40 (13.14)	16.23 (11.73)	16.73 (12.57)	0.234

Table 15: Summary statistics by treatment and the role assigned.

The table shows summary statistic by participants. *Gender* denotes a dummy with value one if the participant is female. *Age* indicates the participants' age. *Ethnicity* denotes the participants' ethnicity. *Degree* indicates the participants' highest achieved degree. *Job* indicates the participants' current job. *Income* indicates the participants' household income in 2019. *HoursWorkOnline* indicates the number of hours a participant spends on online work per week.

B.2 Active participants and dropouts

Table 16 shows the number of active players in each round by treatment and role assigned. Table 17 shows the number of dropouts in each round by treatment and role assigned.

Round	aA	aB	bA	bB	nA	nB
1	120	120	132	132	119	119
2	117	117	132	132	116	116
3	116	116	132	132	115	115
4	113	113	132	132	113	113
5	111	111	132	132	112	112
6	111	111	132	132	112	112
7	111	111	132	132	111	111
8	110	110	132	132	111	111
9	110	110	132	132	110	110
10	110	110	130	130	110	110
11	110	110	130	130	110	110

Table 16: Active participants by treatment.

Round	aA	aB	bA	bB	nA	nB
1	0	0	0	0	0	0
2	2	2	0	0	2	1
3	0	1	0	0	0	1
4	1	2	0	0	0	2
5	2	1	0	0	0	1
6	0	0	0	0	0	0
7	0	0	0	0	0	1
8	0	1	0	0	0	0
9	0	0	0	0	0	1
10	0	0	1	1	0	0
11	0	0	0	0	0	0

Table 17: Dropouts by treatment.

B.3 Post-experimental questions

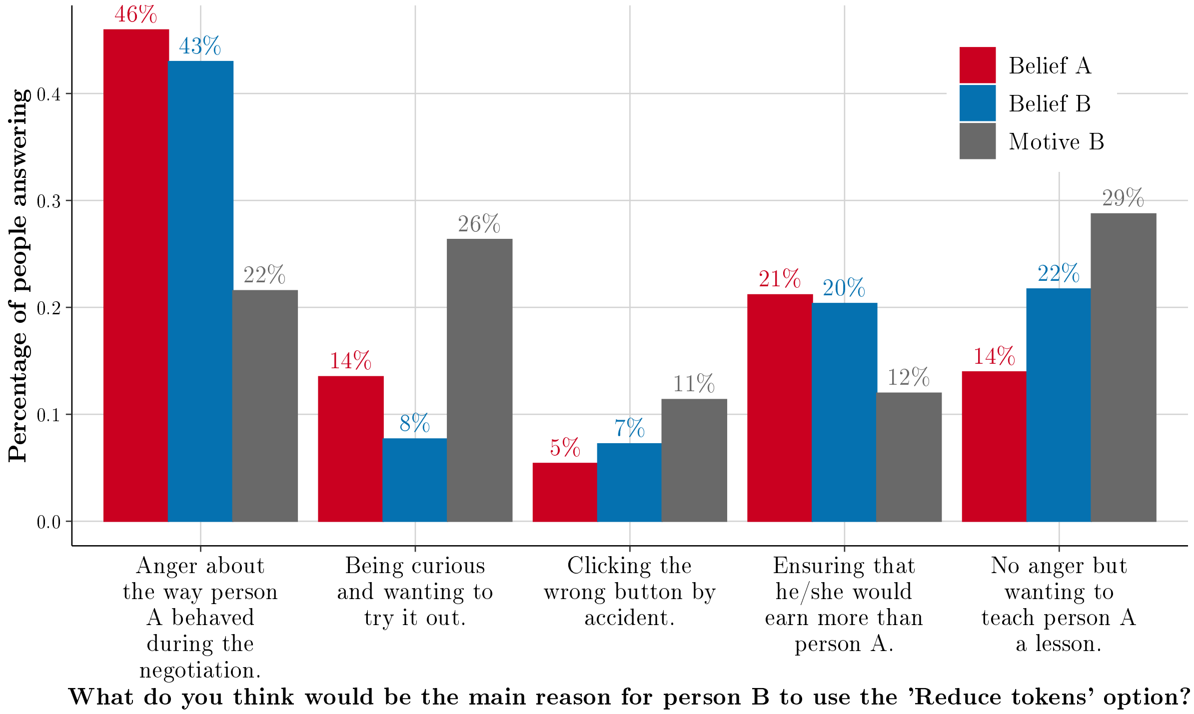


Figure 15: Post experimental questions.

Note: The figure depicts the motives for players B triggering of the nuclear option. Red bars denote the belief of players A on what they thought players B main motive for triggering the nuclear option is. Blue bars denote the belief of players B on what they thought *other* players B main motive for triggering the nuclear option is. Grey bars denote the response of players B on what they thought *their own* main motive for triggering the nuclear option is. The response "other" has been omitted from the figure.

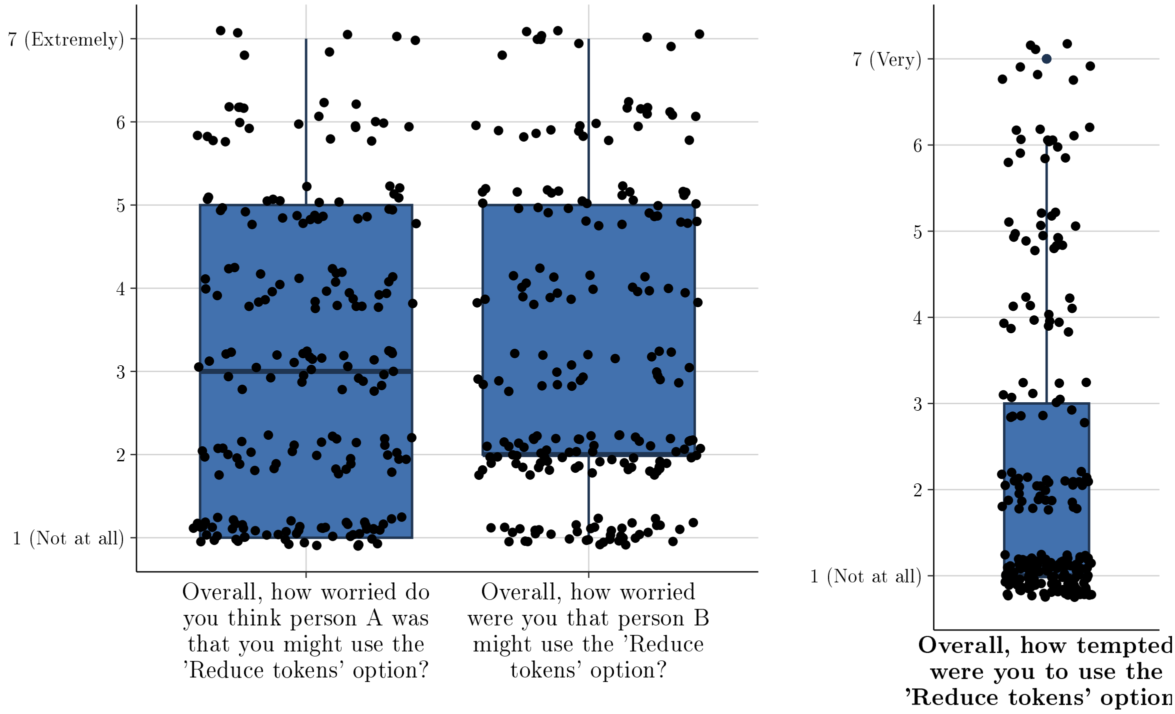


Figure 16: Post experimental questions.

Note: Blue boxes denote boxplots of the responses. Horizontal thick line denotes the median responses. Black dots denote individual responses (the responses are slightly jittered to improve visibility). The left panel depicts the responses of players A, and B to the question "Overall, how worried were you that person B might use the 'Reduce tokens' option?" on the left and "Overall, how worried do you think person A was that you might use the 'Reduce tokens' option?" on the right, respectively. The right panel depicts the response of players B in the nuclear treatment to the question "Overall, how tempted were you to use the 'Reduce tokens' option?".

In the post-experimental questions, we asked participants assigned the role of B (the nuclear-option bearer) whether they felt tempted to use the nuclear option, to get an idea of how many participants would consider using the nuclear option. In the right panel of Figure 16 we see that most participants (56 %) in the role of B were not at all tempted to use the nuclear option. We also obtain a very high correlation between participants being tempted and participants using the nuclear option ($r = 0.463$, $p < 0.001$).

More interestingly, we asked participants in the role of B in the nuclear treatments whether they thought that A-players would be worried that B-players might use the nuclear option, and we also asked A-players how worried they have been that B-players might use the nuclear option. In the left upper panel of Figure 16 we report the responses. We see that there is a lot of variability in the responses and we don't see any clear focal answer. While it seems like players A were slightly more worried than players B anticipated, the difference was not significant $p = 0.68$). The median response of players A to the question how worried they have been is "very little worried". The median response of players B to the question how worried the think A-players have been is "little worried". Thus, we can conclude that while there was no big concern on the side of players A, there were still a considerable worry that players B might use the nuclear option.[23]

[23] It is noteworthy, that participants worry reflect their belief after they have reacted. Thus, if players A are ex-ante very worried of B using the nuclear option, they would ensure this not to happen and therefore would be very little worried after they ensure that all negotiations were successful.

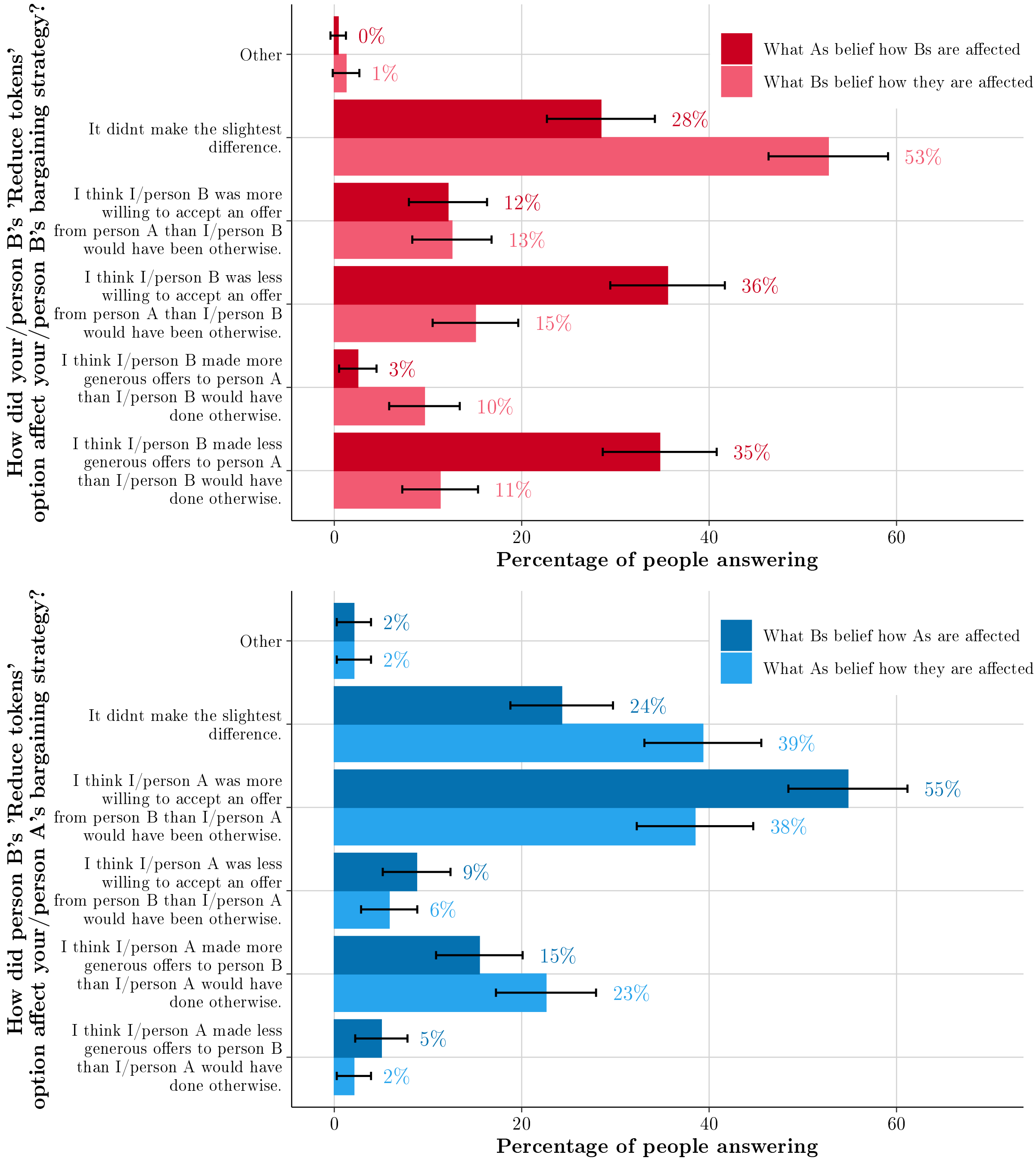


Figure 17: Post experimental questions.

Note: Bars denote the percentage of participants responding in the given way to the posed question. Error bars denote 95% confidence intervals. The top panel denotes the response to the question "How did your/person B's 'Reduce tokens' option affect your/person B's bargaining strategy?". Dark-red bars denote the belief of players B on how they are affected in their behavior by the nuclear option. Light-red bars denote the belief of players A on how players B are affected in their behavior by the nuclear option. The bottom panel denotes the response to the question "How did person B's 'Reduce tokens' option affect your/person A's bargaining strategy?". Dark-blue bars denote the belief of players A on how they are affected in their behavior by the nuclear option. Light-blue bars denote the belief of players B on how players A are affected in their behavior by the nuclear option.

To get a better grasp on the perceived effect of the nuclear option, we asked participants how they thought the nuclear option affected their own strategy, and how they thought the nuclear option affected their partners. Figure 17 reports the responses. The top panel compares Bs' responses on their own strategy to As' beliefs over Bs' strategy. We clearly see that the responses of As and Bs differ ($\chi^2(5)= 31.6$, $p < 0.001$). Most players B responded that the nuclear option did not make any difference in their bargaining behavior (53%) while only 28% of players A believed this to be the case. Players A believed that the players B were less willing to accept requests from As (36%), and that players B made less generous requests (35%), which is in stark contrast to the responses given by players B, where only 15% thought that they were less willing to accept requests from As and only 11% believed that they made less generous requests. Thus, while most players B believed they didn't change their behavior, most As felt that players B were less generous and expected higher requests from As.

The bottom panel compares As' responses on their own strategy to Bs' beliefs over As' strategy. We see that the responses of As and Bs differ, but not as stark as in the previous graph ($\chi^2(5)= 9.5$, $p = 0.089$). Most players A responded that the nuclear option changed their bargaining behavior (61%) (76% of players B believed this to be the case). More interestingly, 55% of B-players believed that players A were more willing to accept a request from B than they would have otherwise, and similarly 38% of players A thought this to be the case. We can also see that 23% of players A thought that they made more generous requests, and similarly 15% of players B believed this to be the case for players A. Thus, while some players believed that the nuclear option did not affect them, a majority of players A (as well as players B) believed that players A were more generous with their requests and were more willing to accept a request from players B.

This provides a tentative mechanism for the changed results. Both players A and B seem to agree that players A became more generous, and were more willing to accept requests. It also seems like As and Bs disagree on the effect on players B: while players B believed not to have changed their behavior, most players A believed that players B turn more greedy. Thus, the change in bargaining behavior is very likely driven by the change in behavior by players As. The behavior of players B is probably less driving the changes as they themselves believed not to be affected, considered the nuclear option as not all all tempting, and believed that players A would not be worried that the option will be triggered. Thus, players B were not fully aware of their increased bargaining power, while players A (who are the ones mostly negatively affected by the nuclear option) are the ones being somewhat afraid of the nuclear option, and therefore reacting to it.

C Instructions

Note: The instructions indicate in grey to which treatments the particular parts apply.

Both treatments:

Thank you for your participation! This is a research study on how people negotiate over splitting an amount of money in a very simple setting.

Important: This study consists of two parts!

For participating (and completing!) this current **part 1** you will receive a reward of $1. For participating and completing **part 2** you will receive an additional $1, plus a bonus of up to $10. You can join part 2 only if you have completed part 1.

Part 2 will take place tomorrow (Tuesday) at 10am PDT (Pacific Daylight Time).

Why is part 2 set at a fixed time? In part 2 you will be matched *live* with other people on MTurk and interact with them online. For this to work we need all participants to be logged in at the same time.

Please proceed with this part 1 only if you can also make it to part 2. If you are not sure that you are available between 10am and approximately 10.30am PDT tomorrow (Tue) please cancel this HIT now and return it. Participating in part 1 is pointless if you do not also participate in part 2.

What is this about?

This current part 1 is mainly about preparing you for part 2 which is the main part of our survey. In part 2 you will negotiate with other people on MTurk over splitting an amount of money (with one person at a time) for a total of 11 rounds. These negotiations will be non-verbal (further details below). Each round will take only up to 1 minute but there will be some additional questionnaires and possibly some delays in between rounds. We think that the entire part-2 HIT will take approximately 25 minutes.

In each of the 11 rounds you will have the opportunity to earn tokens. After part 2 is complete the computer will randomly select ONE of the 11 rounds and will convert the tokens you have earned in that round to an amount in dollars. The exchange rate will be 1 tokens = $0.10. Thus, the more tokens you earn, the more money you will be paid as a bonus (on top of the other payments). How many tokens you earn will depend both on your own decisions and on the decisions taken by the people you negotiate with.

In the following we will explain the setting and other rules. Please read these instructions very carefully.

Person A vs. Person B

During the negotiations you will either be in the role of Person A or in the role of Person B. If you are Person A you will be matched only with people who are in the role of Person B and vice versa. The role assignment is done by the computer at the beginning and is completely random. You will remain in the same role for the entire HIT.

Please note also that the identity of the person you are matched with will change from round to round because the computer is programmed to match "Person A" people and "Person B" people at random at the beginning of each new round.

Negotiation

In each round you will have 60 seconds to negotiate with the other person how to split a pie of 100 tokens. (The remaining time will be constantly shown on the screen). During the negotiation you will have the opportunity to send proposals of how to split the pie to your negotiation partner or to accept a proposal made by your negotiation partner. This is the only form of communication.

To make a proposal there will be a **slider** on your screen, which you use to choose your proposal, and a **"Submit" button**, which you then click to send your proposal to your partner. Whenever you send a new proposal, it replaces any proposal you have made before and only your new proposal will be relevant. Likewise, the computer will always display only the most recent proposal submitted by your partner. To accept your partner's current proposal just click on the **"Accept" button**. You can propose any possible division of the 100 tokens apart from an exact 50-50 split.

As soon as either you or your negotiation partner hits the **"Accept" button** the negotiation ends and the division of the 100 tokens is implemented according to the accepted proposal. The round is complete and you just have to wait for next round to begin.

What happens if there is no agreement?

In the event that the 60-second time limit is reached without either partner accepting a proposal, there is a **“no agreement”** payoff for Person A and for Person B. These payoffs will be visible on both partners’ screens from the start and during the negotiation. Please pay attention to these because *they will change from round to round* and they will often not be the same for Person A and for Person B.

Nuke treatment:

Person B’s final decision

Finally, in case of no agreement, Person B (and only Person B) has to make one more choice. Person B can either:

- Click on the **“Continue”** button to finish the round. Person A and Person B receive their **“no agreement”** payoffs, as explained above.
- Or click on the **“Reduce tokens” button** to finish the round. In this case Person A and Person B **do not** receive their “no agreement” payoffs. Instead, Person A obtains **0 tokens** and Person B obtains **9 tokens** as their earnings for this round.

Both treatments:

Practice rounds

Before the actual negotiations begin, the computer will take you through 5 practice rounds. Their sole purpose is to familiarize you with the setting and the computer interface. In these practice rounds you will not be matched with real people. Instead, your partners will be simulated by the computer, and the outcomes in the practice rounds are not relevant for your earnings today.
The very first round is a tutorial round in which the computer instructs you which choices to take. In the remaining four practice rounds you can make your own choices to try out the interface. You will be reminded when the practice rounds are over and the real rounds begin.

Nuke treatment:

The prediction game

Before we start with the real rounds, we ask you to play the prediction game.
This works as follows. We are running this study today with a large number of people, and we expect there to be instances in which the two negotiation partners do not come to an agreement. The basic question is this:
How often do you predict will people in the role of Person B choose the “Reduce tokens” option after a no agreement outcome?
As your prediction may depend on the no agreement payoffs, which will change from round to round, we ask you to make predictions for all 11 rounds.
When tomorrow’s part 2 has finished, the computer will randomly select **one** participant. We will then compare the predictions of that participant to what Person B people actually have done after a no agreement. **If you get selected, we will pay you up to \$100 on top of your other earnings!** Your payment will depend on how accurate your predictions have been. If your predictions deviate by on average no more than 1% from what really happened you’ll receive \$100. If they deviate by up to 2% you’ll get \$90 and so on. If your predictions are off by more than 10% on average, you will not receive

any additional payment.
Note that we will compare your predictions only to those no agreement cases where you have not been involved yourself (either in the role of A or B). Thus, you do not have to worry that your own actions or those of your negotiation partner affect the accuracy of your predictions.
Please enter your predictions as percentage figures (integers between 0 and 100).

- Entering 0 means: *I predict that for these "no agreement" payoffs, no B-person in this situation will choose the "Reduce tokens" option.*
- Entering 100 means: I predict that for these "no agreement" payoffs, all B-persons in this situation will choose the "Reduce tokens" option.
- Entering 47 means: I predict that for these "no agreement" payoffs, 47% of all B-persons in this situation will choose the "Reduce tokens" option.
- and so on

By the way, note that the order of the 11 rounds in the main part of the survey may differ from the list shown here.
Good luck!

Nuke treatment, B players:

Predicting own behavior

Please predict how likely it is you will use the option in each of the cases.

Welcome to part 2

Part 2 will begin in a few moments. Remember that you will interact live with other people who will be logged on at the same time. This means that you will be required to stay online for the entire duration of part 2 and you will be prompted to make choices throughout. There will be no time for toilet breaks etc. If you stop responding to the prompts, the computer will disconnect you and you will not be paid.
Don't let this happen! Don't become a dropout!
Each screen you will see will have a time limit. 60 seconds are available for a negotiation, as explained in part 1. If time runs out here, this simply means that you and your partner cannot agree on how to share the 100 token pie. This may very well happen on occasion. However, there are also (clearly visible) time limits for the screens **in between the negotiation rounds**. Please make sure that you respond to the prompts here.
If you wait for the countdown timer to reach 0 here, the computer will assume that you have left the survey and will terminate the connection. If this happens you will not be paid!
Provided that you do complete part 2 we will pay you within 48 hours and in two ways:

- First, you will receive a $1 reward for completing part 2.
- Second, after this HIT the computer will randomly select ONE of the 11 rounds and will convert the tokens you have earned in that round to an amount in dollars. The exchange rate will be 1 tokens = $0.10. Thus, the more tokens you earn, the more money you will be paid. How many tokens you earn will depend both on your own decisions and on the decisions taken by the people you negotiate with.

What happens if your assigned partner drops out during part 2?
If this happens the computer cannot let you continue with the survey because the software requires an even number of active participants at all times. However, we will still pay you!

D Demographics & Questionnaire

D.1 Control question

After reading the instructions participants will be asked to answer the following control questions. Individuals who fail more than one control question are routed to the end of the experiment and cannot take part in the experiment.

Please answer the following control questions. Please note: if you fail more than one question you will be excluded from any further participation and any payment.

1. How many rounds are you going to play after the practice sessions and how many of these rounds are going to determine your payoff?:

(a) 11 rounds with 11 rounds determining your payoff
(b) 1 round with 1 round determining your payoff
(c) 15 rounds with 1 round determining your payoff
(d) 11 round with 1 round determining your payoff
(e) 15 rounds with 15 rounds determining your payoff

2. Which role can you have during today's experiment?

(a) Either A or B. But my role will be fixed.
(b) Either A or B. I will play both roles.
(c) Either A or B or C. I will play all three roles.
(d) Either A or B or C. But my role will be fixed.
(e) Either A or B or C. But I will play only two of the three roles.

3. What is your task?

(a) To negotiate how to split a pie of 100 tokens
(b) To decide how much of your 100 tokens the other play will receive
(c) To negotiate how to split a pie of 150 tokens
(d) To decide how much of your 150 tokens the other play will receive
(e) To negotiate how to split a pie of 300 tokens
(f) To decide how much of your 300 tokens the other play will receive

4. What happens in case of an agreement?

(a) The negotiation ends immediately and the division of the 100 tokens is implemented according to the accepted proposal
(b) The negotiation ends immediately and the division of the 150 tokens is implemented according to the accepted proposal
(c) The negotiation ends immediately and the division of the 300 tokens is implemented according to the accepted proposal

(d) The negotiation ends immediately and the other player receives the accepted amount from your 100 tokens
(e) The negotiation ends immediately and the other player receives the accepted amount from your 150 tokens
(f) The negotiation ends immediately and the other player receives the accepted amount from your 300 tokens

5. What happens in case there is no agreement?

(a) Each player obtains their individual personal numbers from stage 2 as their payoff.
(b) Each player obtains a random number between zero and 150 as their payoff.
(c) Each player obtains a payoff of zero.
(d) You can keep your 100 tokens and the other player obtains a payoff of zero.
(e) You can keep your 150 tokens and the other player obtains a payoff of zero.
(f) You can keep your 300 tokens and the other player obtains a payoff of zero.

6. What is the last step in a round when no agreement has been reached?

(a) C can reduce the payoff of A or B.
(b) A can change the results.
(c) B can change the results.
(d) Nothing. The round finishes.
(e) A and B can change the results together.

D.2 Post-experiment questionnaire

1. Were the instructions clear?

(a) Yes
(b) No

2. In case you experienced any hitches during the survey, please describe them here.

3. On a scale from 1 to 7, is this a sort of situation in which people ought to "play fair" in your view or is it acceptable if people try to get the best outcome for themselves and make the best use of their bargaining power?

4.Please give a brief description of your bargaining strategy.

Nuke treatment, A (B) player:

5. Overall, how worried were you that person B (do you think person A was that you) might use the "Reduce tokens" option? (Overall, how worried do you think person A was that you might use the "Reduce tokens" option?)

- 1 (Not at all), ..., 7 (Extremely worried)

(5B. Overall, how tempted were you to use the "Reduce tokens" option?)

- 1 (Not at all), ..., 7 (Extremely tempted)

6. What do you think would be the **main** reason for person B (other people in the role of person B) to use the "Reduce tokens" option?

(a) Being curious and wanting to try it out.
(b) Anger about the way person A behaved during the negotiation.
(c) No anger but wanting to teach person A a lesson.
(d) Clicking the wrong button by accident.
(e) Ensuring that he/she would earn more than person A.
(f) Other reasons:

6a. Any **further** reasons? (You can tick multiple boxes.)

(a) Being curious and wanting to try it out.
(b) Anger about the way person A behaved during the negotiation.
(c) No anger but wanting to teach person A a lesson.
(d) Clicking the wrong button by accident.
(e) Ensuring that he/she would earn more than person A.
(f) Other reasons:

(6B. When you chose or considered to choose the "Reduce tokens" option, what was the main reason for doing so?)

(a) Being curious and wanting to try it out.
(b) Anger about the way person A behaved during the negotiation.
(c) No anger but wanting to teach person A a lesson.
(d) Clicking the wrong button by accident.
(e) Ensuring that he/she would earn more than person A.
(f) Other reasons:

(6Ba. Any **further** reasons? (You can tick multiple boxes.))

(a) Being curious and wanting to try it out.
(b) Anger about the way person A behaved during the negotiation.
(c) No anger but wanting to teach person A a lesson.
(d) Clicking the wrong button by accident.
(e) Ensuring that he/she would earn more than person A.
(f) Other reasons:

7. How did person B's (the) "Reduce tokens" option affect your bargaining strategy? (You can tick multiple boxes.)

(a) It didn't make the slightest difference.
(b) I think I was **more willing to accept** an offer from person B (A) than I would have been otherwise.
(c) I think I was **less willing to accept** an offer from person B (A) than I would have been otherwise.
(d) I think I made **more generous offers** to person B (A) than I would have done otherwise.
(e) I think I made **less generous offers** to person B (A) than I would have done otherwise.
(f) Other:

8. How do you think did the Reduce tokens' option affect person B's (A's) bargaining strategy? (You can tick multiple boxes.)

(a) It didn't make the slightest difference.

(b) I think person B (A) was **more willing to accept** one of my offers than he/she would have been otherwise.
(c) I think person B (A) was **less willing to accept** one of my offers than he/she would have been otherwise.
(d) I think person B (A) made **more generous offers** to me than he/she would have done otherwise.
(e) I think person B (A) made **less generous offers** to person B (A) to me than he/she would have done otherwise.
(f) Other:

E Deviation from the pre-registration.

The reported experiment follows the pre-registration reported at https://aspredicted.org #41190 and #89754 very closely. However, we deviate from it in the following details:

1. In our pre-registration (#41190), we initially outlined plans to conduct five subsequent conditions targeting different channels. However, upon completing the main experiment, we realized that this plan was overly ambitious and technically unfeasible. Specifically, due to the limited pool of Mturkers, the number of participants required for each additional condition would have been prohibitively large. As a result, we decided to focus solely on the main condition of interest and did not proceed with any additional treatments, with the exception of the Threat treatment, which was registered at #89754.

2. We preregistered a set of variables relevant to the bargaining process (e.g. Time of last offer, Identity of initiating party, etc.). Even though we clearly mark them as secondary in the pre-registration, we use those variables only minimally in our current paper, as they do not provide much interesting insight.