

External Interventions: Bluffing and Adaptive Learning in Civil Wars

Seden Akcinaroglu, Sebnem Bora & Elizabeth Radziszewski

To cite this article: Seden Akcinaroglu, Sebnem Bora & Elizabeth Radziszewski (18 Oct 2024): External Interventions: Bluffing and Adaptive Learning in Civil Wars, *Civil Wars*, DOI: [10.1080/13698249.2024.2358287](https://doi.org/10.1080/13698249.2024.2358287)

To link to this article: <https://doi.org/10.1080/13698249.2024.2358287>



Published online: 18 Oct 2024.



Submit your article to this journal [↗](#)



Article views: 130




View related articles [↗](#)



View Crossmark data [↗](#)



External Interventions: Bluffing and Adaptive Learning in Civil Wars

Seden Akcinaroglu ^{a*}, Sebnem Bora^{b*} and Elizabeth Radziszewski^{c*}

^aPolitical Science, Binghamton University, Binghamton, USA; ^bComputer Engineering Department, Ege University, Kazimdirik, Turkey; ^cSTART, Maryland University, College Park, USA

ABSTRACT


This paper examines the effectiveness of bluffing about external intervention in shaping civil war dynamics using an agent-based model. It argues that warring parties' initial information levels about interveners and their learning processes determine the success of bluffing. The model explores three scenarios: 1) accurate initial expectations and informative signals 2) inaccurate initial information and uninformative signals and 3) heterogeneous information sets and varied signal updates. Findings suggest bluffing is most effective when both sides start with inaccurate expectations and update slowly. The model's implications are analyzed through cases in El Salvador, the Warsaw Uprising and South Vietnam negotiations.

ARTICLE HISTORY Received 27 September 2022; Accepted 18 May 2024

Introduction

Approaching Warsaw in 1944, the Soviets assured Polish insurgents of impending assistance as the Home Army fighters prepared for a battle to free their capital from German occupation. Yet the promise so vital to the success of the insurgency never materialised. Convinced, almost until the final stages of the fight, that Soviets would join the Poles in their struggle, the Germans braced for a long battle against an outnumbered but potentially troublesome opponent. Why did the Soviets bluff about their intervention and to what extent did such a strategy prove effective in fulfilling their goals? More broadly, when do third parties bluff about intervention, and how do warring groups' information levels and ability to learn about third-party intentions affect the success of the strategy?

Third parties, including external interveners, make their mark visible in civil conflicts either by facilitating resolutions or deliberately prolonging the fighting (Balch-Lindsay and Enterline 2000, Regan 2002). This paper investigates conditions under which strategic bluffing about intervention can work. Whether they

CONTACT Seden Akcinaroglu  sakcinar@binghamton.edu

*The authors have contributed equally to this manuscript.

© 2024 Informa UK Limited, trading as Taylor & Francis Group

aim to support the rebel group or seek conflict expansion, third parties may manipulate information levels of the warring actors to affect the combatants' preferences over strategies in civil wars. Depending on the fighting groups' initial information levels and ability to correctly learn about third-party goals in the conflict, bluffing may have a varying success rate. By employing an agent-based model that allows us to examine the convergence of beliefs about third-party intentions, we can determine when bluffing is likely to be useful.

Previous works on strategic third parties and civil wars have examined the incentives of external actors to affect the terms of negotiations in conflict by exploiting the combatants' private information about preferences on settlement (Kydd 2006, Rauchaus 2006). Yet while previous research provides insights about motivations to manipulate information in mediation, current studies offer limited ideas about the precise conditions under which specific strategies may work to the advantage of an external actor. The strategic motivations of the third party can then exacerbate informational problems of warring parties and alter war dynamics. The paper advances existing scholarship by showing that the potential for deception is largely dependent on the warring groups' ability to learn and uncover the strategy.

The development of a novel, learning-based approach to show conditions under which third-party bluffing about intervention will be most effective as a strategy in civil wars is the main contribution of the paper. We employ an agent-based model to capture the uncertainty arising from third party intentions, and we capture the complexities of civil war dynamics by showing how varying information levels of combatants and their ability to update expectations of intervention affect the utility of bluffing and yield different war outcomes. Specifically, we examine the success of bluffing when a) warring groups possess the same information about third-party intentions and update with informative signals, b) both have the same information but update with uninformative signals and c) only one of the parties has superior information and the actors update using different types of signals. In doing so we show, for example, that a strategic actor hoping to prolong the war can do so by manipulating warring groups' preferences over outcomes if both possess the same information about the third party and update erroneously. Thus, by discovering the conditions for successful bluffing, we can determine when a third party would be able to shape conflict dynamics in a substantial way.

The paper proceeds in the following manner. First, we present a brief review of the current literature on third-party bluffing. The theoretical argument on the strategic use of third-party bluffing follows, aided by simulations to show how and when warring parties discern the bluff. The theory is demonstrated using three distinct cases: the El Salvador Civil War, the 1944 Warsaw Uprising and the 1973 secret negotiations with South Vietnam. The paper concludes with policy recommendations for conflict resolution.

Third-Party Bluffing and Civil Wars: Why Focus on Strategic Action?

Bluffing, characterised by deceptive communication with the intent to deceive, is primarily linked to James Fearon's (1995) explanation for negotiation breakdown and the onset of interstate war. It entails leaders misrepresenting preferences or private information strategically for optimal outcomes. While research has largely focused on bluffing between two warring states in international disputes, less attention has been given to the role of third-party bluffing in civil wars, despite evidence suggesting their strategic manipulation of warring groups for desired outcomes (Sciolino 1994).

Understanding the role of third-party bluffing or manipulation of private information in civil wars is particularly important given greater information asymmetries may exist between the combatants in intra rather than inter-state conflicts. While leaders of states are often uncertain about each other's capabilities, they are familiar with their own fighting capacity including the size of the army, weapons possession and even levels of resolve. In contrast to state leaders, rebel commanders experience greater difficulty trying to secure committed recruits and funds for fighting. The constant uncertainty related to fighter recruitment and resource acquisition contributes to greater information gaps in civil wars (Walter 2006, Mattes and Savun 2010). Not surprisingly such gaps create ample opportunities for third parties to step in and exploit information in a way that could affect warring parties' decisions to settle or continue the fight. Expectations of intervention alter the onset, duration, and settlement of war and yet we know very little about how third parties might be using this strategic tool to advance their interests (Akcinaroglu and Radziszewski 2005, Thyne 2006, Jackson *et al.* 2020).

Research on third-party bluffing in civil wars predominantly examines biased mediators aiming to manipulate the private information of combatants to end the conflict and influence the outcome (Kydd 2006, Rauchaus 2006). Failed peace attempts often result from the third party shaping the agreement terms to serve its own interests. However, studies rarely explore the potential for third parties to strategically pursue alternative outcomes, such as prolonging the conflict, and manipulating information asymmetries regarding their intentions. Our research contributes to the understanding of third-party bluffing by examining the dynamic relationship between external interveners and warring parties. We explore strategic motivations for intervention and investigate how rebels and the government learn and update information about the third party's role in the conflict. By specifically examining when and how combatants uncover the third party's private information about its intentions, we can identify conditions under which bluffing may achieve outcomes desired by external actors. Using an agent-based model,

we not only uncover the process of learning but also predict the speed of learning, shedding light on the limited window of opportunity where successful bluffing occurs, albeit for a short duration.

Computational models are widely used in the social sciences to capture the complexity of agents in dynamic settings like war and conflict (Findley and Young 2007, Bennett 2008, Pecheninka and Bennett 2017). Agent-based models, which can treat warring parties as boundedly rational and capable of adaptive learning, offer insights into how beliefs and preferences evolve throughout the course of war. In the context of civil wars, where access to funding is critical in rebels' decision-making, preferences regarding settlement can change over time based on perceptions of third-party involvement. However, the use of dynamic approaches that incorporate evolving preferences and new information has been limited, with Zott (2002) being an exception in studying bargaining failures between two conflicting parties. Consequently, our understanding of the conditions under which bluffing becomes an effective strategy, particularly when employed by external interveners in civil wars, remains limited. This study aims to investigate the circumstances in which third parties find bluffing advantageous, highlighting the significance of how quickly rebels and the government uncover the true intentions of the intervener. By employing an agent-based model, we can capture the intricacies of war dynamics, including the evolution of beliefs and learning speed, and assess the model's effectiveness by adjusting agent characteristics and replicating real-life scenarios.

Third-Party Bluffing and Its Effectiveness as a Strategy

Whether they intervene to settle the conflict (Regan 2002) or deliberately prolong it (Balch-Lindsay and Enterline 2000), third parties have an interest in minimising intervention costs. They can employ bluffing and strategic misrepresentation of private information to achieve desired policies with minimal effort. For instance, by convincing warring parties of substantial aid to one side, they can incentivise the other party to settle early rather than endure heavy fighting costs. Alternatively, a third party with a vested interest in destabilising the central government may manipulate information about its involvement in the conflict to deliberately prolong the war (Akcinaroglu and Radziszewski 2005).

Given its cost-efficiency in achieving desired outcomes, bluffing emerges as a strategic tool for interveners, especially in civil wars where information gaps abound. Asymmetric information, a recognised cause of failed bargaining in conflicts (Fearon 1995), leads to misjudgements about power and costs, resulting in unsuccessful negotiations (Slantchev 2003, Smith and Stam 2004, Powell 2004). War becomes a means to acquire essential information, with

parties inclined to settle once they understand each other's positions. Our focus extends to informational asymmetries, particularly the misinterpretation of third-party aid, complicating settlement processes. Third parties, driven by their strategic goals, may distort information about their support, challenging warring groups to accurately assess collective capabilities. As uncertainty grows, bluffing's effectiveness as a strategy increases, allowing third parties to exploit informational gaps to their advantage. This manipulation hinges on the perceived credibility of the bluff, influencing parties' settlement preferences. Our model emphasises the role of information asymmetry, particularly regarding third-party intentions, as a central factor in conflict dynamics and resolution efforts. We contend that the believability of a third party's bluff – defined as the extent to which combatants find the third party's deceptive claims or threats plausible – hinges on two key factors: their initial information sets about the third party's intentions, and the pace at which they can uncover the actual strategy. When the combatants have access to information on critical factors that affect external actors' decisions about interventions, they are less likely to fall prey to deception. If warring groups, however, update information inaccurately about potential intervention, third parties will be more successful in sustaining the bluff and achieving their desired outcome.

Determination of Information Sets

We argue that warring parties possess initial information sets to gauge the probability of third-party intervention and make decisions accordingly. However, these information sets are endogenous. Our focus is on two aspects: the transparency of signals and the bias of warring parties in perceiving and evaluating these signals. The former pertains to signal transmission quality, while the latter concerns signal reception quality.

Transparency of Signals

The transparency of signals, crucial for enabling combatants to accurately update their information on third-party intentions, is influenced by various factors related to domestic context and leadership. Democratic governments, for example, tend to offer higher levels of information access, thereby reducing opportunities for deception and enhancing signal transparency (Schultz 2001). This is further supported by the presence of democratic institutions that limit the potential for misleading information. Executive and political constraints also play a significant role in the information flow (Thyne 2012). In scenarios where decision-makers are restricted by polarised coalition governments or institutional checks that limit executive power, such as in divided governments, their capacity

to fulfil promises is diminished. Additionally, in democracies, foreign policy is often a reflection of domestic preferences, with political costs like protests against intervention influencing policy decisions. These domestic pressures can serve as credible indicators of a decision-maker's constraints. The concept of 'responsible' leaders, or those who have initiated interventions, adds another layer to this dynamic. Such leaders may be reluctant to concede defeat or change policies due to the risk of domestic or elite repercussions, indicating that a change in policy by a third-party leader without this 'responsible' status may carry lower costs (Prorok 2016). Lastly, the principle of convergence dictates that time reveals information and helps converge expectations (Filson and Werner 2002). To understand this, we examine the history of interventions, which can shed light on past interactions, the reputation of the leader, and the third-party state. Shorter alliances tend to present larger informational gaps and create uncertainties about the interests and intentions of the third party.

Biases and Perception of Signals by the Warring Actors

Actors frequently employ cognitive biases and heuristics, which assist in filtering information to manage uncertainty during a political crisis. In turn, these biases influence the warring actors' perception and evaluation of signals coming from third parties. The role of the ally in bringing the leader to power, access to secret information, and the responsibility of the leader are all factors that can affect cognitive beliefs. The alliance relationship builds and perpetuates leaders' beliefs. If the ally is responsible for the leader's initial access to power, this creates an anchoring heuristic whereby the leader favours the initial positive assessment of the third-party state which he associates with the desired emotional state of 'being secure in power'. Thus, he may be more inclined to perceive the ally as reliable even when the signals are transparent and highlight the extent of constraints on third-party intervention. Similarly, access to secret negotiations/meetings with the ally may make the leader believe they have access to information (availability heuristic) that is not accessible to others. In this case, the leader may overvalue the information received from the ally due to the availability, proximity and exclusivity of this information. Lastly, leaders responsible for the initiation of war may want to avoid discordant information that shows evidence of policy change or failure and instead hold onto the belief that the status quo can and will continue. All of these are consistent with the concept of cognitive dissonance and confirmation bias, where individuals seek out information that confirms their pre-existing beliefs and discount information that contradicts them.

The Agent Based Model

This section describes an agent-based model and its simulation results, focusing on conditions for successful bluffing. The model involves a third-party bluffing about external support, with the government and rebels, who possess imperfect information, as agents. The objective is to identify conditions for successful bluffing based on the type and speed of information update among the warring parties. An external actor, potentially backing a faction, is interested in the conflict's outcome. The government and rebels are in contention over issue K, and the conflict persists until either a faction is defeated, or successful negotiations transpire.

Model Parameters and Variables

Capabilities of the government and the rebel are denoted by M_g and M_r , respectively. We assume that both warring parties form expectations, shaped by initial information sets, that a strategic third party will intervene on behalf of the government (a new or continued intervention). Government and the rebels' expectations about third-party intervention are represented by $expected_g$ and $expected_r$, respectively. The true state of intervention by the external intervener is depicted by $actual$. For example, if the third party has no intention of intervening, the value of $actual$ will be 0. Unless the actual intentions of the third party are perfectly forecasted by the government and the rebels, there is room for bluffing and the third party may benefit from creating informational asymmetry. We let $expected_g$, $expected_r$, and $actual$ to be in the range of [0,1] to reflect probabilities of intervention. We sequentially denote these as xx , yy , and zz , representing the values for $actual$ third-party intervention. The variables in $expected_g$ and $expected_r$, namely xx_g , yy_g , zz_g , xx_r , yy_r , and zz_r , are influenced by the initial information set of each warring party. This information set is determined by the transparency of signals (t) and the warring party's bias in interpreting them (b). To the extent that signals are transmitted poorly and received with a notable bias, expectations will deviate from $actual$. We list the parameters and variables of the model and their definitions in [Table 1](#). The initial values of xx , yy , zz drawn uniformly¹ in predefined intervals, normalised and added to yield probabilities of intervention ($expected_g$, $expected_r$, and $actual$) that fall in the range of [0,1]. The below equations show the composition of $actual$ and $expected$ intervention.

$$actual = xx + yy + zz \quad (1)$$

$$expected_g = (txx_g + bxx_g) + (tyy_g + byy_g) + (tzz_g + bzz_g) \quad (2)$$

Table 1. Parameters and variables of the model.

Model Parameters	Definition	Scope	Range	Default Value
M_g	Capability of the government	Individual/Government	0.0 to 1.0	0.7
M_r	Capability of the rebel group	Individual/Rebels	0.0 to 1.0	0.3
actual	Decision of the third-party to intervene	Individual/Third-party	0.0 to 1.0	0.0 to 0.5
$expected_g$	Government's expectation of third-party intervention	Individual/Government	0.0 to 1.0	0.0 to 1.0
$expected_r$	Rebels' expectation of third-party intervention	Individual/Rebels	0.0 to 1.0	0.0 to 1.0
xx	First category (economics) that represents actual third-party incentive to intervene	Individual/Third-party	0 to 200	Uniformly drawn in [0,40]
yy	Second category (politics/military) that represents actual third-party incentive to intervene	Individual/Third-party	0 to 200	Uniformly drawn in [0,40]
zz	Third category (social) that represents actual third-party incentive to intervene	Individual/Third-party	0 to 200	Uniformly drawn in [0,20]
$xx_{g,r}$	First variable that represents government/rebels' perception of third-party constraint	Individual/Government/Rebels		
$yy_{g,r}$	Second variable that represents government/rebels' perception of third-party constraint	Individual/Government/Rebels		
$zz_{g,r}$	Third variable that represents government/rebels' perception of third-party constraint	Individual/Government/Rebels		
$txx_{g,r}$	Transparency of emitted signals for the government/rebels in assessing the first category of third-party intervention	Individual/Government/Rebels	0 to 200	Uniformly drawn in [0,60]
$tyy_{g,r}$	Transparency of emitted signals for the government/rebels in assessing the second category of third-party intervention	Individual/Government/Rebels	0 to 200	Uniformly drawn in [0,60]
$tzz_{g,r}$	Transparency of emitted signals for the government/rebels in assessing the third category of third-party intervention	Individual/Government/Rebels	0 to 200	Uniformly drawn in [0,20]

(Continued)

Table 1. (Continued).

Model Parameters	Definition	Scope	Range	Default Value
$bx_{g,r}$	Government/rebel bias in assessing the first category of third-party intervention	Individual/Government/ Rebels	0 to 200	Uniformly drawn in [0,20]
$by_{g,r}$	Government/rebel bias in assessing the second category of third-party intervention	Individual/Government/ Rebels	0 to 200	Uniformly drawn in [0,20]
$bz_{g,r}$	Government/rebel bias in assessing the third category of third-party intervention	Individual/Government/ Rebels	0 to 200	Uniformly drawn in [0,20]
cycle	The number of simulation ticks. It increases by 1 at each new simulation tick.	Integer	0 to 20,000	0

$$expected_r = (txx_r + bxx_r) + (tyy_r + byy_r) + (tzz_r + bzz_r) \quad (3)$$

The first equation details the actual intervention, which is determined based on three broad dimensions: Social, Political/Military and Economic, as outlined in Appendix Section 1, [Table 1](#).

Initially, 'cycle', representing the count of simulation ticks, is set to 0. Then, the simulation starts, and 'cycle' is increased by 1 at each new cycle of the simulation. We then compute $expectation_g$ and $expectation_r$ and we assign new values to these variables drawn from the learning intervals to mimic new signals. Thus, with each new cycle, the variables are continuously updated.

Informational Asymmetries at the Bargaining Stage

Assuming that incomplete information about relative capabilities can cause bargaining failures and subsequently, war, we define the beliefs of the government (p_g) and rebels (p_r) about this ratio.

$$p_g = M_g / (M_g + M_r) \quad (4)$$

The equation represents the relative capabilities of each party, where the government aims to obtain at least p_g percent of K , while the rebels aim for $(1-p_g)$ percent. However, this equation only considers the parties' capabilities without factoring in third-party aid. Since third-party aid enhances the capabilities and resolve of each party, altering their assessment of power distribution, we modify Equation (4) as follows:

$$p_g = [M_g * (1 + expected_g * 0.1) + expected_g] / [M_g * (1 + expected_g * 0.1) + expected_g + M_r] \quad (5)$$

Equation (5) incorporates expected third-party aid ($expected_g$) to augment the government's capability, reflecting its potential boost from troops or funding. Here, the term $expected_g * 0.1$ highlights that even the mere expectation of third-party support can enhance the rebels' resolve, leading to increased determination in their fight (Akcinaroglu and Radziszewski 2005). In this context, the initial relative capabilities (as defined in equation 4) that determine each party's bargaining power are altered in favour of the armed actor expecting external support. With the anticipation of aid, when calculating the relative bargaining power, both parties will first consider the expected material and military assistance from a third party's ally, $expected_g$ and second, small increases in resolve, $expected_g * 0.1$ that this support will contribute to the recipient's efforts in the conflict.

$$p_r = (M_r) / [M_g * (1 + expected_r * 0.1) + expected_r + M_r] \quad (6)$$

In this model, assuming third-party assistance to the government, Equation (5) includes the term $M_g * (1 + expected_r * 0.1) + expected_r$ in the denominator.

We replace $expected_g$ in (5) with $expected_r$ in (6) to represent the rebels' belief in relative power based on their own expectations of third-party aid. The acceptance or rejection of bargaining stage offers depends on the beliefs of both sides (equations (5) and (6)). If the absolute difference between the government's maximum concession $(1-p_g)K$ and the rebels' minimum acceptance p_rK exceeds 0.006, negotiations collapse, and the conflict persists.²

$$|(1.0 - p_g) - P_r| < 0.006 \quad (7)$$

The other condition to achieve a settlement is as follows:

$$actual = expected_g = expected_r \quad (8)$$

Notice that, (7) can be satisfied without (8), this is the situation where both parties have similar beliefs on the distribution of power, that is on third party aid, yet these beliefs do not reflect the true state of the world. For example, both may have positive beliefs that the third party will aid the government (e.g., $expected_g=expected_r=0.8$) but such beliefs may exceed the actual probability of intervention (e.g., $actual=0.2$). In this case, negotiation occurs with false beliefs on third party intervention, that is, without convergence of beliefs on the actual value of intervention. In that case, the third party can achieve its objectives, such as increasing an ally's bargaining power and stake in issue T, without intervention costs. This is termed *settlement by bluff*. Alternatively, a *sincere settlement* occurs when both warring parties understand the third party's true intentions and accurately assess the power distribution, enabling a resolution. The simulation concludes once a settlement is reached, either by bluff or sincerely. If not, the war persists, leading to the next simulation cycle.

Warring and Adaptive Learning

When negotiations fail, the two parties return to war which provides them with the means to learn adaptively on third party interventions. Every time negotiations fail the two sides have the means to gather more information on the intentions of the third party. This means that third parties can not fool forever. We start the simulation with a random draw for expectations on third party aid, $expected_g$ and $expected_r$. Once negotiations fail due to the satisfaction of (7) or (8), the parties have the opportunity to receive new signals on third party intentions. This is depicted by drawing new values for $expected_g$ and $expected_r$, where the learning intervals narrow with each cycle. We define the bounds of learning intervals for the components of $expected_g$ and $expected_r$ as $[(x-(A/cycle)), (x+(A/cycle))]$, $[(y-(B/cycle)), (y+(B/cycle))]$, and $[(z-(C/cycle)), (z+(C/cycle))]$ where A, B, and C refer to predefined values that are chosen to depict the quality of the information sets of the warring parties.

For example, if the actual value of $x = 20$ and $A = 40$ the government will get a new signal at cycle one from $[20 - 40, 20 + 40] = [0, 60]$ given that the value of each variable is bounded from below by zero, in cycle two however the new signal will be drawn from $[20 - 20, 20 + 20] = [0, 40]$ as $A/2 = 20$. Thus, with each cycle, signals become more informative as interval bands narrow.

- Two types of information sets, correct and wrong, are considered to define learning intervals. Parties with correct information sets lack complete insight into third-party actions but have relatively accurate interpretation about third-party intentions. These cases present narrow learning intervals, allowing informative signal-based information updates. Conversely, parties with wrong information sets hold beliefs that significantly deviate from reality, resulting in wide learning intervals and largely uninformative signals. With time, these signals will improve. The rules for updating variable values are as follows:
- The learning interval is updated with each cycle, narrowing according to the given rule. Rapid narrowing occurs when actors possess correct initial information sets.
- If a variable falls within its learning interval in a cycle, its value remains unchanged.³ For example, if the initial draw for x_g by the government is fifty and the new learning interval for cycle one is $[0, 60]$, if fifty falls within that range, the government will not receive a new signal for x_g in that round.
- If a variable's value falls outside its learning interval during a given cycle, the actors receive a new signal. For instance, consider the initial draw for x_g in cycle one as seventy, which is outside the learning interval of $[0, 60]$. In this case, a new draw is made, let's say yielding a value of forty. As changing beliefs is challenging, the two parties are not allowed to discard their old information entirely. Instead, the old information $x_g = 70$ is partially retained along with the new information $x_g = 40$, and the parties compute the average of the two signals. Therefore, the government's updated value for x_g would be $(70 + 40)/2 = 55$.
- After computing new variable values, $expected_g$ and $expected_r$, p_g and p_r are automatically updated at each cycle.

Simulation Details

We implement our agent-based model (Repast 3.1) to look at three distinct scenarios based on who initially possesses correct or wrong information sets on third party intentions. We observe how varying information sets of warring parties and their ability to update expectations on third party intervention affect the utility of bluffing for the third party. In the first scenario, we simulate the model when both warring parties have the correct initial

information sets about third-party intentions, and they receive good and informative signals for updating (right learning rule necessitates narrower learning bands). In the second scenario we simulate the model where both parties have incorrect initial information sets and receive uninformative signals for updating (wrong learning rule necessitates broader learning bands). In the third scenario, we simulate the model where only one of the warring parties has the correct information while the other one does not.

Homogeneous Information Sets with the Right Learning Rule

In this scenario, we study the effectiveness of bluffing in conflicts where both sides have homogenous information sets and can update correctly. We assume the probability for third-party intervention is low (*actual*= low), with both the government and rebels mostly identifying the factors influencing this decision accurately (*expected_g* and *expected_r*=*low*). In this scenario, the variables *xx* and *yy* of the actual values are uniformly drawn from the range [0, 40], and the variable *zz*, is uniformly drawn from the range [0, 20]. After normalisation and addition, the final values fall within the range [0, 0.5], indicating a probability of less than half for third-party intervention. Similarly, initial values in *expected_g* and *expected_r* are drawn from the same ranges, with *txx_{g,r}*, *tyy_{g,r}* being drawn from U[0, 40], and *tzz_{g,r}* drawn from U[0, 20]. To add to the accuracy of the information sets, we introduce a minimum bias by drawing *bxx_{g,r}*, *byy_{g,r}* and *bzz_{g,r}* from the narrow range of U[0, 5].

Reflecting the common reality of rebels being weaker than governments in civil wars, we initially set *M_g* at 0.7 and *M_r* at 0.3. For informative future signals, we set A, B, and C values at forty in the learning intervals, aligning initial expectations with the correct information set. Figure 1 displays the simulation results.

The simulation graph illustrates convergence (condition (8)) occurring around forty cycles.⁴ This highlights that with accurate initial information sets and good signals, both the rebels and the government can promptly update information about external intervention, leading to swift convergence towards complete information. Such quick recognition can undermine the effectiveness of a third-party strategy, making bluffing unproductive. Parties then arrive at a genuine settlement mirroring the real power distribution.

Homogeneous Information with the Wrong Learning Rule

In this situation, we observe the success of bluffing when both sides have homogeneous but incorrect, learning rules about intervention. Here, both parties begin with false expectations about intervention and update their

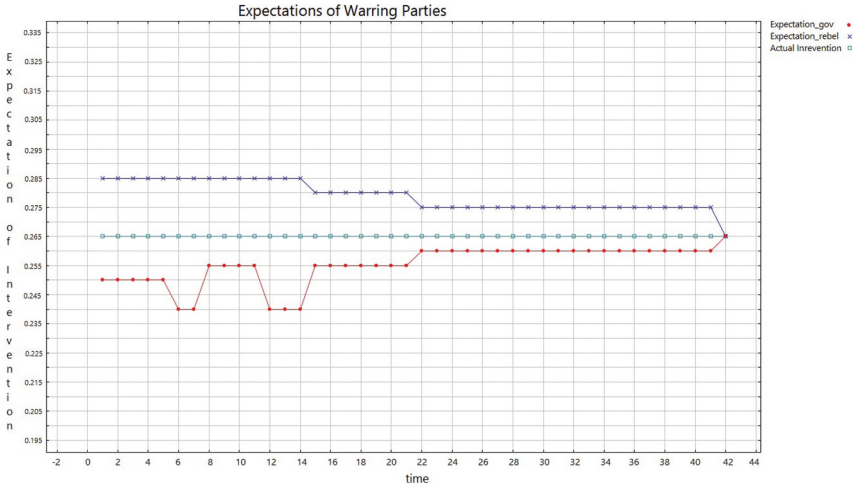


Figure 1. Updating when parties have Homogenous Information Sets with the Right Learning Rule.

information inaccurately. We maintain the assumption of low actual intervention intent from the third party, and both government and rebels misjudge the determining variables for third-party intervention. Consequently, both anticipate third-party intervention in this context.

In this intervention scenario, the initial values for the variables xx and yy are drawn from $U[0,40]$, and zz from $U[0,20]$, resulting in a low intervention probability. The initial expected values, $txx_{g,r}$ and $tty_{g,r}$ are drawn outside the actual values, from $U[50, 60]$ while the values of $tzz_{g,r}$ are drawn from $U[10, 20]$. A notable bias is incorporated for $bxx_{g,r}$, $byy_{g,r}$, $bzz_{g,r}$ by drawing each from $U[10, 20]$. This, coupled with transparency values, t , outside the actual range, signifies both parties' initial misinformation levels. Overall, expectations of intervention ($expected_g$, $expected_r$) are overly optimistic, ranging from 0.7 to 1.0.

We continue to assume that the values of M_g and M_r are 0.7 and 0.3, respectively. A, B, C in the learning intervals are given the large values of 200, 200, 100 respectively to reflect the uninformative nature of the new signals. In this case, we also set the learning intervals for the three variables to be divided by $cycle * 0.4$ instead of $cycle$ to reflect the difficulty of learning once parties start with bad information sets. Multiplying the cycle with 0.4 slows the pace of learning, that is, the rate at which learning intervals narrow. The simulation results are shown in Figure 2.

As depicted in Figure 2, convergence in this scenario is slow, requiring around 500 simulation cycles for both parties' expectations to align with the actual intervention values, providing sufficient time for the third party to

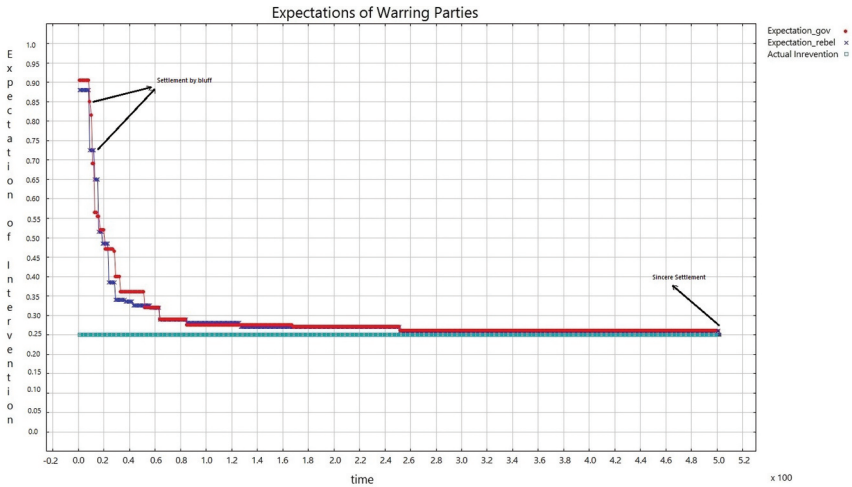


Figure 2. Updating when Parties have Homogeneous Information Sets with the Wrong Learning Rule. We expanded the range of the x axis such that $\text{time} = x * 10^2$, this means that 3.0 on the x scale is actually 300 cycles.

meet its goals. For instance, a third party could bolster its ally's negotiating power by inducing a settlement based on inflated perceptions of third-party assistance. These elevated expectations can enhance the ally's relative power and control of K without significant costs. Figure 2 shows opportunities for both parties to reach a settlement by bluff under shared, but incorrect beliefs about third-party aid within zero to five hundred cycles. Given the slow updating process in this scenario, the third party can achieve its objectives well before beliefs converge to actual values.

Heterogeneous Information

In this scenario, bluffing proves successful when the warring parties have heterogeneous information sets, with one party holding more insight into third-party intentions. We assume low intervention intent from the third party, with the rebels misjudging and the government correctly identifying intervention-determining variables. The draw parameters for *actual* remain unchanged: xx and yy are from $U[0,40]$, while zz is from $U[0,20]$. In this scenario, for $expected_g$, initial values of txx_g and tty_g too, are drawn from $U[0,40]$, and tzz_g from $U[0,20]$. For $expected_r$, initial values of txx_r and tty_r are drawn from $U[50, 60]$, and czz_r from $U[10, 20]$. To create minimum government bias, we draw bxx_g , byy_g , bzz_g from $U[0,5]$ while for the rebels, we create a notable bias by drawing bxx_r , byy_r , bzz_r from $U[10,20]$. After normalisation and addition of the three variables, the value of *actual* and $expected_g$ fall in

the range of $[0,0.525]$ indicating a low probability of intervention, while the value of $expected_r$ lies in $[0.7,1.0]$. Again, M_g and M_r are set to 0.7 and 0.3, respectively. For the rebels, we retain narrower learning intervals of A, B, C as 40 while for the government the numbers are 200, 200, 100, respectively to reflect uninformative signals⁵. The simulation results are shown in Figure 3.

In a scenario with heterogeneous information sets, the convergence rate falls between those of the first and second scenarios, requiring approximately 400 simulation cycles for parties' expectations to align with actual third-party intervention values. Learning happens quicker than when both parties are uninformed, potentially making the strategy viable. However, it's slower than when both groups have accurate information, implying a brief window of opportunity for the third party. Given the significant information disparity between the parties, a settlement is more likely when condition (8) rather than just (7) is met. The government, starting with superior information, quickly discerns third-party intentions, while the rebels continue learning from negotiation failures and third-party actions. This makes a bluff settlement harder to achieve, with a sincere one more probable if a settlement occurs at all. However, the third party may still weaken the opposition, in this case, the rebels, by reinforcing its ally's resolve to continue fighting.

In the following section, we offer three cases to highlight the varying degrees of success in third-party bluffing during war. These cases aren't for rigorous empirical testing, but rather to illustrate our model's processes, and serve as an initial validation of the model's empirical predictions.

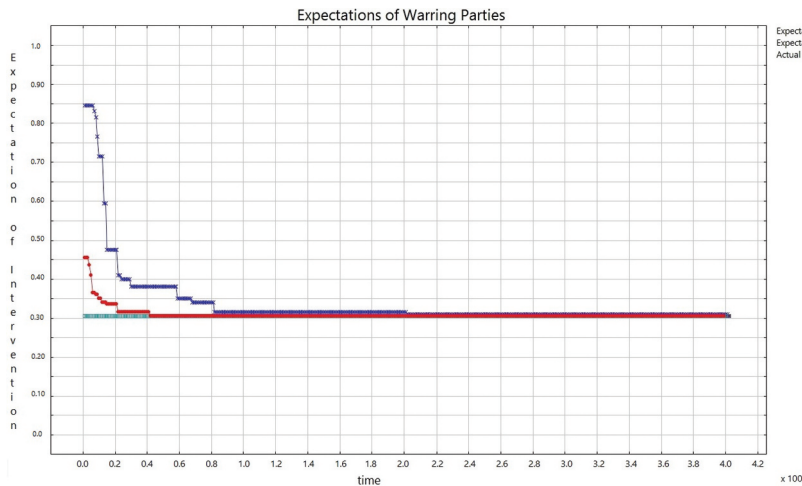


Figure 3. Updating when Parties have Heterogeneous Information Sets. We expanded the range of the x axis such that $\text{time} = x \times 10^2$, this means that 3.0 on the x scale is actually 300 cycles.

Case: Homogenous Information and Correct Learning About Intervention – Civil War in El Salvador, 1980–1992

Our prediction suggests that third-party bluffing is the least effective strategy when warring groups possess correct information (Table 2) about external actor's intentions and the right learning rule. In order to show that the strategy failed, we need to demonstrate the existence of bluffing, its quick discovery by both parties and the tactic's ineffectiveness in altering conflict dynamics as envisioned by an external intervener.

The sustained conflict in El Salvador can be attributed to the substantial military aid provided by the United States to the country since 1980. Beginning with President Carter, the U.S. contributed nearly six billion dollars to bolster the government's forces, which would likely have collapsed within six months without U.S. assistance (Federal News Service 1988). Therefore, the cessation of U.S. aid played a crucial role in the victory of FMLN, a Marxist movement committed to transforming the political and social landscape of El Salvador. Supported by Cuba, Nicaragua and to a lesser extent Honduras, all opponents of the central government, the FMLN possessed tactics and weapons to wage their war (Montgomery 1995). Initially strong, aid dwindled after the Cold War, impacting FMLN's fighting momentum. Despite their determination, the group shifted its strategy from direct offence to resistance (Prisk 1991). The declining aid from rival governments and ongoing U.S. support of the central government thus played a crucial role in shaping the rebels' strategy in the late 1980s and 1990s.

At the same time, U.S. support for the government began to change. With the slaying of Jesuit priests by the army and poor human rights record, domestic debate about aid to the increasingly abusive government began to divide the country. Protests erupted in several U.S. cities, prompting Congress to deliberate on the future of intervention. With the strong support

Table 2. Information Sets in El Salvador.

Third-Party State		Warring Actors		
Transparency of Signals	USA	Signal Bias	FMLN	Government
Domestic Costs	Yes, Protests	Political and Executive Constraints	No, Centralized Command	Limited ⁹
Political and Executive Constraints	Yes, Divided Government			
Transparency	Yes, Democracy	Leader Affiliation with the 3 rd party	No	No
History of Intervention	Long >10	Secret Negotiations	No	No
Responsible Leader	No, Bush	Responsible Leader	No, Ceren	No, Christiani
Transparency of the Signals?	Yes	Bias in Perceptions?	No	Limited

Outcome: The clear and transparent signals regarding U.S.'s constraints on intervention, combined with the lack of evident cognitive bias, enabled both parties to form accurate initial information sets.

of the Democrats, Congress at last cut U.S. support to El Salvador by half, with a possibility of renewing full assistance (Drink 1990). In declaring that full aid could be restored if rebels failed to negotiate, President Bush assured the government of El Salvador that their expectations of assistance should continue, a step that would be critical in bringing two sides to the bargaining table. The domestic developments in the country, however, provided little evidence that the President's promise could be fulfilled.

To what extent did the rebels and the government believe that aid could be restored if rebel attacks persisted? In other words, was the U.S. threat to the rebels credible, and what impact did it have on war dynamics? Despite U.S. assurances of aid restoration to the government, both warring groups had doubts about the actualisation of full aid. The government of El Salvador was concerned about aid being at risk, even after approval for non-lethal materials to be shipped to the country (Goshko 1991), which increased its willingness to seek a settlement. These concerns were justified due to domestic divisions within the U.S. and concerns that fully supporting the government would send contradictory signals regarding U.S. commitment to prosecute those responsible for the Jesuit killings. Providing aid became a polarising issue in the U.S. Evidence suggests that the government of El Salvador and the rebels were aware of the growing tensions within the U.S (Prisk 1991). Comparing their situation to Vietnam, FMLN leadership believed that bipartisan divisions over aid would favour their cause. Additionally, the rebels relied on a lobbying organisation in Washington to undermine U.S. aid (Prisk 1991). The rebels held the belief that U.S. aid would not immediately diminish, but they were convinced that ongoing support for the government was not a credible signal in the long term. The transparency of domestic developments in the U.S. enabled both parties to accurately perceive the constraints faced by the third party, making bluffing non-credible.

The case suggests that once the bluffing was quickly uncovered, it had limited success in completely breaking the rebels' resolve, as intended by the third party. The rebels remained determined to continue their struggle and had faith in the possibility of societal transformation in the late 1980s. Despite facing internal support losses and resource deterioration, the FMLN chose to engage in dialogue with the government as a means to regroup for the next phase of the conflict. This strategic decision aligned with their belief that U.S. aid would eventually cease, providing an opportunity for the rebels to win. In essence, the rebels' beliefs about expanding the conflict were consistent with our model's predictions that bluffing about continued U.S. assistance would have limited effectiveness in altering their calculations. The temporary dialogue between the combatants was perceived by the rebels as part of their ongoing plan to achieve victory. Recognising the importance of improving their overall position in the

conflict, the rebels opted for negotiations, viewing it as an opportunity to enhance their image and buy time for their preparations to resume the struggle (Prisk 1991).

The rebels anticipated that ARENA's election win might prompt a military clash and a destabilising coup, risking chaos in El Salvador and diminishing the likelihood of sustained U.S. intervention, contrary to earlier promises (Prisk 1991). Their strategic outlook, sceptical of U.S. commitment, matched model predictions that bluffing has limited effects in scenarios where intentions are quickly discerned. U.S. efforts to undermine rebel resolve were partly unsuccessful, leading the rebels to engage in temporary negotiations within their broader strategy. Initial talks in the mid-1980s faltered, culminating in a 1992 peace agreement beneficial to the rebels. Despite a weakened military stance, the FMLN secured significant gains, including integration into the National Civilian Police and recognition as a political party (Prisk 1991).

Case: Homogeneous Information Levels and Wrong Learning About Intervention Insurgency Struggle in the Warsaw Uprising, 1944

This case allows us to illustrate the effectiveness of bluffing when both parties possess wrong information about intervention and learn erroneously (Table 3). Our prediction suggests that bluffing will have the greatest chance of success when combatants know little about third-party's true intentions in the war and update information incorrectly. In this context bluffing will also alter group calculations, and in doing so leave its mark on the conflict.

Table 3. Information Sets in Warsaw Uprising.

Third-Party State		Warring Actors		
Transparency of Signals	USSR	Signal Bias	Polish Resistance	Germany
Domestic Costs	No	Political and Executive Constraints	Yes, 1-Lublin Committee (USSR initiated) vs. AK (loyal to Polish gov. in exile, London), 2-constraints within Polish gov. in exile ¹⁰	No
Political and Executive Constraints	No			
Transparency	No, Authoritarian	Leader Affiliation with the 3 rd party	No	No
History of Intervention	Small <5	Secret Negotiations	Yes	No
Responsible Leader	Yes, Stalin	Responsible Leader	Yes, Mikołajczyk ¹¹	Yes, Hitler
Transparency of the Signals?	No	Bias in Perceptions?	Some	Limited

Outcome: The clear and transparent signals regarding U.S.'s constraints on intervention, combined with the lack of evident cognitive bias, enabled both parties to form accurate initial information sets.

The case of Polish uprising against the Germans in 1944⁶ illustrates the role of bluffing in helping the Soviets achieve their goal in Poland by relying on the warring groups' erroneous information levels and updating. Expecting assistance from the Soviets, whose arrival in Warsaw appeared imminent, the Poles began an armed struggle against the Germans in the capital. With approximately 40,000 soldiers armed mostly with light weapons, the Polish Home Army could not expect to last long in the fight against the powerful occupier equipped with tanks, artillery, and planes (Siekierski 2004). The urgency to initiate immediate action without notifying Stalin was due to strategic considerations. The Warsaw uprising was perceived as a way for the pro-London Poles to gain control. For that, Soviet entry into Warsaw had to be preceded by the AK seizure of the city (PSZ III 1950, p.656). Yet, the decision to begin the insurgency was based on the belief that the Red Army would join the Poles in liberating Warsaw (D'Este 2004). General Pelczynski expressed this clearly, *'We were convinced that the occupation of Warsaw was one of the aims of the Soviet offensive. On about 21 July we became convinced that the Russians would cross the line of the middle Vistula without any strong opposition from the Germans'*. (Pelczynski's oral statement, Chiechanowski, 1974)

Though Polish-Soviet relations were deteriorating in 1943 when the graves of Polish officers murdered by the Soviets were discovered in Katyń, the Polish underground still believed the Red Army would join them in fighting the common enemy. Reports from the Soviet Foreign Office announcing that the country's troops had launched the liberation of Polish territory with an objective to 'squash the enemy', and 'to help Polish people re-establish an independent, strong, and democratic state' coupled with Stalin's promise during PM Mikolajczyk's August 9 visit signalled to the insurgency that expectations of aid were indeed credible (PSZ III 1950, 826–830). In addition, the Poles believed erroneously, as would later become evident, that the U.S. and Great Britain would pressure Stalin to aid the insurgents who were allied with the West and provided valuable war intelligence (Applebaum 2004).

Just like the Poles, the Germans expected the Soviets to join the insurgents to liberate Warsaw. They were already battling the Red Army and were certain of the approaching aid to the resisters (Davies 2003). Even when the Soviets reached the capital but refused to join the insurgents, the Germans failed to consider that withholding assistance to the Poles was deliberate. Instead, they believed the Red Army would provide the anticipated relief. The army log, for example, interpreted the Soviet attack on Radzymin, a town near Warsaw, as an attempt to break through and help the insurgents (9th Army Report, 1944). It was thus clear that both sides expected the Poles to receive military support from the Soviets. Finding the bluff convincing, the Poles were motivated to engage in battle, while the Germans were keen on quashing the insurgency by all means, viewing it as a prime example of European defiance they aimed to eradicate

once and for all (Heinz 1944). Anticipated aid came as no shock. Given the advance of the Soviets, Warsaw was critical in keeping the west-east connection open (Richie 2013). In fact, Hitler had declared Warsaw a fortress just days prior to the Warsaw Uprising, implying no one would leave (Heinz, 1944).

The U.S.S.R was counting on the Germans to stay and confront the Poles. The Soviets did not anticipate German retreat from Warsaw (PSZ III 1950, 826); rather, their bluff hinged on altering Polish incentives. Given that the Germans had already commenced counter attacks against the Soviets close to the capital just a day before the insurrection erupted, even AK commander, Bor-Komorowski had revised his estimation, indicating that a German withdrawal from Warsaw would not be part of their response strategy (oral evidence, Ciechanowski 1974). Eventually, AK calculations rested on Soviet aid which they believed would settle the outcome.⁷ The effectiveness of the bluff, however, would be undermined if it was swiftly uncovered. The Soviets employed mixed signals to keep the bluff alive. On the evening of September 10th, the Soviets finally approved the Western Allies' use of shuttle bases. Additionally, the sounds of battle echoed from across the Vistula as the Soviets initiated an attack on Praga and Soviet planes re-emerged over the city. A few days later, the extensive U.S. air fleet, having awaited Soviet landing permission for a month, finally appeared over Warsaw to drop supplies. Consequently, the AK Commander who was about to surrender to the Germans ended the armistice discussions (Cienciela 1955, Stawski 1950). In the words of Churchill, *'They wished to have the non-Communist Poles destroyed to the full, but also to keep live the idea that they were going to their rescue'* (Churchill 1953).

To what extent did bluffing as a strategy benefit the third party? Newly accessible Soviet archives offer an understanding of Stalin's complex reasoning behind withholding aid. Aware that the Polish resistance had regularly organised uprisings during Operation Tempest whenever the Red Army was in proximity, coupled with their enduring opposition against Russia following liberation, Stalin was not prepared to let a similar situation unfold in Warsaw. Stalin's stance was clear: the Poles could achieve victory, but only if they acted in conjunction with the Red Army (Mukhina 2006). Stalin's objective was to establish a Soviet satellite in Poland. Bluffing about aid was thus an effective strategy for the Soviet leader, allowing him to accomplish his political goal of eliminating Polish fighters that could jeopardise Soviet dominance of the country. The case of the Warsaw Uprising demonstrates the impact of credible bluffing and learning on conflict dynamics. Expectations of aid initially inspired the insurgents to take up arms against the enemy. When the bluff was eventually discovered, and the Poles realised their chances of success had vanished, most of the Home Army fighters were already killed. In this sense, Soviet bluff retained its credibility long enough to eliminate potential opposition to Stalin's takeover of Poland, thus rendering the strategy effective.

Case: Heterogeneous Information Levels and Learning About Intervention – Nixon’s Secret Negotiations During Vietnam War, 1973

The model demonstrates the somewhat limited utility of bluffing when one party has superior information than the other (see Table 4), but the less informed group eventually discovers the bluff. To show the strategy’s mixed effectiveness, we need to prove the existence of bluffing, the warring groups’ divergent perceptions of the bluff and the strategy’s partial success in securing third-party interest.

After more than a decade of active involvement in Vietnam, the United States was ready to disengage from the conflict. As domestic pressure intensified and waves of anti-war protests paralysed the country, the Johnson and then the Nixon administrations realised that continued presence in Vietnam was a dangerous and politically costly venture. Not surprisingly, before his re-election in 1972, Nixon pursued a peace agreement between the North and South Vietnamese that would terminate U.S. presence in the area and boost the candidate’s popularity (Small 1989). Getting the two parties to the negotiating table, however, was difficult as the North Vietnamese sought a decisive victory and refused to negotiate with Thieu, the South Vietnamese president. The latter, on the other hand, felt that a peace deal in 1973 would diminish his power. In order to encourage negotiations, Nixon thus attempted to change the warring groups’ preferred strategies. To ensure the North Vietnamese’s acquiescence, Nixon ordered the bombings of Hanoi hoping that a signal of U.S. resolve would prompt the Northern enemy to comply. Aiming to sway Thieu’s position on negotiating, Nixon secretly

Table 4. Information Sets in Vietnam.

Third-			Warring Actors	
Transparency of Signals	USA	Signal Bias	S.Vietnam	N.Vietnam
Domestic Costs	Yes, Protests	Political and	Limited	Limited
Political and Executive Constraints	Yes: Divided Government	Executive Constraints		
Transparency	Yes, Democracy	Leader Affiliation with the 3 rd party	Yes	No
History of Intervention	5<Medium <10	Secret Negotiations	Yes	No
Responsible Leader	No, Nixon	Responsible Leader	No, Thieu	Yes, Le Duan
Transparency of the Signals?	No	Bias in Perceptions?	Some	Limited

Outcome: Despite the transparency of signals regarding the U.S’s intentions, their interpretation was distorted due to South Vietnam’s cognitive bias, resulting in a flawed information set. In contrast, the clear signals were strengthened in the case of North Vietnam due to the limited cognitive bias, leading to an accurate information set.

pledged full U.S. assistance in case of the North's violation of the agreement (Wicker 1975).

The North and the South differed in their perceptions of the signals. The South believed that the U.S. would in fact resume its involvement in the conflict at the first sign of agreement violation (Wicker 1975). Despite domestic opposition to war, Nixon issued secret promises to Thieu that played a vital role in fulfilling the administration's goal of ending the conflict. The North, however, doubted credibility of the promise but decided to eventually negotiate (Bibby 1985). In light of the strong anti-war protests in the U.S., it was clear by 1973 that U.S. presence in Vietnam would diminish (Small 1989). At the same time, North Vietnam's leadership agreed to negotiations to secure a prompt U.S. withdrawal and continue its struggle without facing the external intervener (Bibby 1985).

To what extent was the U.S. administration relying on bluffing to secure the negotiations? And what was the strategy's success in light of the combatants' varying information levels about U.S. intentions? First, evidence shows that Nixon's secret promises of assistance in case of the North's attack, which were fundamental in bringing Thieu to the negotiating table (Wicker 1975), could not be carried out. The level of domestic and Democratic Congressional opposition before the peace deal suggested that the U.S. presence in Vietnam was coming to an end with dim hopes for renewed presence (The New York Times 1975). In light of strong anti-war momentum in 1973, even popular presidents could not be in a position to maintain U.S. presence in Vietnam for long. Clearly a bluff, the secret promise of U.S. assistance to the South was effective in selling the negotiations to Thieu, a step guaranteeing the strategy's initial success. This finding is consistent with our model since it shows that when the warring groups have different information levels about third-party intentions, bluffing will offer a glimpse of success.

Second, the case suggests that bluffing was only partially effective because North Vietnam's leadership was confident that U.S. presence in Vietnam was coming to an end in light of the costly domestic opposition. Correctly perceiving U.S. intentions, in contrast to Thieu, North Vietnam quickly violated the peace agreement. When U.S. aid to the South failed to materialise by 1975, South Vietnam's resolve was diminished and paved the way for capitulation. U.S. bluffing was thus successful in achieving initial peace agreement and providing the country with an exit strategy. At the same time, the tactic failed to maintain the peace accord that would guarantee the survival of South Vietnam's government, all of which were in the interest of the U.S. In conclusion, the case suggests that bluffing under heterogeneous information offers, at best, a short-term solution to a strategic third party.⁸

Discussion

First, our study defines bluffing as a deliberate deception, different from cases where entities are willing but ultimately unable to meet their commitments. We provide evidence of this deception through examples of leaders' verbal commitments to aid allies, made despite their knowledge of existing constraints. Such leaders, knowing the impracticality of fulfilling these promises, would not have proceeded with their assurances if acting in good faith. Addressing the possibility that third parties might have believed in their ability to intervene despite being unaware of their limitations, we assume that third parties typically possess a more accurate understanding of their domestic constraints (transparency) and the degree to which these constraints limit them (bias) than do combatants. This premise is based on the understanding that all political changes, including leadership or regime shifts, are, to a large extent, endogenous and predictable. Our model, therefore, is built on the premise of asymmetric information, suggesting that leaders are aware of the feasibility of their commitments and the limitations thereof, even considering the impact on their successors. This concept is underscored by the situation in Vietnam, where, despite the potential for reputational harm, the strong anti-war sentiment in 1973 meant that no U.S. leader could maintain the nation's military engagement indefinitely, emphasising the realistic bounds of political commitments. For future studies, we recommend investigating how modifying biases in third parties' perceptions of their own ability to intervene might affect conflict dynamics differently compared to bluffing.

Second, our models build on the concept that decisions made by actors are shaped by rather path-dependent information sets, implying that their updates are contingent upon information acquired in prior periods. This insight leads us to regularly update the initial bias (b) and transparency (t) values to underscore our assertion of the enduring nature of leader bias once it is established. As a result, our framework does not support the scenario where an initially incorrect information set is accurately corrected. This necessitates initialising the values of t and b as in the case with homogeneous but incorrect information sets, but then permitting agents to update at narrow intervals. We model this scenario and present our findings in the Appendix, Section 7. The results indicate that when we move away from the assumption that one's starting point dictates their learning pace, convergence occurs fast, almost similar to scenarios with homogeneous and correct information sets with the right learning rule. This suggests that the rate of updating may be more crucial than the initial starting point.

Conclusion

The paper uses agent-based models, simulations and case studies to examine the effectiveness of third-party's bluffing in civil wars, noting its success varies with combatants' initial knowledge and ability to identify the true intentions of external intervenors. It highlights that bluffing can significantly impact the strategic calculations of rebels and governments moving beyond a static analysis. The key finding is that bluffing is most effective under specific conditions: when combatants have incorrect information about third-party actions and misjudge their updates of expectations, offering a dynamic view of strategic behaviour in conflicts.

Additionally, the paper offers ample policy suggestions for conflict management, showing that bluffing about intervention may be a strategy worth risking, for it delivers substantial benefits to the external party under the right conditions. For example, a third party who seeks to destabilise the central government may bluff about providing aid to the rebels, increase their resolve, and in turn inspire more fighting. Bluffing about external intervention could also be utilised by third parties as an effective strategy of conflict resolution under specific conditions. For example, bluffing by a non-democratic/less transparent state might be more successful because the group that finds the bluff credible might be unable to discover its true nature before deciding to end the fighting. Contrary to common belief, such states might then be useful in bringing the rebels and the government to the negotiating table (settlement by bluff), albeit with more favourable terms for its ally, and, in doing so, pave the way for settlement.

Lastly, our case studies encompass diverse scenarios, including insurgencies as seen in El Salvador, and insurgencies occurring in conjunction with interstate wars, as in Vietnam and the Warsaw Uprising. We examine instances where third parties back either rebels or governments, and though not featured in our cases, scenarios with multiple external interveners could also occur. These situations may contain nuanced differences. For example, overcoming informational hurdles could be more challenging in contexts involving both interstate and civil wars, or multiple third parties, where the cost-benefit analysis of intervention can become murkier as the number of involved actors increases. Furthermore, the factions allied with a third-party state may be more prone to cognitive biases. Despite these complexities, these case studies collectively demonstrate bluffing's effectiveness in a broad range of circumstances.

Notes

1. We choose a uniform distribution due to its simplicity and the absence of prior expectations regarding outcomes. In this case, all possible states of intervention

are assumed to be equally likely. We use the same rationale when selecting the uniform distribution from which we draw expectations of intervention for warring actors.

2. We use 0.006 instead of 0 for easier convergence. We also test alternative numbers (0.005 and 0.007) in the Appendix, Section 6 for sensitivity analysis.
3. Our decision to keep the value unchanged and to refrain from permitting small updates is influenced by our perception of the informational landscape in social and political arenas as chaotic, characterised by significant error margins, randomness, challenges in learning, and the persistent presence of learning biases. Allowing minor adjustments would not alter the model's forecasts. This would however necessitate narrowing the learning intervals in our models resulting in longer convergence rates across all scenarios.
4. We do more illustrative runs, simulating our models with alternative values of t_g , r and $b_{g,r}$ in Appendix, Section 3. In these runs, the convergence rate ranges between 20 and 40 cycles. In our models, a 'cycle' is conceptualised as an abstract unit of simulation time, crafted without direct correspondence to real-world time periods such as days, months, or years. This design choice allows cycles to function as theoretical intervals during which actors process and react to information. The notion of cycles being 'context-bound' yet 'consistent across models' underscores this adaptability, ensuring that the framework can be applied uniformly despite varying scenarios. It is thus crucial to understand that referencing '40 cycles' within a specific scenario, like El Salvador, does not equate to a protracted period similar to rounds of negotiations or specific chronological sequences. In the framework of our models, '40 cycles' signifies a swift period of adjustment for actors to shifts in bias (b) and transparency (t). If the scenario shifts to one of low transparency and high bias, mirroring a situation where the information set is inaccurate and not properly updated, we would anticipate the number of cycles needed to detect a deception to increase significantly, up to 500. This example underlines the necessity of grasping the relative differences in the number of cycles needed across various scenarios within the same context (for instance, comparing 40 cycles to 500 cycles). It stresses the importance of making consistent scenario comparisons to accurately interpret the outcomes of our model. It's not the absolute number of cycles that matter but the comparative analysis of how different settings and strategic dynamics influence the speed and nature of decision-making and strategy adaptation.
5. We also adjust the uninformed actor's learning intervals by dividing by $\text{cycle} \times 0.5$ instead of $\text{cycle} \times 0.4$. This is slightly faster to account for potential learning from both third-party actions and those of the informed enemy. See Appendix, Section 2, Table 2 for all the initialisation values and learning intervals in the experiment.
6. The conflict, though not labelled a civil war, shares its characteristics, making it relevant for testing our theories. The Warsaw Polish Underground, with its distinct political ideology and governance, rebelled against the Nazi-imposed government in a manner akin to a civil war, where governments face organised insurgencies (Ney-Krwawicz 2004/2005). Our model extends beyond civil wars to encompass internal conflicts broadly.
7. At a 25th July Bor-Komorowski Staff meeting prior to the insurgency, calculations were based on how much they could withstand German assault which could even be a fortnight if they could augment their arsenal from German

captures or British air drops. Ultimately, he reasoned that the outcome would be settled with the arrival of the Russian army (Cienciela 1955, Skarzynski 1969).

8. In Appendix, Section 4, we simulate the three case studies by initialising values with the information from Tables 2, 3, and 4, along with approximate real-world relative capabilities.
9. The Vdem dataset shows some executive constraints (5) in 1989 but no political constraints (0).
10. The Polish Government in exile included pre-war Opposition Parties. Furthermore, some coordination problems hinted at discord among the President, Cabinet and Supreme Command.
11. Since Mikolajczyk served as the Deputy PM under PM Sikorski, he is responsible by affiliation.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Seden Akcinaroglu is a Professor at Binghamton University. Her research focuses on international rivalries, internal wars and strategies of violent non-state actors. She is the author of *Private Militaries and the Security Industry in Civil Wars: Competition and Market Accountability* (Oxford University Press, 2020) and *Battle for Allegiance: Governments, Terrorist Groups and Constituencies in Conflict* (University of Michigan Press, 2020). Her work has been published in peer reviewed journals such as *Journal of Conflict Resolution*, *Journal of Peace Research*, *International Interactions* and *Conflict Management and Peace Science*.

Sebnem Bora is an Associate Professor at Ege University where she teaches complex adaptive systems and agent-based modelling. Her research focuses on complex systems, agent-based modelling and simulation, and artificial intelligence. She has published many research articles on multi-agent systems, agent-based modelling of biological systems, and optimization of complex systems. Her work has been published in peer-reviewed journals such as *Cluster Computing*, *Simulation*, *The Journal of Supercomputing* and *Informatica*.

Elizabeth Radziszewski is an Associate Research Scientist at START where she co-leads the Global Responses to Asymmetric Threats project. Her research focuses on irregular warfare, international rivalries, foreign policy, and creativity/innovation. She is the author of *Private Militaries and the Security Industry in Civil Wars: Competition and Market Accountability* (Oxford University Press, 2020) and *Social Networks and Public Support for the European Union* (Routledge, 2013). Her work has been published in several peer-reviewed journals and policy magazines such as the *Wilson Quarterly*, *Journal of Conflict Resolution*, *Small Wars & Insurgencies*, *Foreign Policy Analysis*, *Journal of Global Security Studies*, *International Interactions* and *Political Research Quarterly*.

ORCID

Seden Akcinaroglu  <http://orcid.org/0009-0004-3062-2385>

References

- 9th Army Report, 10 September 1944, WIH t 312, roll 346.
- Akcinaroglu, S. and Radziszewski, E., 2005. Expectations, rivalries, and civil war duration. *International Interactions*, 31 (3), 349–374. doi:10.1080/03050620500303449
- American interests in El Salvador, 1988. *Federal News Service*, 30 December.
- Applebaum, A., 2004. In Warsaw, a 'good war' wasn't. *The Washington Post*, 2 August.
- Balch-Lindsay, D. and Enterline, A., 2000. Killing time: The world politics of civil war duration, 1820–1992. *International Studies Quarterly*, 44 (4), 615–642. doi:10.1111/0020-8833.00174
- Bennett, S.D., 2008. Governments, civilians, and the evolution of insurgency: Modeling the early dynamics of insurgencies. *Journal of Artificial Societies and Social Simulation*, 11 (4), 1–7.
- Bibby, T., 1985. *Vietnam: The end*, 1975. Available from: <https://smallwarsjournal.com/documents/bibby.pdf> [Accessed 10 September 2022]
- Churchill, W., 1953. *Triumph and tragedy*. London: Houghton Mifflin.
- Ciechanowski, J.M., 1974. *The Warsaw rising of 1944*. New York: Cambridge University Press. (Pelczynski, T.'s oral statement, Czwarte Zebranie Relacyjne w inst. J. Piłsudskiego w Londynie 18/4/1951: Decyzja Podjęcia Walki o Warszawę w 1944 r', APUST).
- Cienciela, A.M., Aug 1955. *The Warsaw rising of 1944 in the light of Polish-Soviet relations during world war II*. Thesis (PHD). Univ of Mc Gill.
- D'Este, C., 2004. Rising '44: Betraying Warsaw. *The New York Times*, 25 July.
- Davies, N., 2003. *Rising '44: The battle for Warsaw*. New York, NY: Penguin.
- Drink, J., 1990. The Senate voted to slash aid in half. *The Associated Press*, 19 October.
- Fearon, J.D., 1995. Rationalist explanations for war. *International Organization*, 49 (3), 379–414. doi:10.1017/S0020818300033324
- Filson, D. and Werner, S., 2002. A bargaining model of war and peace: anticipating the onset, duration and outcome of war. *American Journal of Political Science*, 46 (4), 819–837. doi:10.2307/3088436
- Findley, M. and Young, J.K., 2007. Fighting fire with fire? How (Not) to neutralize an insurgency. *Civil Wars*, 9 (4), 378–401. doi:10.1080/13698240701699482
- Goshko, J., 1991. U.S. To begin sending Salvadoran aid. *The Washington Post*, 27 June.
- Heinz, R., *Kurze Schilderung des Kampfes um Warschau, August–Oktober 1944*, p. 22.
- Jackson, J.A., San-Akca, B., and Maoz, Z., 2020. International support networks and the calculus of uprising. *Journal of Peace Research*, 57 (5), 632–647. doi:10.1177/0022343319885181
- Kydd, A., 2006. When can mediators build trust? *American Political Science*, 100 (3), 449–462. doi:10.1017/S0003055406062290
- Mattes, M. and Savun, B., 2010. Information, agreement design, and the durability of civil war settlements. *American Journal of Political Science*, 54 (2), 511–524. doi:10.1111/j.1540-5907.2010.00444.x
- Montgomery, T.S., 1995. *Revolution in El Salvador: from civil strife to civil peace*. Boulder, Colorado: Westview.
- Mukhina, I., 2006. New revelations from the former Soviet archives: The Kremlin, the Warsaw uprising, and the coming of the Cold War. *Cold War History*, 6 (3), 397–411. doi:10.1080/14682740600795584

- Ney-Krwawicz, M., 2005. *The Polish underground state and the home army*. Warsaw Uprising Website. Available from: <http://www.warsawuprising.com/state.htm>. [Accessed 26 September 2022].
- Pecheninka, A.O. and Bennett, D.S., 2017. Violent and non-violent strategies of counterinsurgency. *Journal of Artificial Societies and Social Simulation*, 20 (4), 11. doi:10.18564/jasss.3540
- Powell, R., 2004. Bargaining and learning while fighting. *American Journal of Political Science*, 48 (2), 344–361. doi:10.1111/j.0092-5853.2004.00074.x
- Prisk, C.E., 1991. *The comandante speaks: memoirs of an El Salvadoran guerrilla*. Boulder. Colorado: Westview Press.
- Prorok, A.K., 2016. Leader incentives and civil war outcomes. *American Journal of Political Science*, 60 (1), 70–84. doi:10.1111/ajps.12199
- PSZ, 1950. *Polskie Siły Zbrojne w Drugiej Wojnie światowej Krajowa, Armia*. III
- Rauchaus, R.W., 2006. Asymmetric information, mediation and conflict management. *World Politics*, 58 (2), 207–241. doi:10.1353/wp.2006.0027
- Regan, P.M., 2002. Third-party interventions and the duration of intrastate Conflict. *The Journal of Conflict Resolution*, 46 (1), 55–73. doi:10.1177/0022002702046001004
- Richie, A. 2013. *Warsaw 1944: Hitler, Himmler, and the Warsaw uprising*. Farma, Straus and Giroux. (Fernschreiben 4 August 1944 to the AOK, [BA-MA H-12-9/5])
- Schultz, K.A., 2001. *Democracy and coercive diplomacy*. Cambridge: Cambridge University Press.
- Sciolino, E., 1994. U.S. Policies under siege. *The New York Times*, 16 April.
- Siekierski, M., 2004. Remembering the Warsaw uprising. *Hoover Digest*, 4, Available from: <https://www.hoover.org/research/remembering-warsaw-uprising> [Accessed 26 September 2022].
- Skarzynski, A., 1969. *Polityczne przyczyny powstania warszawskiego*. Warsaw: Państwowe Wydawn. Nauk. <https://catalog.libraries.psu.edu/catalog/245539>
- Slantchev, B.L., 2003. The principle of convergence in wartime negotiations. *The American Political Science Review*, 97 (4), 621–632. doi:10.1017/S0003055403000911
- Small, M., 1989. *Johnson, Nixon, and the doves*. New Brunswick: Rutgers University Press.
- Smith, A. and Stam, A.C., 2004. Bargaining and the nature of war. *Journal of Conflict Resolution*, 48 (6), 783–813. doi:10.1177/0022002704268026
- Stawski, T., 1950. Wrzesniowy Czyzys Powstania Warszawskiego - September crisis of the Warsaw rising. *Kultura*, (9/35), 121–125.
- Thyne, C., 2006. Cheap signals with costly consequences: The Effect of interstate relations on civil war. *Journal of Conflict Resolution*, 50 (6), 937–961. doi:10.1177/0022002706293675
- Thyne, C., 2012. Information, commitment, and intra-War Bargaining: The effect of Governmental constraints on civil war duration. *International Studies Quarterly*, 56 (2), 307–321. doi:10.1111/j.1468-2478.2012.00719.x
- Walter, B.F., 2006. Information, uncertainty, and the decision to secede. *International Organization*, 60 (1), 105–35. doi:10.1017/S0020818306060048
- Wicker, T., 1975. Mr. Thieu tells his side of it. *The New York Times*, 22 April.
- Zott, C., 2002. An agent-based explanation of inefficient bargaining under private information. *Journal of Conflict Resolution*, 46 (6), 727–775. doi:10.1177/002200202237927