Harm Intended: conflict-generated utility as a consumption good

Introduction

Can an individual derive enjoyment from conflict? Does an individual derive psychological utility from harm inflicted on an adversary? The existing rationalist bargaining literature on war is built upon a fundamental assumption that conflict is inherently costly and inefficient. Thus, without a bargaining failure, both the winning and losing sides of a dispute should prefer to reach a bargain, avoiding the costly conflict. Challenging the conventional inefficiency assumption, this paper offers an alternative individual-level psychological theory of conflict to explain why individual actors are willing to engage in conflict despite its costliness. The theory incorporates Conflict-Generated-Utility (CGU) into individual actors’ expected utility function as a consumption good. Specifically, the CGU theory makes three hypotheses: (1) for individual actors, ex post efficient conflict is possible. In other words, individual can derive greater utility from conflict than a bargained settlement (2) In conflict, an individual can derive greater utility from inflicting more harm on an outgroup adversary even when holding the value of the dispute constant. (3) The size of CGU depends on the depth of individual identification to his/her ingroup.

Thus, this paper is organized as follows. The first section reviews five main strains of the bargaining literature of war including information asymmetry, the commitment problem, the principal-agent problem, issue indivisibility, and costly peace. I hope to show through this literature review that five strains of the existing bargaining theories are all variants of an overarching bargaining inefficiency explanation for war, which explain the occurrence of costly conflict as a function of inefficiencies in the bargaining process. The possibility of conflict efficiency, on the other hand, has been prematurely dismissed. I then highlight recent advances in behavioral economics and experimental social psychology that suggest the potential existence of CGU.

In the second section, I articulate my theory and hypotheses in detail, explaining the empirical expectation of a CGU theory. In this section, I also offer a stylized theoretical framework to illustrate how individual-level psychological utility could aggregate into state-level policy outcomes under different regime types using a principal-agent framework.

Finally, the paper outlines two conjoint experiment designs closely structured around the bargaining model to test empirically the existence of conflict-generated utility. The model-based experiment also aims to assess the aggregation of CGU with different risk, cost, and utility conditions. Moreover, the experiment employs a priming treatment to randomly manipulate an intermediate variable – national ingroup identification, so as to explore how the effect of CGU is contingent on nationalism. Lastly, I also discuss ways of integrating cash lottery incentive into the experiment instrument so as to alleviate some of the external validity concern associated with experiments that use hypothetical conflict scenarios.
Literature Review: The Bargaining Models of War

The Inefficiency Puzzle:

Though contemporary bargain literature flourished following the publication of Fearon’s seminal article in 1995, the comparison between war and bargain is not new. Throughout Arms and Influence, Schelling (1966, 7) made the analogy between war and bargaining, suggesting that diplomacy is essentially a bargaining process to produce an outcome acceptable to two states; Thus, war is a “dirty, extortionate, and reluctant” bargaining process. Schelling added that a state capacity for violence is essential its bargaining power. Waltz (1979, 114) makes an analogy between war and collective bargaining. He points out that strikes occur despite being costly to both labor and capital. He goes on to suggest that the existence of strike as a rare but possible option influences day-to-day labor relations just as the shadow of war influences peacetime international relations.

The contemporary formal analyses of war as bargaining are often inspired by Rubinstein’s (1982) bargaining model, originally designed to model the partition of a disputed good in an iterated noncooperative game. In particular, the Rubinstein model highlighted the role of uncertainty over cost in determining bargaining outcomes. Thus, at its core, the bargaining literature is an endeavor to model disputes in which two players are trying to strike a zero-sum bargain while being uncertain about their opponents’ probability of winning and resolve. Bargaining theories of war since has been based on a revised Rubinstein model that contains conflict as an outside option in the form of a costly lottery (Powell 2002, 6). Let the value of a dispute be u and chance of winning for one player be p, and the cost of a conflict be c, the model formalizes to:

\[ E = p \cdot u + (1 - p) \cdot 0 - c \]

\[ E = p \cdot u - c \]

As Fearon points out in his seminal paper, if we assume the cost of conflict is always positive, the expected payoff from conflict \( E = p \cdot u - c \) should always be lower than the payoff of a bargain that reaches the same result \( E = p \cdot u \). This model reveals that conflict should result in inefficient outcome for both the winner and the loser. One state or another may prefer fighting to the status quo, but no rational state could prefer fighting to a negotiated settlement that reach the same outcome. This model highlights an key puzzle—if conflicts are inherently inefficient, why do self-interested state engage in these conflicts? This inefficiency puzzle lead to an explosion of research since the mid-1990s.

Spurred by the inefficiency puzzle, rationalist bargaining literature have developed five main strains of bargaining inefficiency theories regarding the cause of war: asymmetric information (Fearon 1995), commitment problem (Powell 2006), principal-agent problem (Jackson and Morelli 2007; Mesquita et al. 2005), issue indivisibility (Goddard 2006; Hassner 2003; Toft 2014), and more recent investigation into the mechanism of costly peace (Coe 2012). These theories explained war as a result of uncertainty and changes in two opposing states’ winning probability (p), and resolve (c).
Information Asymmetry

The most prominent bargaining theory of war centers around the problem of information asymmetry. The crux of the information asymmetry is that because states are uncertain about each other’s probability of winning (p) or resolve (c) in conflict, bargaining sometimes fail. Fearon (1995, 1997) and others point out that because resolve provides bargain power, states have the incentive to misrepresent their resolve. Consequently, states resort to costly strategies such as sinking cost and tying hands to provide credible signals of their resolve. When credible signals are conveyed, peaceful compromise can be achieved through bargaining. When signals are not seen as a credible, war starts as a result of bargain failure. A variety of theories on costly signal mechanism has grown out of the information asymmetry problem focusing on two types of signaling strategy: tying hands and sinking cost. Both signaling strategies increase the credibility of signals by increasing the risk of war.

A typical hand-tying signaling strategy involves generating audience cost (Fearon 1994). Assuming that leader who backs down from a public threat will be punished by her domestic audience, audience cost conveys information on the true resolve of a state. As a hand-tying strategy, if the audience cost is large enough, the payoff structure of an unresolved state can change sufficiently to become a resolved one. Therefore, audience cost is a tool for leaders of State 1 (S1) to credibly signal S1’s resolve in an international bargain with State 2 (S2) (Slantchev 2011, 48). War occurs when S2 deems S1’s audience cost signal not credible, seeing the leader of S1 as unaccountable to the domestic audience, thus not vulnerable to domestic punishment. Following this logic, democratic leaders that are vulnerable to losing election are capable of generating greater audience cost than their invulnerable authoritarian counterparts (Fearon 1994, 581; Schultz 1999; A. Smith 1998). Authoritarian leaders’ threats are hence more likely to be dismissed as not credible, increasing the chance of conflict (Schultz 1999). Weeks (2014, 44) revises the audience cost theory, pointing out that authoritarian states with nonpersonalist civilian machines also face domestic accountability and thus are capable of generating credible audience cost signal.

Another type of information asymmetry concerns uncertainty over the probability of winning. Fey and Ramsay’s (2011, 161) game-free analysis shows that uncertainty over the probability of winning (p) as a major source of bargain failure. One strategy to reduce this second type of uncertainty is military mobilization. Military movements, such as mobilization, forward deployment of troops can convey credible signal through two different mechanisms, sinking cost and enhancing a state probability of winning. A military mobilization and deployment can be financially costly, by incurring an ex-ante cost, a resolved state can demonstrate its value of the dispute (Fearon 1997). Moreover, appropriate physical deployment of troop by S1 can increase military readiness, increasing S1’s probability of winning, reducing S2’s probability of winning, thus deterring S2 from fighting by changing the payoff structure of both sides (Slantchev 2011, 66). As a result, a state that does not seek war nevertheless needs to engage in activities such as military movement which simultaneously make war more likely and deterrence signal more credible.
Commitment Problem

Commitment problem is the second source of bargaining failure. According to this theory, the lack of binding, enforceable agreement results in bargain failure and even the absence of bargain (Fearon 1995). A key source of commitment problem is the shifting power balance between states. For example, S1 with growing relative power would have decreasing incentive to uphold a bargained agreement with S2. Knowing S1’s propensity to renege, S2 is less likely to reach a bargain with A in the first place. Changing distribution of power incentivize declining S2 to wage preventive war, effectively locking-in S2’s share of the disputed good before S1 became too powerful to beat (Powell 2006). Tingley’s (2011) iterated resource division game experiment offered empirical support for commitment problem theories linking changing bargaining power and inefficient outcomes. The commitment problem also speaks to earlier IR literature on power transition and hegemonic transition as a cause of major war, a process in which rising and declining power find it difficult to reach a binding compromise (Copeland 2000; Gilpin 1981; Kugler and Organski 1989). Another related manifestation of commitment problem is pre-emptive war generated by first strike advantage. When there is first strike advantage, a state chooses bargain over fighting essentially reduces its own chance of winning, giving its opponent an opportunity to exploit the first strike advantage. Thus, first strike advantage could reduce the chance of reaching a bargain, making war more likely.

Commitment problem is exacerbated if bargaining is a cumulative process (Powell 2006, 195). In other words, if concession to an adversary would strengthen the adversary’s position to extract more concession, bargaining is more likely to fail. For example, British and French concession to the German demand for Sudetenland gives Germany an even stronger military position in subsequent crisis bargaining. By gaining Sudetenland, Germany not only gained additional population and economic power but also deprived natural defensive barriers of Czechoslovakia and Poland. Thus, the 1938 Munich concession empowered Berlin to extract additional territorial concessions from the Czech rump state and Poland in 1939. Because the probability of winning (p) is constantly shifting, inefficient conflicts rooted in commitment problem can occur even when states have complete information.

Domestic politics can also exacerbate the commitment problem. Powell’s analysis shows that power shift among domestic political factions can make an internationally efficient bargain domestically unviable. For example, assuming a state’s foreign policy gain is distributed between two domestic political factions, a ruling faction, and an opposition faction. The ruling faction gets a larger share of the pie than the opposition faction. A successful bargain that involves compromise with foreign adversaries might adversely affect the current ruling faction’s hold on domestic political power enhancing the chance of an opposition takeover. But the current opposition faction cannot credibly commit to offering the current ruling faction an adequate share of the benefit if there is a change of government (Powell 2006, 190–92). As a result, the current ruling faction would prefer the larger share of a smaller pie by going to war while in office than the smaller share of a larger pie generated by peaceful bargain after losing office.

Leadership turnover can also exacerbate the commitment problem. Wolford (2007, 2012) has advanced a model of leadership turnover driven commitment problems. When leader turnover occurs, a successor may decide to overturn commitment made by her predecessor. Different leaders might have varying levels of resolve. A more resolved leader may demand greater
concession from foreign adversary than her predecessor, thus demand a renegotiation of an inherited agreement. Curiously, Wolford (2012, 524) finds the combination of an irresolute incumbent and a resolved successor in S2 might encourage S1 to offer preemptive appeasement, bolstering the domestic political position of an irresolute incumbent in S2. In other words, irresolute leaders sometimes improve the chance of reaching a peaceful bargain. Nonetheless, leadership turnover creates two additional paths to war. First, when a resolved incumbent leader in S2 faces domestic political challenges from a similarly resolved successor, the incumbent might demand more than S1 is willing to concede so as to increase her chance of domestic political survival, resulting in bargain failure. Secondly, S1 may anticipate the leadership turnover in an adversarial state, S2, that would result in a less resolved leader. Prefer to bargain with a less resolved successor than the more resolved incumbent, S1 may hold out from negotiating with the current leader of S2 or decide to fight S2 so as to facilitate an advantageous leadership turnover in S2. Therefore, the prospect of leadership turnover provided two additional paths to war associated with commitment problems.

Principal-agent Problems

A third source of bargaining inefficiency is the principal-agent problem. In an international bargaining context, the principal-agent problem occurs when the leader of a state has a payoff structure that is different from the state (Bueno de Mesquita et al. 1999). Smith’s (1996) formal analysis provided a framework to understand the leader’s perverse incentive in diversionary international military adventure when conflict affects election outcomes. Empirical research provided support for this perverse incentive model. For instance, Trager and Vareck’s (2011) experimental research shows successful extraction of foreign concession in crisis bargain increases domestic support for a president. In addition, a president backing down from a threat faces more severe punishment than a president who followed through a threat but lost a war, indicating a domestic political incentive for a leader to start a war that they cannot win (Trager and Vavreck 2011, 536).

Jackson and Morelli (2007) offered a general formal model of leader bias in a principal-agent framework. According to this model, when a leader’s benefit-cost ratio in a conflict is greater than that of the state, the leader has a positive bias towards conflict. A positively-biased leader would go to war even if they are completely informed about their low probability of winning. Paradoxically, when it is possible for states to extract concession in bargaining, states have an incentive to select positively-biased leaders who would strengthen a state’s bargaining position (Jackson and Morelli 2007, 1356). One example of the principal-agent problem is diversionary war (Levy 1989). Richards et al. (1993), Hess and Orphanides (1995), and Tarar (2006) have proposed a number of formal models explaining diversionary conflict as competence signal in domestic politics; They argue that leaders weigh the office-holding benefit against the cost of war and other factors. Using an explicit principal-agent framework, Downs and Rocke (1994) contend that when a leader’s risk of losing office is high, risky war provides an opportunity for leaders to engage in “gamble for resurrection” that could potentially help them regain domestic support.

In addition, the degree of principal-agent problem in crisis bargaining contingent on regime types. It is generally assumed that democratic leaders are more faithful agents of the domestic principal than autocratic leaders (Bueno de Mesquita et al. 1999; Downs and Rocke 1994, 363;
Therefore, autocrats are more likely to have their own interests related to an international dispute beyond the dispute itself such as their own political survival. An empirical examination of office-seeking incentive and conflict proposed by Chiozza and Goemans (2003) suggests that, in contrast to diversionary theories, higher risk of losing office makes it less likely for a leader to initiate conflict; They argue that the high risk of losing office, help explain why democratic leaders initiate less conflict. Goemans (2000) offered a different theory of leader incentive combining risk of losing office and leader’s personal fate after losing office. He points out that unlike democratic leaders, who enjoy personal safety after losing office, or autocratic leaders who have a low risk of losing office, leaders of mixed regimes could lose office after failed conflict and tend to suffer extreme punishments such as imprisonment and death when they are deposed. Therefore, leaders of mixed regimes have particularly strong incentive to fight losing war (Goemans 2000). Though many of these principal-agent theories are framed as audience cost theories, they are primarily concerned with the divergence of incentive between leader agent and domestic public principal as causes of war. Seen in this light, unfaithful agent with perverse incentive creates bargaining inefficiency, shrinking the margin of mutually acceptable outcome between two rational states.

**Indivisibility**

Indivisibility theories perhaps represent the greatest departure from the canonical bargaining model. It highlights an extreme form of bargaining inefficiency—states are completely unwilling to bargain over certain issues. If a disputed issue is indivisible, war became the only way to settle the dispute. Territory closely linked to national and ethnic identity is considered a typical indivisible good, which is difficult to be bargained over. The prime example of an indivisible territory is the Israeli-Palestine dispute over the sacred sites of Jerusalem. Based on cases of dispute over sacred religious sites, Hassner (2003) developed a theory of indivisibility in conflict. He argues sacred space cannot be compromised because any settlement less than the whole, significantly reduced the value of an indivisible good. In addition, indivisible goods are nonfungible, which means that they cannot be substituted or exchanged for side payment. Toft (2005, 2006) proposes an issue indivisibility theory of war, asserting that whether a territorial dispute is closely linked to ethnic identity determines whether a dispute would escalate into a violent conflict. War is also more likely when any amount of territorial concession is seen by a state as a threat to their overall sovereignty. Likewise, ethnic groups or state are more likely to consider historical homeland an indivisible issue and any competing claim as fundamentally illegitimate even if the same actor has an interest in settling the dispute (Goddard 2006, 44). More recent empirical studied also provide some support for the notion of indivisibility. Ginges et al. (2007) conducted a series of vignette experiments, presenting hypothetical side-payment offers to Palestinians and Israelis to settle their long-standing territorial dispute. These side payment offers are not only ineffective in promoting compromise, but they also produce a blowback effect that makes treated subjects more supportive of violence as a solution to the disputes.

Territorial dispute is not the only type of potentially indivisible issue. A side payment experiment in Iran shows that side payment treatment increase Iranian subjects’ disapproval of a compromise on its nuclear program among subjects who considered the development of a nuclear program a sacred issue (Dehghani et al. 2009). Moreover, Weigiger (2013) argues that sometimes state believes the leader of an opposing state are undeterrable “war-lover”. As a
result, nothing short of the complete overthrowal of a war-loving regime through an unlimited war could address the threat. Although Weisiger characterized his theory as a “dispositional commitment problem”, his argument is fundamentally an indivisibility theory revealing the impossibility of an compromising bargain when the minimally acceptable result for one side is the total annihilation of an adversary. Lastly, the claim of indivisibility can be used as a bargaining technique (Tarar and Leventoğlu 2009). Since states attach extremely high value to an indivisible good and are resolved to fight, all states have the incentive to misrepresent that a disputed issue is considered indivisible good. Thus, similar to other types of information problem, indivisibility generates additional uncertainty, increasing the chance of conflict when states mistake a truly committed actor for a pretender.

**Costly Peace**

Finally, some researchers have also noticed that bargained peace comes with its own costs. Earlier works by Powell (2006, 192) categorized this cost of preserving the status quo as a special type of commitment problem that involves a tradeoff between short-term deterrence and long-term bargaining position. At time t, S1 that spends $D_{1t}$ amount of resource on defense and deterrence would enjoy a stronger position in bargaining with S2, if $D_{1t} > D_{2t}$. However, by spending more on defense at t, S1’s economic growth suffers, resulting in a lower amount of overall resource at time t+1 (Powell 2006, 172). Combining deterrence cost with the commitment problem, investment in defense and bargaining in the short term undermine a state’s long-term security. Hence, S1 has incentive to start a war before t+1, capitalizing its temporary advantage.

A concrete example would be the Soviet Union’s arms race with the United States during the Cold War. In order to obtain parity and advantage over the US in both nuclear and conventional armament, the Soviet Union invested heavily in defense between the 1960s and the 1970s at the expense of its economic development. The growth of Soviet capability strengthened its bargaining position in the 1970s. However, by favor guns over butter, the growth of the Soviet economy suffered, exacerbating its overall resource constraint vis-à-vis the United States in the 1980s. Consequently, Moscow can no longer keep arming in the 1980s, weakening its bargaining position. Thus, desirable peace at present is achieved at the expense of desirable peace in the future. Similarly, if the cumulative cost of defense outweighs the cost of immediate war, a rational state would also prefer fighting. Slantchev’s (2011, 229–30) formal analysis shows a paradoxical result— the high cost of war incentivizes states to pay more for deterrence and maintain peace, but more deterrence simultaneously makes peace more expensive. Similarly, Coe (2012) points out that peace is sometimes more costly than war given the cost of arming, imposition and predation. Recent work of Fearon (2018, 554) proposed a similar costly peace theory, suggesting a state’s peacetime investment in defense can end up reducing its value of the status quo. By revising the cost side of a utility function, these costly peace theories reveal that conflict might provide net gain when the cost of armament or deterrence become exceedingly high.

**Missing Piece of the Puzzle: conflict-generated-utility**

In sum, the bargaining literature has developed five strains of theories to explain the causes of war. To varying extents, all five strains can be categorized as variants of a bargaining inefficiency explanation for war. For information theorist, war occurs because of the inefficiency
in the signaling process in an uncertain environment where states have the incentive to misrepresent their resolve and power. For researchers of commitment problem, bargaining is sometimes an inefficient process to reach a binding enforceable agreement, particularly in the contexts of power shift or cumulative bargaining. For researchers of principal-agent problem, bargaining inefficiency stems from the inconsistent interests of the state principals and their leaders, who might prefer risky or losing war, to losing office. For proponents of issue indivisibility, war occurs because bargaining is not useful at all in settling sensitive disputes central to national or religious identity. Finally, researchers of costly peace highlight dynamics that makes the bargaining process, such as investment in armament and deterrence, more expensive than war itself. All five theories explain the occurrence of costly conflict by describing sources of inefficiencies in the bargaining processes. However, none of these bargaining inefficiency theories have revised the utility parameter in the expected utility function or have considered the alternative possibility of conflict efficiency.

What is missing from this debate is the consideration of conflict efficiency and the role of psychological utility in an individual’s payoff structure. Apart from passing references, the bargaining literature of war and broader IR has not seriously considered the idea that individual might enjoy the process of conflict, thus, deriving utility from war. Though much existing research including Fearon’s (1995) seminal work noticed the possibility that people enjoy the experience of conflict, almost all dismissed this possibility by presenting a simplistic “why are we not all dead” critique (Jervis 1978, 170). For example, noting conflict-generated utility as an alternative explanation to bargaining inefficiency theories, Weisiger (2013, 15) dismissed the idea of conflict as a consumption good. He asserts that if people enjoy the experience of fighting, wars would occur all the time which defies the empirical reality that wars are rare. However, this critique takes the idea of conflict as consumption good out of the rational decision-making context, even when people derive utility from conflict processes, the overall gain from fighting still needs to outweigh the cost of conflict. Thus, following the rule of price and demand, if the cost of conflict is high, we should expect individual actors, including those deriving utility from conflict, to engage in less conflict. Moreover, if conflict is a consumption good, it should follow the rule of diminishing marginal utility. Therefore, we should expect peace after prolonged, major conflict rather than constant fighting.

Granted, beyond the rationalist bargain literature, there has been IR research that explored cognitive sources of disagreement over the value of disputes. In particular, behavioral economics offers insights on endowment effect which may suggest the owner of a disputed good, such as a piece of territory, values the same good more than the challenger. But endowment effect only suggests there are different pricing of dispute for each side (Kahneman, Knetsch, and Thaler 1990, 1991; Levy 2007, 95). Prospect theory continue to regard utility as exogenous to conflict processes. If we assume conflict is not a consumption good, conflict remains inherently inefficient even for the owner who is under the spell of the endowment effect. As a study of information processing, prospect theory can be seen as an extension of information theories of war. Since behavioral economics insights reveal additional difficulties in bargaining processes, the revision of the bargaining model that draws on prospect theory still falls within the boundary of the existing bargaining inefficiency explanation.

Two areas of research beyond political science lend support to my suspicion that conflict can be a consumption good, from which individual derive utility. Firstly, behavioral economics studies
of bargaining behavior reveals that, shaped by social preference such as fairness, individual actors’ expected utility is linked to payoff of their counterparts (E. Fehr and Schmidt 1999; Levine 1998). In particular, recent behavioral economics experiments have identified individual willingness to engage in costly punishment as an important source of maintaining cooperation within large social groups (Ernst Fehr and Gächter 2000, 2002; Henrich et al. 2006; Rockenbach and Milinski 2006). These recent experimental studies have found that, in contrast to the canonical model of the individual as self-regarding utility maximizer, individual subjects across countries are willing to pay real monetary prices to punish those who defect in third-party punishment game. Some studies also identified potential genetic and neural basis of costly punishment behaviors (McDermott et al. 2009; Sanfey et al. 2003). Studies have also found that costly punishment is sustained even when less costly means of norm enforcement is available (Rockenbach and Milinski 2006). Prosocial individuals strongly identify with their group and community are especially willing to engage in costly punishment out of altruism. These findings beg the question – whether willingness to engage in costly punishment for ingroup cohesion translate into the willingness to engage in costly conflict with the outgroup? Can individuals’ parochial altruism also make them willing to engage in costly conflict? These questions crucial to the study of international conflict remain to be answered.

Secondly, recent social psychology research on schadenfreude (counter-empathy), has discovered that individual can derive psychological satisfaction from observing the misfortunates of political, racial outgroups, as well as randomly assigned minimal groups, even though these misfortunates are costly for themselves (Cikara et al. 2014; Cikara, Bruneau, and Saxe 2011; Combs et al. 2009; Leach et al. 2003). Moreover, fMRI scans show outgroup harm activates the same area of the human brain, Ventral Striatum, that is associated with receiving rewards (Singer et al. 2006; Takahashi et al. 2009). Nonetheless, with its conceptual focus on the individual as passive observers, the schadenfreude literature stops short of exploring if the individual can also derive psychological utility from actively inflicting harm on the outgroup adversaries. The social psychology literature has also been focusing on investigating the determinants of schadenfreude, instead of its consequences. These two characteristics of the schadenfreude research have precluded the analysis of intergroup counter-empathy as a motivator in political conflicts.

While political scientists are searching for answers to the inefficiency puzzle, the missing piece that could solve the puzzle might be in our back pocket all along. If conflict itself can generate utility, conflict is no longer inherently ex-post inefficient; The implication is that conflict can occur between rational actors in the absence of information asymmetry, commitment problem, indivisibility, principal-agent problem, or costly peace. In other words, conflict is not necessarily a result of bargain failure; war can also occur due to the desirability of conflict. Moreover, the study of conflict-generated-utility may help explain anomalous cases of conflict onsets that are difficult to explain using existing bargaining models. It would allow researchers to examine the omitted parameter of psychological utility at the individual level that could lead to a preference for seemingly inefficient outcomes. Building upon insights of behavioral economics and social psychology studies, in the following section of this paper, I offer a psychological theory of individual willingness to engage in costly conflicts.

---

1 In a third-party punishment game (or a third-party dictator game (TP-DG)), A and B plays an ultimatum game, and C act as the third-party punisher. A can “defect” by offering B an unfair division of the cash incentive. C decide whether to pay her own cash incentive to punish A.
Competing Explanation: Risk-Seeking

Another potential explanation for inefficient conflict beyond existing bargaining literature is risk-seeking. To the extent that prospect theory and the bargaining literature has considered the psychological element in individual payoff structure, both literatures have attributed psychological sources of war to risk seeking. Indeed, due to methodological and conceptual limitations of the canonical approach, it is difficult to distinguish risk-seeking from CGU for the two explanations are observational equivalent and conceptually overlapping.

Firstly, in observational studies of historical wars and conflict, CGU and risk-seeking are observational equivalences. The existing research tends to model war-prone leaders as risk-accepting individuals in order to explain why they deviate from the basic bargaining model (Jackson and Morelli 2007). According to this risk-independent risk-attitude theory, leaders such as Napoleon or Hitler prefer costly military gambles to certainty equivalent (Powell 2002, 3). Though it is difficult to tease apart individual risk-attitude and other individual-level psychological utility that drive leaders to war in observational studies, there is an alternative explanation – rather than being risk-accepting, the Napoleons and Hitlers of the world derive utility from conflict processes. Treating conflict as a costly lottery, we can extend the gambling analogy. There are two causal paths that lead an individual actor to engage in more costly gambling: the first path is that the individual is more risk-accepting; the second path is that the individual enjoys the activity of gambling even if she is risk neutral or risk-averse. In other words, the war-proneness can be explained by the utility generated by the activity of conflict rather than risk-acceptance.

Secondly, due to the probability-independent definition of risk-seeking in much of the game theoretical literature (Fehr-Duda and Epper 2012, 569; McCarty and Heirowitz 2007, 39). Risk-accepting behavior is modeled as driven by a negative risk premium independent of probability parameter in the model. As a result, all actor willing to engage in ex post costly lottery (unfair bet) are categorized as risk-accepting. Thus, even if the probability of winning parameter in the bargaining model take on the value of 100 percent, the fact that an actor is willing to pay the cost of war still make such actor “risk-seeking”. Peculiarly, this conception of risk acceptance has nothing to do with risk/probability parameter in the model and is only a reflection of positive cost. In addition, given that risk-acceptance is treated as default explanation for inefficient conflict that are “out of scope” of existing bargaining literature, there is surprisingly little empirical research that has tested this risk-seeking explanation. Moreover, behavioral economics and political science experiments research has shown that probability-independent definition of risk attitude does not capture the reality of human behavior (Starmer 2000). In addition, prospect theory highlights heterogenous risk attitudes created by gain and loss perceptions, revealing patterns of risk aversion when pursing perceived gains and risk-accepting behavior in avoiding perceived loss. This risk attitude difference is linked to endowment effect and reference point updating, which explain why states might decide to fight in order to avoid loss (Levy 2007, 99). A more appropriate characterization of the probability-independent definition of “risk-acceptance” is, thus, “cost-acceptance”.

When political scientist studies conflict through the bargaining model, the substantive interests about risk attitude concerns whether an actor is sensitive to lower probability of winning or are risk-seekers that have a higher willingness to fight under lower probability of victory. Thus, a
more adequate definition of risk preference in the bargaining model of war should be probability-dependent (Fehr-Duda and Epper 2012), which is not only better suits the goal of substantive research, but is also shown to be a more realistic reflection of human behaviors in empirical research (Kahneman, Knetsch, and Thaler 1991; Kahneman and Thaler 2006; Par M. Allais 1953). For example, we should expect risk-accepting actor to engage in more high value gamble with lower probability of winning, compared with a moderate risk gamble with equivalent expected utility. Hence, this paper will seek to tease apart the risk preference explanation and the hypothesis of conflict as a consumption good by using a probability-dependent definition of risk attitude by manipulating the risk of winning in hypothetical conflict scenario, if the risk-seeking explanation is correct, I expect to see some respondents to be indifferent to lower probability of winning, or even prefer the option with lower probability of winning. It is worth noting that the risk-seeking explanation and CGU explanation is not mutually exclusive.
EPW Abstract:

- Literature (the null hypothesis): Bargaining theory of war, conflict is always ex post inefficient
- An alternative Hypothesis: Individuals derive psychological utility (CGU) from harm inflicted on an outgroup adversary, making conflict sometimes efficient.
- Research Design: priming and conjoint experiment.
- For more literature review please see the paper draft.

Theory: Conflict Generated Utility

The core concept of this project is conflict-generated utility (CGU), an individual-level variable. Broadly, I define CGU as the psychological utility an individual derives from conflict processes. This paper investigates one specific kind of CGU, the psychological utility an individual derives from inflicting harm on an adversary outgroup in conflict. The theory focuses on two types of individual actors, individual members of the public, and individual leaders of states. CGU is psychological rather than instrumental. In other words, harm inflicted on an adversary is the end, rather than a means to achieve material or strategic gains.

The connection between harm and war is not entirely new to the IR literature, from classical deterrence theory of Schelling (1966) to the recent formal analysis by Slantchev (2003), researchers of war has long noticed that the capacity to inflict harm plays a central role in interstate security bargain, even when the harm inflicted cannot alter the outcome of a nuclear or conventional war. Yet, few if any study investigated the possibility that, in addition to being means to an end, harm can be an end in itself. To the best of my knowledge, there has not been a theoretical or empirical study that explored if inflicting harm as a goal or motivation behind conflict decision making.

Theoretical Model of Conflict-Generated Utility:

I formalized the CGU theory and its observable implications, by revising Fearon’s (1995) canonical model. As the model suggests, the bargaining theory assumes that because conflict is inherently costly (C > 0), all else being equal, achieving the same result through bargaining is always preferable to conflict. Therefore, the first key hypothesis ($H_{01}$) of the bargaining model is that when the expected value of bargaining ($E_b$) is always greater than expected utility from conflict ($E_c$) (Fearon 1995, 388). In addition, because conflict is not a consumption good (Fearon 1995, 383) and value of the dispute is exogenous to the conflict process, a canonical bargaining model predicts that when conflict does occur, the harm inflicted on the adversary is epiphenomenal to the expected utility in conflict. Harm would only matter as a bargaining leverage or a source of information on power and resolve. In other words, all else being equal, individuals derive no greater utility from a conflict result in 0 adversary casualty and a conflict result in 10,000 adversary casualty ($H_{02}: E_{c1} = E_{c2} | harm_1 > harm_2$).
Bargaining Theory: (Null Hypotheses)  
\[ E = p \cdot U + (1 - p) \cdot 0 - C \]  
\[ E_b = p \cdot U \]  
\[ E_c = p \cdot U - C \]  
\[ H_{0_1}: E_c < E_b \]  
\[ H_{0_2}: E_{ci} = E_{cj} | harm_i > harm_j \]  

CGU Theory:  
\[ CGU = E_c - E_b + C \]  
\[ E_c = p \cdot U - C + CGU \]  
\[ H_{CGU1}: CGU > 0 \]  
\[ H_{CGU2}: E_{ci} > E_{cj} | harm_i > harm_j \]  
\[ H_{CGU3}: E_{ci} > E_{cj} | id_i > id_j \]  

The CGU is defined formally as the difference between expected utility through conflict and bargaining achieving the same result other than the cost of fighting: \( CGU = E_c - E_b + C \). The CGU theory proposes three hypotheses differs from the expectation of a canonical bargaining model. \( H_{CGU1} \): all else being equal, it is possible for individual to derive greater utility from costly conflict than bargaining. \( H_{CGU2} \): all else being equal, individual can derive greater utility from inflicting greater harm on an outgroup adversary. \( H_{CGU2} \) is a mechanistic hypothesis for the broader \( H_{CGU1} \). In other words, harm inflicted on the adversary is one mechanism through which conflict can be ex post efficient even if conflict can achieve the same outcome as bargaining. The third hypothesis is that the size of CGU is mediated by an individual’s depth of identification with one’s ingroup (\( id \)). The more one identifies with her ingroup, the greater CGU she will be able to derive (\( H_{CGU3}: E_{ci} > E_{cj} | id_i > id_j \)). This hypothesis is built on social psychology research that suggest group level emotion can only be experienced by individual that actually identified with one’s ingroup (Mackie, Devos, and Smith 2000). It is worth noting that the CGU theory is not identity-group specific. Given the pervasive ingroup-outgroup dynamics that exists between nations, ideologies, religions, ethnic groups, political parties, any one or multiple intergroup conflicts can provide permissive condition for CGU generation. One particularly important type of ingroup identification relevant to IR is national ingroup identification. Though other types of groups such as ideological and religious group may also have a role in international conflicts, given that states are the primary actors in international relations, national ingroup identification will almost always play an important role in conflict between states. Thus, one empirical expectation of \( H_{CGU3} \) is that more nationalistic individual will, on average, derive greater utility from conflict with state that belongs to an adversarial outgroup.

Aggregation Process:  

To construct a parsimonious model that illustrates how the individual level CGU could aggregate into state-level policy outcomes, I use a simple principal-agent problem framework to model the domestic politics of CGU. The stylized theoretical model is outlined in the 2 by 2 table below. I make a simplifying assumption that categorizes both principals—domestic constituency, and the agents—leaders into two ideal types: prudent and biased. Prudent principals and agents are domestic constituencies and leaders that have the payoff structure resembling the canonical bargaining model. In other words, the prudent types do not derive utility from conflict. Biased principals and agents, on the other hand, are domestic constituencies and leaders that have payoff structures incorporating CGU, who derive utility from conflict.

My theory expects conflict-generated utility at individual level to increase conflict through two pathways of preference aggregation: on the one hand, there is a leader-driven process in which leaders personally derive psychological utility from conflict. Because the domestic public in democracies has an interest in deterring and punishing excessively militant or pacifist leader,
they should place a check on leader behavior (Downs and Rocke 1994). Similarly, regime
insiders form an elite constituency in non-personalist civilian autocracies can form a domestic
constituency that holds the autocratic leader accountable (Weeks 2012). Thus, the first
permissive condition for a purely leader-driven process is that a leader, is unaccountable to its
domestic constituencies, faces little constraint on his/her foreign policy decision making power.
Hence, a purely leader-driven process should occur in personalistic dictatorships. Personalists
can translate their personal belief and wishes into policy with little opposition (Janis 1972;
Peceny and Beer 2002; Tetlock 1998; Weeks 2012). Because distribution of power and resource
revolve around the leader, institutions in a personalistic dictatorship can even amplify dictator’s
personal psychological bias through a “working toward the Fuehrer” mechanism in which state
functionaries will try to competitively outbid each other in fulfilling leader’s preferences
(Kershaw 2014). As a freewheeling agent, a personalist can act in ways that fit his/her personal
payoff structure, ignoring the cost-benefit calculation for the domestic constituency. In addition,
not all leaders behave in the same way within a similar institutional environment (Horowitz and
Stam 2014, 530). Therefore, the second permissive condition for a leader-driven process is that
leader must be able to derive utility from a given conflict. In practice, this means that leaders
have strong ingroup identification in the form of nationalism or Manichean ideological beliefs
(Gries 2014). The connection between national identification and intergroup CGU is intuitive,
highly nationalistic individual should have stronger intergroup biases against adversarial
outgroup states. Manichean ideological identification also enhances intergroup biases by
strengthening the us-versus-them mentality. Leaders such as US President Reagan and Bush, for
instance, think about foreign policy as contests of good versus evil, and their adversaries as “evil
empire” or “axis of evil”, not to be bargained with but to be destroyed. One might suggest that
some leaders such as Reagan uses Manichean language as a bargaining rhetoric to show resolve
and extract concessions from adversaries. However, if leaders of superpowers have incentives to
use the rhetoric of adversary harm as bargaining tools, it suggests that there is real credibility in
asserting adversary harm as a foreign policy goal. The potential strategic use of CGU rhetoric by
leaders would strengthen rather than weaken this theory.

On the other hand, a constituency-driven process of audience reward is more likely to occur
when a significant contingent of the domestic public derive utility from conflict, hence giving
leader political incentive to initiate conflict. In this situation, the domestic constituency, rather
than acting as a prudent principal, abandons its constraining responsibility, permitting and even
rewarding CGU-seeking behavior of the leader. Given that a biased constituency derives CGU in
addition to the value of any dispute, they would punish a prudent leader for being overly pacifist,
consequently rewarding leader for being belligerent. When the domestic public derives CGU,
leaders have an additional incentive to initiate conflict so as to obtain audience reward ($R$), which
formalized to:

**CGU Theory at the Leader Level with Prudent Constituency:**

Prudent Leader: $E_L = p \cdot U - C$

Biased Leader: $E_L = p \cdot U - C + CGU$

**CGU Theory at the Leader Level with Biased Constituency:**

Prudent Leader: $E_L = p \cdot U - C + R$
Biased Leader: $E_L = p \cdot U - C + CGU + R$

Therefore, even a leader who is not personally motivated by CGU would have a greater incentive to initiate conflict when facing a biased domestic audience. One example of this audience-driven process is the onset of the Spanish–American War, despite having a relative dovish leader, President McKinley, who prefers a bargained solution to the Cuban crisis to war, pro-war public opinion played a pivotal role in driving a US towards conflict (Gould 1982; Wilkerson 1967).

But the most war-prone situation involves not only biased leaders who personally enjoys conflict but also a biased constituency. In contrast to the leader-driven process, audience reward is more likely in democracies where public opinion plays an important role in the political survival of leaders. Depends on factors such as nationalism, emotion and war-weariness, public opinion of a state shift between being prudent and biased. If the domestic public in a state does not derive significant CGU, accountability to a prudent principal will provide a check on leaders’ conflict initiation even when leaders are personally biased. There are special cases in which leader that faces de jure constraints, such as elected leaders in democracies, who are able to ignore their domestic constituency, behaving as freewheeling agents. In these cases, these leaders can be treated as if they are personalist in spite of regime type.

It must be stressed that adversarial intergroup context is a crucial scope condition of the CGU theory. The CGU theory does not expect states’ leaders or public would derive utility from initiating conflict with a state that they perceive to be a member of their ingroup. For example, US public or presidents in the postwar era are unlikely to derive utility from conflict with Britain because Britain is seen as an ally and a fellow democracy. Instead, the US public or leader is more likely to derive utility from conflict with Russia, China or Iran because these states belong to adversarial outgroups because history of rivalry, regime type, or ideological differences.

<table>
<thead>
<tr>
<th>Principal: Domestic Constituency</th>
<th>Agent: Leader</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low conflict generated utility (Prudent Principal)</td>
<td>Freewheeling Agent</td>
</tr>
<tr>
<td>High conflict generate utility (Biased Principal)</td>
<td>Biased Personalist</td>
</tr>
<tr>
<td>A. Leader-driven process:</td>
<td>C. CGU play no significant role in conflict initiation</td>
</tr>
<tr>
<td>B. Constituency and Leader driven process:</td>
<td>D. CGU play no significant role in conflict initiation</td>
</tr>
<tr>
<td>Constrained Agent</td>
<td>E. Successful Constrain</td>
</tr>
<tr>
<td>Non-personalist executive And Democracies</td>
<td>F. Audience reward: Constituency-driven conflict initiation</td>
</tr>
</tbody>
</table>
Contributions:

A CGU theory bridges recent psychological approaches to the bargaining theory of war. It could offer a parsimonious framework to integrate a series of psychological motivation for war into the canonical model. Though one might quibble over if deriving utility from conflict is “rational” in the colloquial sense, CGU is compatible with the expected utility model. In fact, when seemingly “irrational” factors have consistent systematic effects on outcome, they can be studied using a methodologically rationalist approach based on strong rationalist assumption of utility maximizing. Even states fighting for non-material gains can still be broadly considered rational. Indeed rational states often fight over non-material gains such as glory, status, reputation (Dafoe and Caughey 2016; Hardie, Johnson, and Tierney 2011, 75; Jackson and Morelli 2011; Kalin and Sambanis 2018; Mercer 2010; Renshon 2016; Yarhi-Milo 2018). Drawing on social psychology research on identity, Sambanis, Skaperdas, and Wohlforth (2015) premise that the very process of war can significantly enhance a state’s international status, national unity, and state capacity. As a result, a peaceful international bargain cannot be reached even if it is materialistically more efficient. Building on these insights, moral psychologist Johnathan Haidt (2014) went so far as to recommend increase US competition with China as a way of reducing political polarization in the United States. Renshon’s (2017) empirical study seem to lend support to the idea that state might derive status gains from conflict. In particular, even losing state in conflict can gain status from war if they fought valiantly against a stronger opponent for exceeding the expectation of observers. Given the positional nature of status, it is unclear how desirable goods such as status can be transferred through peaceful bargain between states. More broadly, from individual voting, political contributions to charitable donations to aid organizations many costly behaviors of rational individuals are motivated by enjoyment of the activity (Andreoni 1990; Ansolabehere, de Figueiredo, and Snyder 2003; Barzel and Silberberg 1973). As Riker (1995, 30) puts it, “human being does many things simply for fun, excitement, and self-expression”; There is no reason why conflict cannot be one of them.

Furthermore, as an endeavor to engaging in cumulative modeling dialogue, this study builds on the insights of the existing bargaining theories of war, particularly the principal-agent theories that links domestic actors’ payoff structure to international security bargaining. A CGU theory provides an alternative analytic framework of studying conflict as a consumption good. The theory goes beyond behavioral economics studies of intra-society costly punishment as a cooperation enforcement mechanism to explore individual decision in a costly conflict between adversarial states. The CGU theory also goes beyond schadenfreude literature’s limited scope of individuals as passive observers, contributing to an emerging research program of outgroup harm as a factor motivating individual actors in violent intergroup conflict (Cikara 2018; Halevy, Bornstein, and Sagiv 2008). Given the nature of international conflict as an intense form of intergroup conflict, conflict generated utility rooted in counter-empathy should be ubiquitous. Afterall, “political power is a psychological relation between those who exercise it and those over whom it is exercised” (Morgenthau 1947, 14). Since political conflict, including international and intrastate wars, are fundamentally intense conflicts between adversarial groups, the CGU theory help integrates a potentially ubiquitous psychological characteristic of conflict into the canonical rationalist model. I hope to show that psychological and rationalist studies of war can be an integrated and cumulative endeavor in which rationalist theory supplies specific, testable baseline models that can be continuously enriched by incorporating essential psychological factors. As Akerlof and Kranton (2000) point out in their seminal paper on
identities in economics, even if we maintain the rationalist assumption that individual are utility maximizers, the formal integration of psychological factors, such as individual identity, significantly change the result of canonical economic models. Therefore, even the most reductionist bargaining model would be incomplete without incorporating conflict-generated-utility into the payoff structure. Moreover, It has been established in game theoretical literature that, in zero-sum game, a strategy that minimize opponent payoff and the strategy that maximize one’s self interest should be consistent (Neumann 1928). In this sense, incorporating adversary harm into the bargaining model of war speaks to some fundamental insights of zero-sum games. The psychological utility generated by adversary harm serves a functional purpose in repeated interaction between adversarial groups.
Research Design:

The following section outlines a conjoint experiment design that tests the CGU theory by randomly assigning harm inflicted on an adversary in addition to parameters of the canonical bargaining model. Further, using priming treatment, a potentially important intermediate variable in intergroup conflict—national ingroup identification—is also randomly manipulated; half of the subjects will be assigned to receive a national ingroup identification prime embedded in the instructions before their choice tasks. Moreover, to enhance external validity of the experiment, cash lottery incentive will be incorporated into the conjoint tasks in ways that aligns with cost of the conflict decisions to provide a quasi-behavioral simulation of cost. The experiment instrument is designed to be implemented through Qualtrics using a convenient US sample recruited through online platform such as Amazon Mechanical Turk or Prolific.

Design Choice:

Advantages:
In studying the effects of psychological variables in IR, a key challenge is the multidimensionality of decision-making process—even if we cast institutions aside, it is unclear how an experimentally manipulated psychological variable would matter in the real world when there are multiple causal variables at play. Infer from the theoretical model, four key parameters could influence conflict behavior: the probability of winning (P), value of the dispute (U), the cost of conflict (C), and the harm inflicted on the adversary (HARM). National ingroup identification is a potentially important intermediate variable affecting CGU effect size (M). Conjoint experiments are designed to mirror the multi-dimensional decision making processes in the real world (Hainmueller, Hopkins, and Yamamoto 2013, 3). For this reason, in economics and marketing research, conjoint experiment is widely used to assess consumer choice in multidimensional decision-making (Gustafsson, Herrmann, and Huber 2007). For the same reason, conjoint experiments have been used to study a variety of multidimensional political decision-making processes ranging from international climate cooperation, attitude towards immigrant, political candidates, to party platforms (Bechtel, Genovese, and Scheve 2016; Bechtel, Hainmueller, and Margalit 2014; Green, Krieger, and Wind 2001; Hainmueller, Hangartner, and Yamamoto 2015; Hansen et al. 2015; Horiuchi, Smith, and Yamamoto 2018). Therefore, conjoint experiment would be particularly suited to tests CGU theory which is based on an expected utility model of conflict behavior with multiple model parameters already identified.

In experimental setting, hypothetical conflict scenarios allow researcher to overcome the selection problem of using observational data, testing several theoretically important variables at the same time. By randomly assigning all critical parameters of a conflict decision, the total effect of each treatment can be estimated as average marginal component effect (AMCE). The direct effect and indirect effect of each treatment with respect to each potential intermediate variable can be separated by estimating the averaged controlled direct effect (ACDE) and the eliminated AMCE (Acharya, Blackwell, and Sen 2018; Hainmueller, Hopkins, and Yamamoto 2013, 10,12), which offer insights into the causal mechanisms of CGU. For this specific study, a key advantage of simulating a multidimensional decision-making process using a conjoint experiment is to distinguish CGU from risk-seeking—two concepts that are observationally equivalent and conceptually overlapping.
In addition, the practical advantage of using a conjoint design, instead of a commonly used vignette experiment when administering multiple treatments, is that by placing variables in table form, it reduces satisficing. By placing treatments in table form, conjoint designs also help make subjects’ decision-making process more structured around variables of interests, reducing mental imputation. A conjoint design also makes a formal-model-based experiment manipulating multiple parameters more cost-effective than vignette experiments. Bansak et al. (2018) recent tests using Amazon Mechanical Turk and Survey Sampling International samples shows that subjects can perform up to a dozen or so choice tasks in a conjoint experiment without serious deterioration of response quality. Hence, researcher can significantly increase the statistical power of their experiment by multiplying the number of choice tasks each respondent will perform.

Tradeoffs:
There are a number of tradeoffs involved in this research design choice. Firstly, the estimation of total effect (ACME) in conjoint experiment requires the assumption of no-profile order effect and the assumption of no carryover effects. A conjoint experiment involving multiple choice tasks is a within-subject design, which create potential challenges such as practice effect or fatigue that can lead to the violation of the two assumptions. Fortunately, the no-ordering effect can be tested by examining if being in placed in option 1 or option 2 influences treatment effect. Likewise, no carryover effect can be tested by estimating the order effect of profiles appearing in difference choice tasks. Secondly, the estimation of ACDE as “direct effect” and eliminated ACME as “indirect effect” will not be able to distinguish two types of causal mechanism, mediation and moderation(interaction). Thirdly, the identity priming treatment will be an imperfect manipulation, which would only able to reveal the direction of indirect effect not its effect size. Thirdly, unlike lab experiment in which subjects’ willingness to engage in costly adversary harm can potentially be measured behaviorally and quasi-behaviorally, such as costly monetary punishment in third-party punishment game or hot sauce penalty, a survey experiment setting would only be able to measure treatment effect in stated preference. Subjects might be less motivated to harm a hypothetical adversary than an actual “adversary” player in lab setting. Lastly, from a theoretical perspective, individual level survey experiment does not capture institutional influence on the aggregation of individual preferences into state policy, which have to be investigated using alternative methods.

Experiment design:
The conjoint experiment designs are centered around a scenario of naval crisis. Before any treatment, the subject receives a very brief introduction of the scenario. The subjects were informed that a small US Navy flotilla is conducting a joint naval exercise with the Japanese Navy in international water 3000-mile west of Hawaii. A randomly assigned traditional adversary of the US, China or Russia, demands US Navy to immediately withdraw from the area, threatening to remove US fleet by force. To amplify the intergroup setting, subjects will be informed that they are responding to a previous choice by a respondent in a mirrored experiment being conducted in Russia or China, and that one of their response will influence the monetary payoff of their opponent. A set of typical Russian and Chinese names, respondent location, survey response date, will be used to describe such fictional survey opponent so as to enhance realism.
The experiment uses the US versus a concrete authoritarian state scenario for two reasons. Firstly, an authoritarian power will be chosen as the potential adversary in this case because being an outgroup adversary is the key scope condition for the CGU theory. A concrete scenario of US versus a traditional adversary drawn on real-world ingroup-outgroup dynamics, which abstract State A vs State B setting cannot achieve. Secondly, as a form of covariate control design, a concrete scenario provides more context for the subjects, which reduces imbalance in subjects’ mental imputation (Dafoe, Zhang, and Caughey 2018, 413).

**Priming Treatment: National Ingroup Identification (M):**
Since this study investigates international conflict as a form of intergroup conflict. The depth of individual identification with her own national in-group would be an important intermediate variable influencing treatment effect. As Acharya et al. (2018) and colleagues point out, in experimental setting, the causal mechanism consists of two components, the causal mediation effect (Imai et al. 2011; Imai, Tingley, and Yamamoto 2012), and the interaction (moderation) effect (Baron and Kenny 1986; A. S. Gerber and Green 2012; VanderWeele 2015). In order to assess the causal mechanism of how CGU affects individual decisions on the use of force, I slightly vary the identity of the target country in the naval crisis scenario for half of the subjects before conjoint tasks. The control group will simply be told that China/Russia demands the “joint fleet” to withdraw from the exercise area. While The treatment group will be briefed that the “US Navy” in particular is being challenged. The intent of this priming treatment is that stronger national in-group identification can be elicited by the framing of the question and influencing subject’s subsequent choices. The treatment would potentially increase CGU effect size. Here, the quantity of interest is the ACDE and eliminated ACME of adversary harm with respect to this identity manipulation. This particular ACDE will provide an estimation of the direct effect of adversary harm fixing ingroup identity priming treatment. The eliminated ACME estimates the indirect effect (this includes both mediation effect and treatment/mediator interaction effect) of ingroup identification as a causal mechanism linking adversary harm with expected utility.

It should be acknowledged that the manipulation of identity will be imperfect at best. Moreover, the description of a military conflict scenario inevitably elicits subjects’ national ingroup identification. Though the precise size of mediation and moderation effect cannot be recovered using this imperfect manipulation, the priming treatment can help discover the manipulation-induced indirect effect (Acharya, Blackwell, and Sen 2018), which provide meaningful evidence for the causal mechanism of a theory. To test if this priming treatment has the intended effect. Survey instrument measures national ingroup identification twice in the pre and post treatment questionnaires. The post-treatment measurement could serve as a manipulation check that validates if, all else being equal, the treatment delivered has moved national ingroup identification in a positive direction. The pre-treatment identity measurement, on the other hand, can serve as a fallback option for conventional heterogeneous treatment effect analysis should the identity manipulation turnout to be inadequate.

---

2 This assumes that the priming treatment does not have a perverse effect of reducing subject’s ingroup identification.
Conjoint Treatments:

The conjoint experiment manipulations are structured as follows. To simplify the interpretation of AMCE, ACDE, and eliminated AMCE, all parameter other than cost only takes on two levels: high and low. For each respondent, the treatment will be randomly ordered but constant throughout all choice tasks to reduce profile order effect.

Value of the dispute (U):
A naval encounter in international water scenario is chosen to avoid a hypothetical territorial loss that potentially produces confounding effects identified by prospect theory and issue indivisibility theory. Instead, the value of the dispute in this scenario lies in two types of deterrence utility: specific deterrence ($U_1$) and general deterrence ($U_2$). Here, specific deterrence refers to US’ ability to deter a specific adversary in the naval encounter. The use of force might be costly and risky, but even a valiant defeat would potentially deter the same adversary from challenging the US in the future. General deterrence, on the other hand, concerns how third-party states observe US crisis management. It is plausible that the use of force by the US, regardless of outcome, could enhance US deterrence of all potential adversaries, as well as its credibility among allies. In order to adjust for these two deterrence mechanisms. I will randomly assign binary treatments of specific and general deterrence successes. In specific deterrence success treatment, subjects will be informed that US use of force is likely to deter the adversary. In general deterrence success treatment, subjects will be told that the international community is likely to view a use of force as an enhancement of US credibility. Specific and general deterrence failure treatments will provide information on the contrary. A subject will receive any combination of the two deterrence treatments.

The probability of winning (p):
The probability of winning treatment randomly assigns likely victory (p = 0.7) or likely defeat (p = 0.3) to each option profile. In contrary to the prediction of bargaining theory that the losing side would prefer a bargained defeat to a costly defeat in war, I expect that, psychologically, even losing a fight might be better than losing without a fight. Moreover, the experimental manipulation of p helps me address two important risk-attitude-based alternative hypotheses. As discussed in the literature review section of this paper, the existing research often model actors as risk-seeking to explain why their behaviors deviate from the basic bargaining model (Jackson and Morelli 2007). According to this risk-attitude theory, leaders such as Napoleon or Hitler are particularly willing to make military gambles because they are risk-seeking. It is difficult to tease apart risk-attitude of leaders and CGU in observational studies. However, by experimentally manipulating the probability of winning, I can adjust for the probability-dependent risk-attitude aspect of decision-making. Combined with random assignment of treatment, risk-attitude differences and CGU can be disentangled.

Secondly, due to ingroup favoritism and outgroup derogation, as well as perceived democratic advantages, subjects may overestimate US’ chance of winning (Hewstone, Rubin, and Willis 2002; Lake 1992; Reiter and Stam 1998a, 1998b). For example, US subjects might impute that authoritarian states are inherently weak and have a low probability of prevailing in a conflict. Offer specific probabilities of winning can help block off this threat to inference.

Cost of conflict (c):
To explore the role the cost dimension of conflict decision making, the experiment randomly assigns three level of costs, no fighting, fighting with low cost, and fighting with high cost, operationalized as a number of casualties on the US side. In choice
profiles, low-cost condition specifies no casualty on US side, and high-cost condition specifies 500 US casualties. The no fighting condition also specifies no US casualty. Apart from aligning experiment design with the bargaining model, the explicit manipulation of cost also helps deal with potential outgroup derogation – namely individual might underestimate the combat effectiveness of the adversary, thus underestimating one’s own cost in fighting.

**Harm inflicted on the adversary (HARM):** This manipulation is achieved through the random assignment of adversary casualty numbers similar to the cost treatment. In option profiles, the low and high harm conditions specify 0 and 500 adversary casualties respectively.

**Paired Conjoint Experiment Design:**

<table>
<thead>
<tr>
<th>Probability of Winning (p)</th>
<th>Value of the dispute (U₁)</th>
<th>Value of the dispute (U₂)</th>
<th>Cost of Conflict (C)</th>
<th>Harm inflict on adversary (HARM)</th>
<th>Intermediate Variable: In Group-Identification (M)</th>
<th>Outcome (E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Win (p = 0.7)</td>
<td>Specific Deterrence Success (U₁ = 1)</td>
<td>General Deterrence Success (U₂ = 1)</td>
<td>No fighting no casualty (c = 0)</td>
<td>Enemy suffers no casualty (HARM = HARM₀)</td>
<td>National Identity Priming (M = 1)</td>
<td>1. Choose Option 1 vs. Option 2</td>
</tr>
<tr>
<td>Lose (p = 0.3)</td>
<td>Specific Deterrence Failure (U₁ = 0)</td>
<td>General Deterrence Failure (U₂ = 0)</td>
<td>Fighting, low casualty (c = c₁)</td>
<td>High enemy Casualty (HARM = HARMₗ)</td>
<td>No National Identity Priming (M = 0)</td>
<td>2. Numerical rating of the two options</td>
</tr>
</tbody>
</table>

In each of the eight choice tasks, subjects are asked to report a preference between two juxtaposing option profiles with two fighting options with varying levels of winning probability, deterrence utility, cost, and harm. Thus, each choice profile constitutes an observation. The dependent variable – preference for a choice profile is coded 1 if preferred and 0 otherwise. In addition, each subject is asked to make numerical ratings of the two options individually on 5-point scales, which provide more nuance evaluations of each profile. The following table provides an example of the choice tasks section. Causal quantities of interest are the AMCEs of fighting and harm attributes. If the CGU hypotheses are correct, I expect the fighting and higher harm treatment conditions to increase the probability that an option will be chosen.

Because the no fighting treatment is the lowest level of cost, it would be unrealistic to have options that would result in adversary casualty with no fighting at all. This requires a restriction to be imposed to avoid the unrealistic counterfactual (Hainmueller, Hopkins, and Yamamoto 2013, 5). It is only possible for subjects who receive c = c₁ or c = cₕ treatments to see HARMₗ. The restricted attributes combination that will not appear is shown in the table below. Given this restrictions, subjects assigned high adversary harm, will only be compared with other subjects assigned fighting condition so as not to violate the profile randomization assumption (Hainmueller, Hopkins, and Yamamoto 2013, 9).

**Restricted attribute combination**

<table>
<thead>
<tr>
<th>Cost of fighting</th>
<th>“If the US does not use military force against the adversary, no US Navy sailors will be killed in fighting”</th>
<th>Adversary Harm</th>
<th>“500 adversary’s Navy sailors will be killed in fighting”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributes</td>
<td>Option 1</td>
<td>Option 2</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Probability of Winning</td>
<td>“If force is used, there is a 30% chance for the US to win”</td>
<td>“If force is used, there is a 70% chance for the US to win”</td>
<td></td>
</tr>
<tr>
<td>Specific deterrence</td>
<td>“The use of force is unlikely to deter this adversary”</td>
<td>“The use of force is unlikely to deter this adversary”</td>
<td></td>
</tr>
<tr>
<td>General deterrence (reputation)</td>
<td>“The international community pays close attention to the crisis, use of force is likely to demonstrate US credibility in the world”</td>
<td>“The international community pays little attention to the crisis, use of force is unlikely to demonstrate US credibility in the world”</td>
<td></td>
</tr>
<tr>
<td>Cost of Fighting (Casualty)</td>
<td>“If the US uses military force against the adversary, 500 US Navy sailors will be killed in fighting”</td>
<td>“If the US does not use military force against the adversary, no US Navy sailors will be killed in fighting”</td>
<td></td>
</tr>
<tr>
<td>*By choosing this option: You enter a lottery to win 50 dollars</td>
<td>*By choosing this option: You enter a lottery to win 60 dollars</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HARM</td>
<td>“500 adversary’s Navy sailors will be killed in fighting”</td>
<td>“No adversary’s Navy sailors will be killed in fighting”</td>
<td></td>
</tr>
<tr>
<td>*By choosing this option: Your opponent enters a lottery to win 50 dollars</td>
<td>*By choosing this option: Your opponent enters a lottery to win 60 dollars</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

On a scale from 1 to 5, where 1 Indicates that the United States should absolutely not choose option 1 and 5 Indicates that the United States should definitely choose option 1. how would you rate option 1?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

On the same scale, how would you rate option 2?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>
Incorporating cash lottery incentive:

A long-standing concern about the use of survey experiments to simulate conflict decision has been the problem of stake. Purely hypothetical survey experiment typically involves no real stake, subjects are paid a fixed amount irrespective of their decision (Ding, Grewal, and Liechty 2005; Hainmueller, Hangartner, and Yamamoto 2015; Riker 1995, 31); There is neither reward nor cost for their decisions in experiment. Since experiment subjects usually do not suffer real consequences, their willingness to fight might be interpreted as unrealistic cheap talks. In order to increase the external validity of the experiment by raising the stake, I align lottery-based cash incentive to the cost conditions (J. M. Smith 1982).

When making the fighting versus not fighting decision, every subject will be informed that their choice entails a chance to choose between different lotteries: The lottery incentive is aligned with cost conditions of the two options. A choice between two options presented is also a choice between two lotteries. No fighting condition is associated with a 60-dollar lottery, the equivalent of a day’s wage in the United States, based on US national average hourly wage. Fighting with no casualty condition is tied to a 55-dollar lottery, fighting with high casualty is tied to a 50-dollar lottery. If a subject is assigned with two options with the same cost conditions, they enter the same lottery regardless of their choice. For instance, if they face two high casualty fighting options, both options are tied to a 50-dollar lottery. Thus, fighting is always costly. According to the existing bargaining model, very few (if any) individual would pick fighting to begin with. And less subject should pick fighting under high-cost condition than under low-cost condition. A similar but fictional lottery will be tied to the adversary harm conditions, the subject will be informed that the lottery payoff of their “opponent” is tied to their choices.

Additional questions and procedures:

In order to capture potential heterogeneities in treatment effects, a number of questions measuring the demographic and political characteristics are included at post-treatment portion of the questionnaire. For example, one demographic characteristic that could lead to heterogeneous treatment effect is personal connection to members of the US military. Consequently, a dichotomous response question assessing if the subject has a family member or close friend currently serving in the US armed forces is included in the post-treatment questionnaire. Other demographic and other political covariates that will be collected include gender, age, ordinal income level, education, race/ethnicity, party identity, as well as subjects’ level of interest in foreign policy, and a question assessing whether the subject values maintaining US’ global primacy. In addition, because each subject will be asked to perform 8 choice tasks. The effect of national in-group identification prime and other treatments may be reducing with every new conjoint task performed. Moreover, subjects might be less focused on conjoint profile variations due to increasing fatigue. To account for these practical issues, the order in which a conjoint task appeared will be collected.

In order to address potential inattention. An attention check question is included immediately after the naval exercise situation briefing (before an adversary challenge is presented), asking respondent to identify which two countries are holding a joint naval exercise. The second line of this attention check question instructs respondents to ignore the question and do not respond. Subjects failing this attention check will have their experiment terminated and would be excluded from the sample before any treatment is administered. A pre-treatment exclusion of
inattentive subjects is preferable to post-treatment exclusion of subjects that involves controlling for post-treatment variable, such as response time. As a tradeoff, pre-treatment exclusion would change the definition of the sample. But given that a convenient sample would be used in the pilot experiment, such tradeoff is acceptable.

Analysis:

I hope to achieve five objectives through the analysis of experiment results.

(1) The first goal is to estimate the main quantity of interest—the ACME of adversary harm treatment. I expect adversary harm to have a positive ACME. While the null hypothesis, inferred from the canonical bargaining model expect the adversary harm treatment to have no effect. In conjoint design ACME of the adversary harm manipulation is also the total effect of harm treatment (Acharya, Blackwell, and Sen 2018, 370).

(2) The second objective is to estimate eliminated ACME of harm conditional on national ingroup identification manipulation. Applying Acharya et al. framework of causal mechanism to this conjoint experiment, the eliminated ACME of HARM (ACME of Harm minus ACME conditional on national identification manipulation) provides evidence for ingroup identification as a causal mechanism which includes both the mediation effect of manipulated identity and the interaction between harm and identity manipulation. Given that the identity manipulation is imperfect, it would only show whether ingroup identification have the expected direction of effect, rather than capturing the size of indirect effect.

(3) The third objective is to test the competing risk-seeking hypothesis by estimating two quantities: (1) the ACDE of cost conditional on probability of winning manipulation, (2) the ACDE of harm conditional on probability of winning manipulation. The average controlled direct effects (ACDE) of cost and harm are essentially the ACME of cost and HARM manipulation fixing ingroup identification, which represent the direct effect of cost and harm. (3) the eliminated ACME of cost conditional on probability of winning manipulation, (4) the eliminated ACME of harm conditional on probability of winning manipulation. If risk-seeking alone provide an explanation for conflict-prone individual, we should expect the both ACDEs (1), (2), the measurement of direct effects, to be insignificant or minute while both eliminated ACME (3), (4), the indirect effects, to be large and significant, indicating that the treatment effect of fighting and adversary harm can be entirely attributed to risk-seeking. However, if the CGU theory of adversary harm is correct, we should expect to observe significant positive ACDE of Harm fixing probability of winning manipulation.

(4) The fourth objective to estimate a series of eliminated ACMEs of adversary harm for each parameter in the canonical bargaining model of war. The eliminated ACME is the difference between total effect (ACME) and average controlled direct effect (ACDE) of harm manipulation. This quantity provides insights into causal mechanisms of CGU that are related to existing bargaining model parameters. For instance, I hope to explore if high cost of conflict leads to different levels of CGU. In this example, eliminated effect is the difference between the effect of a high adversary harm (versus a low adversary harm option) under low US casualty condition and the same effect when US casualty is high. The potentially different subject sensitivities to adversary harm caused by canonical model parameter treatments could provide indications of
mechanisms, such as the adversary harm’s substitution or additive relations with the probability of winning or its potentially compensating effect for high cost when a nation suffers high casualty. Conversely, favorable winning probability might allow psychological motivations, including adversary harm to exert greater influence on decision-making.

(5) Finally, the last objective of analysis is to explore heterogenous treatment effects based on subjects’ demographic, and political characteristics. For example, neuroscience studies of schadenfreude has identified potentially more pronounced counter-empathy among male subjects (Singer et al. 2006). Thus, there might be gender heterogeneity in the treatment effect of adversary harm. Since different groups of subjects with strongly heterogenous treatment effects can result in a misleading overall ACME size (Horiuchi, Smith, and Yamamoto 2018, 199), these heterogeneity could still be important despite being observational.
Conclusion:

If we see the canonical bargaining model as a blank canvas, the development of bargain literature on war in the past two decades can be seen as a process of cumulatively extending and revising this canonical model, incorporating factors that influence the model parameter. The five main strains of the bargaining theory of war highlighted the information asymmetry, commitment problem, principal-agent problem, issue indivisibility, and costly peace as sources of bargaining failures and causes of war. However, all five existing strains of the rationalist literature dismissed the possibility that individuals can derive utility from conflict. Due to the assumption that conflicts are inherently costly, the benefit side of the utility function was left unrevised. Thus, there is a demand for a conflict-generated-utility theory. Moreover, recent advances in behavioral economics and experimental social psychology has provided new theoretical and methodological tools to meet this demand.

Drawing on these recent developments as well as the existing bargaining model of war, I proposed a theory of conflict-generated-utility to incorporate the psychological payoff of conflict into individual incentive structure. I also proposed a stylized theory on how individual-level CGU can aggregate into state-level policy under personalist and non-personalist regime types. Lastly, treating the canonical bargaining model as the null hypotheses, the paper has outlined two conjoint experiment designs based on a revised bargaining model to tests the CGU hypotheses empirically. The experiment design proposed ways of experimentally manipulating all bargaining model parameters, adversary harm as well as national ingroup identification moderator. The experiment also suggested cash lottery procedures that align subject’s monetary payoff with hypothetical costs of conflict in the experiment.

Finally, it is worth noting the potential implications of a CGU theory that motivates the present inquiry. Numerous research questions in the social sciences are efforts to explain seemingly inefficient behaviors of rational actors. If the proposed experiment design confirms the CGU hypotheses, the result could have broad implication for research beyond international crisis bargaining. The same CGU model can be applied to the study of civil war, violent ethnic conflict and terrorism, in which CGU motivates ostensibly irrational behavior such as genocidal violence (Kaufman 2006). Given that CGU can only be obtained through fighting, this type of utility simultaneously makes an actor more resolved and bargained settlement less likely, which is a curious dynamic for game theoretical modelling of conflict. The CGU model may also apply to polarized partisan politics in democracies. Instead of maximizing vote shares or representing the material interests of their constituencies, deriving enjoyment from antagonizing and undermining the opposing party might be an important motivation behind policy and electoral agendas of politicians and candidates (E. R. Gerber and Morton 1998; Layman, Carsey, and Horowitz 2006). As a result, politicians are rewarded by their political ingroup for polarizing positions, leading to suboptimal political outcomes. Therefore, the finding of this proof-of-concept study can be extended by game theoretical modeling of strategic interaction under uncertainty, as well as empirical testing using observational data in a variety of subject areas involving individual actors in conflict situations.
Bibliography:


Schultz, Kenneth A. 1999. 53 Source: *International Organization* *Do Democratic Institutions Constrain or Inform? Contrasting Two Institutional Perspectives on Democracy*.


