Bargaining under polarization: The case of the Colombian armed conflict

Sigifredo Laengle

Department of Management Control and Information Systems, University of Chile

Gino Loyola

Department of Management Control and Information Systems, University of Chile

David Tobón-Orozco

Department of Economics, Universidad de Antioquia

Abstract

A bargaining framework and a measure of conflict polarization are developed from two elements: (i) hatred-based negative externalities experienced by the parties to the conflict, and (ii) penalties the parties impose on their delegated negotiators when concessions are made in the bargaining process. The framework establishes agreement and disagreement regions and it is shown that a necessary condition for a negotiated solution is the adoption of a dual policy that combines dissociative political and military strategies. This analytical approach is applied first to polarized conflicts generally and then to the specific case of the internal conflict in Colombia between that country's government and the FARC guerrilla group. The model provides a rationale for the complex dynamic of Colombia's current peace process, which has involved a preliminary agreement and its subsequent rejection in a national referendum. Our analysis highlights the successful dissociative political-military strategy followed by the negotiators that enabled them to reach the agreement and the negotiators' underestimation of the hatred levels that led the majority of the Colombian society represented in the referendum to vote the agreement down because they considered the concessions made by the government too generous to be acceptable.

Introduction

Peace talks, disarmament discussions (to decommission terrorist groups), political transitions and many other negotiation-related processes take place in highly polarized environments. The parties involved in these processes often prefer to let them stagnate into a deadlock or break down altogether, implying that they are willing to assume the costs of failing to reach an agreement.

There is an abundant literature based on game theory that attempts to explain the trade-off between conflict and cooperation and the conditions for an agreement/disagreement. Examples of such studies are the classic conflict and cooperation analyses by Hirshleifer (1991), Skaperdas (1992) and the bargaining approaches of Anderlini & Felli (2001), Crawford (1982), Fernandez & Glazer (1989) and Haller & Holden (1990). However, none of these works consider the roles played by negative sentiments like hatred or envy and the resulting polarization in either the breakdown of a negotiation process or the emergence of an equilibrium without cooperation.

To fill this gap, we propose a modified version of the classic Nash demand game in which negotiators delegated by the parties to the game conduct a bargaining process in a polarized environment. This polarization

Corresponding author:
david.tobon@udea.edu.co

results from the hatred each party feels for the other, a sentiment they transfer to their respective negotiators by means of penalties imposed on the latter for any concessions made during the bargaining.

The outcomes of this polarized negotiation are characterized using a four-region graph based on the values taken by two parameters: (i) the level of polarization in the conflict (the product of the parties' hatred levels), and (ii) the cost of the conflict (the difference in the size of the surplus between a situation of peace and a situation of war). The graph thus contains two agreement regions and two breakdown regions. This setup provides a rich analytical framework for the theoretical evaluation of different policies that could potentially achieve a negotiated solution in a polarized conflict. It will be shown that the four-region analysis leads to two main conclusions. First, a proper evaluation of strategies aimed at resolving a polarized conflict should include the most accurate possible identification of the conflict's starting point, that is, the stalemate region where the bargaining process is initially stuck. Crucial to this correct identification are good estimates of both the polarization level and the conflict cost. Second, a necessary condition for reaching an agreement in a polarized conflict is the implementation, ideally by both parties, of a dual policy consisting of a *dissociative* mixture of a *political* strategy affecting the polarization level (rhetoric or 'words') and a *military* strategy influencing the conflict cost ('facts'). Any policy not built around this dissociation will fail to bring about a resolution of the conflict. As for the characteristics of the agreement itself, we will show that there is a continuum of Pareto efficient negotiating solutions constituting an interval whose minimum and maximum equilibrium shares depend on the level of hatred of the parties. When the conflict cost is positive, these hatred levels, mediated by penalties, build a sort of commitment device for negotiators and thus ultimately constitute a source of bargaining power. The negotiator most likely to obtain the largest share in an agreement is the one who incurs the biggest penalty.

The proposed four-region framework is also useful for describing and evaluating peace initiatives adopted in real situations. Excellent examples of this are the various negotiation processes that have attempted to settle the internal armed conflict in Colombia, which only in recent years has made major advances towards a peaceful solution. This decades-long conflict between the guerrillas – primarily the group known as the Revolutionary Armed Forces of Colombia (FARC) – and the Colombian government has been highly polarized from the very beginning. This polarization is clearly evident in the many intransigent postures and inconclusive negotiations that have long prevented the achievement of a lasting solution.

We will therefore use our theoretical framework to assess three Colombian government policies adopted at different times in an attempt to solve the conflict. Two of them – the Demilitarized Zone plan and the Democratic Security policy – are past initiatives that failed while the third is the current Peace Process, launched in 2012. We conclude that this last policy succeeded in reaching a preliminary agreement in August 2016 because of the parties' willingness to combine open dialogue conducted in moderate rhetoric with an active military strategy. We also find that the rejection of this agreement in the September 2016 referendum can also be explained by our model. More specifically, our analysis suggests that the negotiators understated the hatred level of the median voter, misleading the government into making concessions to the FARC that went beyond the maximum the majority of the Colombian society represented in the referendum were prepared to tolerate.

The present study contributes to the growing literature in political economy devoted to the theoretical relationship between polarization and the incidence and intensity of conflicts (see Esteban & Ray, 1994, 1999, 2008, 2011; Esteban & Schneider, 2008).[1] These works use a notion of social polarization based on concepts of alienation and intergroup antagonism that resembles the concept of hatred or envy underlying our conflict polarization measure. The authors' polarization metric affects non-linearly either the intensity or the incidence of the conflict, which is reminiscent of our results on polarization and the occurrence of a negotiation breakdown. Despite the similarities, however, their approach and the one presented here differ in two crucial aspects. Whereas they model the *conflict* scenario as a rent-seeking contest, we model the *peace* scenario as a distributive bargaining problem. Moreover, their results do *not* depend on the level of antagonism but *do* on the size of each interest group, contrasting with our framework in which the hatred levels are crucial in characterizing the relationship between polarization and conflict incidence.

As for studies specifically on the Colombian case, our analysis is also related to the relatively recent literature taking a game-theoretic approach to the search for a

---

[1] Other related contributions include the proposal of an alternative polarization index (Reynal-Querol, 2002) and the study of evidence on the relationship between polarization and the incidence of civil wars (Montalvo & Reynal-Querol, 2005).

peaceful resolution. This research generally models solutions in terms of classical games comparing the economic benefits/costs of maintaining the conflict versus reaching an agreement (Gorbaneff & Jacome, 2000; Zambrano & Zuleta, 2017). Other articles have examined the circumstances under which Colombian society can continue to survive the conflict and its costs, and the determinants of the transfer that would be required to reach an agreement (Salazar & Castillo, 2001; Zuleta, Villaveces & Andonova, 2013). Although these models describe various characteristics of the conflict, they neither account formally for the polarized environment nor consider the latter's crucial role in the failure of previous peacemaking efforts.

The remainder of this article is organized as follows. Section 2 proposes a model of bargaining under polarization, analyzing the main properties of its equilibrium and its implications for conflicts in general; Section 3 applies the model to the conflict in Colombia by analyzing the dynamic behind some relevant peacemaking attempts; and Section 4 presents a summary of our main conclusions. All the technical proofs are collected in the Online appendix.

A model of negotiation under polarization

In this section we develop a general model of a static distributive bargaining game between two delegated negotiators, each one representing one of the two parties to a conflict. Since both parties exhibit a hatred-based externality, each subjects his or her own negotiator to negative pressure proportional to the gain obtained by the other party's negotiator. This pressure takes the form of a penalty imposed for ceding any part of the good in dispute. From these penalties we derive a measure of the conflict's polarization level.

The bargaining game
The proposed game is based on the classic Nash demand game (Nash, 1953). The players are negotiators A and B, each representing one of the two parties as they attempt to resolve an ongoing conflict by means of a bargaining process modelled as the distribution between them of a pie of size π. The pie can be thought of as a set of economic and political claims in dispute measured in a common unit. Each negotiator simultaneously demands a share of the pie expressed by a number $x_i \in [0, \pi]$ where $i = A, B$. If $x_A + x_B \leq \pi$, an agreement is reached and each player gets the amount he or she demanded. If, on the other hand, $x_A + x_B > \pi$, there is a disagreement (negotiation breakdown), which can

be interpreted to mean the conflict continues. In this latter scenario each negotiator gets a conflict payoff (i.e. his outside option), which will be described in more detail later.

Polarization in this context of delegated negotiation is modelled through externalities and penalties as follows. We assume that the two parties represented by the negotiators each experience a negative externality – a kind of envy or hatred – to the extent they are made better off by the surplus they obtain and worse off by the share of the pie captured by their counterpart.[2] We also assume that each party transfers (completely or partially) the externality they experience to their negotiator. This transfer takes the form of a penalty proportional to the concessions made by each negotiator to their counterpart, which are interpreted by the parties as a betrayal of their fundamental principles.

In the case of an agreement, these assumptions are captured by the payoff function for negotiator i given by $U_i(x_i, x_j) = x_i - \lambda_i x_j$, where $\lambda_i \geq 0$ is the marginal penalty imposed on negotiator i for the share he or she concedes to negotiator j, for all i, j = A, B; $i \neq j$, and reflects party i's negative externality, that is, their individual hatred level.[3] In the case of a disagreement, the *net* conflict payoff $\underline{U}_i$ to negotiator i is given by $\underline{U}_i = V_i - \lambda_i V_j$, where V_i, V_j represent the *gross* conflict payoffs to the negotiators, that is, the payoffs in the absence of negative externalities and penalties.[4] Note that these gross conflict payoffs can be negative, which would make sense, for example, in the context of a war.

In order to obtain feasible solutions, we assume that $\underline{U}_i \geq -\lambda_i \pi$ for all i.[5] To simplify the presentation of our main results we adopt the two definitions given below.

---

[2] This notion of envy is used by Varian (1974) in the context of social choice to define an equitable distribution as one in which envy is absent. A similar concept in a negotiation context is employed by Kirchsteiger (1994) in an ultimatum game model.

[3] See the Online appendix A2 for a formal treatment of these concepts of externalities, delegated negotiation and penalties.

[4] This assumption differs conceptually from the one made by Esteban & Ray (2008) regarding the role played by alienation or envy in payoffs from an alternative scenario. Whereas in our framework a preference externality is present even if negotiations fail, in Esteban and Ray's model alienation coefficients are absent in the peaceful status quo scenario. A justification for our assumption is that in addition to the penalty in the agreement case, each party also penalizes their negotiator if their counterpart obtains a payoff in the conflict scenario. This may occur if the payoff is interpreted as a failure of the military strategy also delegated to (and therefore conducted by) the same negotiator.

[5] This condition ensures that equilibrium partitions are non-negative and smaller than the pie up for division (see proof of Proposition 1).

In subsequent sections they will also prove useful for deriving a framework that allows us to perform both normative and descriptive analyses of a polarized conflict.

The first definition expresses the conflict cost as the difference in the size of the surplus between a situation of peace and a situation of war.

Definition 1: Y, the conflict cost, is defined as $Y \equiv \pi - (V_A + V_B)$.

The second definition, based upon our notions of externalities and penalties, proposes a polarization metric in the context of a negotiation.

Definition 2: Λ, the conflict polarization level, is defined as $\Lambda \equiv \lambda_A \lambda_B$.

Polarization is typically defined as a division into two groups with sharply contrasting sets of opinions and beliefs, which have both *causes* and *consequences*. Our proposed polarization metric is related to this definition. Regarding causes, we posit that a plausible source of polarization is the negative sentiments such as hatred or envy arising from ideological, religious, historical or ethnic issues. This approach does not imply that these sentiments are a necessary condition for polarization but rather that they may be a reasonable sufficient condition. As for consequences, we assert that a concrete manifestation of sharply contrasting opinions in the context of a bargaining situation is exemplified by positions that are very difficult to reconcile through an agreement. We show formally that in the classical situation in which conflict cost is positive, negative sentiments and penalties result in negotiators taking more polarized positions as they introduce a source of aggressiveness additional to the one arising from the zero-sum nature of the Nash demand game. In fact, if the parties' hatred levels are sufficiently high, the result may be a disagreement.

The equilibrium

Let X^* be the agreement set (Nash equilibrium) of the bargaining game just described. We then say that the shares (x_A^*, x_B^*) belong to this set if x_i^* solves the following program for negotiator i:

$$\max_{x_i} \; U_i(x_i, x_j) \qquad (1)$$

subject to

$$0 \leq x_i \leq \pi \qquad (2)$$

$$U_i(x_i, x_j) \geq \underline{U}_i \qquad (3)$$

$$x_i + x_j \leq \pi \qquad (4)$$

where i, j = A, B with $i \neq j$. Condition (2) is the program's feasibility constraint ensuring that each negotiator's demand is non-negative but smaller than the total pie at stake. Condition (3) is the individual rationality constraint, according to which a negotiator is willing to participate in the bargaining process only if he or she obtains a utility equal to or greater than the net payoff he or she got in the conflict scenario. Lastly, condition (4) is the agreement constraint imposed by the rules of the game.

The following proposition characterizes the set of agreements X^* computed on the basis of the above program.

Proposition 3: Consider the following conditions: (C1) $\Lambda \leq 1$ and $Y \geq 0$; (C2) $\Lambda \geq 1$ and $Y \leq 0$. If either of conditions C1 or C2 holds then the agreement set is not empty and is characterized by

$$X^* = \left\{ (x_A^*, x_B^*) : x_A^* \in [\underline{x_A}, \overline{x_A}], \; x_B^* \in [\underline{x_B}, \overline{x_B}], \; \text{and} \; x_A^* + x_B^* = \pi \right\}$$

where

$$\underline{x_i} \equiv \frac{\lambda_i \pi + \underline{U}_i}{1 + \lambda_i} \qquad (5)$$

$$\overline{x_i} \equiv \frac{\pi + \underline{U}_j}{1 + \lambda_j} \qquad (6)$$

and $i, j = A, B; i \neq j$.

Thus, there are two cases in which the outcome of the negotiation is an agreement. The first case arises if the level of polarization is moderate and the cost of the conflict is positive. This scenario is intuitive because when the size of the pie in a situation of peace exceeds its size in a situation of war ($Y \geq 0$), a concession-making process will be successful if the penalties incurred by both negotiators are sufficiently low ($\Lambda \leq 1$).

The second agreement case emerges when the polarization level is sufficiently high and the conflict cost is negative. In contrast to the first case, this scenario is much less intuitive because when polarization is sufficiently high ($\Lambda \geq 1$) a deal can be reached even if the pie in a situation of peace is smaller than that for a war ($Y \leq 0$).

When there is an agreement and the cost of conflict is positive (i.e. condition C1 in Proposition 3 holds), the own penalty λ a negotiator incurs is a source of bargaining power. In fact, the negotiator who is more likely to

obtain the largest fraction of the pie in a possible agreement is the one who gets the largest penalty, that is, the one with the largest value of λ.[6] Thus, although a high λ_i would appear at first glance to put negotiator i in a weak bargaining position, our results demonstrate that in a polarized bargaining situation, such a weakness can in fact become a source of strength. The intuition behind this result is that a high λ_i must be interpreted 'as if negotiator i's hands were tied relatively more tightly behind his or her back, in effect making penalties similar to the *commitment* tactics characterized by some authors in the bargaining literature (Crawford, 1982; Muthoo, 1996). Thus, if a settlement is reached it will be negotiator i's counterpart (j) who will make the bigger concessions.

From Proposition 3, it follows that an agreement will exist if the equilibrium interval characterized there is non-empty, that is, if $\underline{x_i} \leq \overline{x_i}$ for all i. It can be shown (see the Online appendix) that this condition is equivalent to

$$(1 - \Lambda)Y \geq 0, \qquad (7)$$

which thus constitutes a necessary condition for a negotiated solution. We can therefore perform the following comparative statics exercises over the agreement equilibria. If condition C1 in Proposition 3 holds, it is clear from (7) that any increase in λ_A or λ_B (and thus the polarization level Λ) makes reaching a settlement more difficult. The same is true of high levels of V_A or V_B (and therefore low levels of conflict cost Y).

Alternatively, if condition C2 in Proposition 3 holds, it follows from inequality (7) that increases in λ_A or λ_B (and thus the polarization level Λ) make negotiation easier. Similarly, high levels of V_A or V_B (meaning high levels of conflict cost Y in absolute value terms) also increase the likelihood of an agreement.

The next result follows directly from Proposition 3.

Corollary 4: If either $\Lambda = 1$ or $Y = 0$ (or both) holds, there is a unique agreement given by

$$X^* = \left\{ (x_A^*, x_B^*) : x_A^* = \underline{x_A} = \overline{x_A} \text{ and } x_B^* = \underline{x_B} = \overline{x_B} \right\}.$$

Another important property of the game is that the existence of an agreement is not guaranteed. This property is stated in the following proposition.

Proposition 5: Consider the following conditions: (C3) $\Lambda < 1$ and $Y < 0$; (C4) $\Lambda > 1$ and $Y > 0$. If either one of conditions C3 or C4 holds, the agreement set X^* is empty and thus there is a disagreement.

This statement implies that there are two cases in which bargaining powers become incompatible and the negotiation breaks down. The first case arises if the polarization level is moderate and conflict cost is negative. This disagreement scenario is intuitive because as long as both negotiators incur sufficiently low penalties and the polarization level is thus sufficiently low ($\Lambda < 1$), the conventional logic applies and it will be impossible to reach an agreement if the size of the pie in a situation of peace is smaller than the gross payoffs obtained by the two parties in a situation of war (Y < 0). In fact, this case has already been studied in the literature, albeit for situations with no polarization.[7]

The second negotiation breakdown case emerges if the polarization level is sufficiently high and conflict cost is positive. This novel scenario is particularly interesting because it occurs despite the fact that the size of the pie in a situation of peace exceeds that for a situation of war ($Y \geq 0$). The level of polarization is *too high* because the political penalties the delegated negotiators incur when they make concessions to their counterpart are also too high ($\Lambda > 1$). No compromise can then be reached, the negotiating process fails and a conflict scenario arises.

An alternative explanation for the negotiation breakdown can be derived from Equation (5). It is easily shown that if either one of conditions C3 or C4 in Proposition 5 holds, then

$$\underline{x_A} + \underline{x_B} > \pi.$$

Thus, if either the polarization level is very high or conflict cost is negative (but not both), the positions taken by the negotiators become so *extreme* that even the sum of their *minimum* equilibrium positions is higher than the surplus at stake. As a result, no agreement is possible.

The interpretation behind some of the negotiation outcomes described above can be improved by adopting the following two definitions. The first one defines a measure of the total welfare involved in a possible agreement:

---

[6] One way to see this is by assuming symmetric outside options such that $V_A = V_B$. In this environment, it can be shown that when $Y \geq 0$, if $\lambda_A > \lambda_B$ then $\underline{x_A} > \underline{x_B}$ and $\overline{x_A} > \overline{x_B}$.

[7] Adapting Myerson's (1991) terminology, this case might be called a non-essential bargaining problem.

Definition 6: U^*, the total agreement payoff, is defined as $U^* \equiv U_A(x_A^*, x_B^*) + U_B(x_A^*, x_B^*)$.

The second definition is the measure of the total welfare involved in a disagreement:

Definition 7: $\underline{U}$, the total net conflict payoff, is defined as $\underline{U} \equiv \underline{U}_A + \underline{U}_B$.

Based on these two definitions we establish the following equivalence between the conditions for an agreement/disagreement:

Corollary 8: The condition $U^* \geq \underline{U}$ is equivalent to either of conditions C1 or C2 in Proposition 3.

Given this equivalence, we can then restate the results of Propositions 3 and 5 as follows. Although the presence of externalities and penalties modifies the original Nash demand game, the conditions for an agreement/disagreement in a negotiation conducted in a polarized environment can be rewritten in terms of the classic condition comparing the total agreement payoff U^* with the total net conflict payoff $\underline{U}$.

We may therefore conclude that despite the apparent contradiction between the two agreement scenarios characterized in Proposition 3, the two in fact share the same underlying economic rationale once we incorporate the agreement and conflict payoffs *modified* by polarization. According to Corollary 8 the condition $U^* \geq \underline{U}$ holds in both of these cases, which means that the total payoff of an agreement exceeds the net payoff of the conflict.

Similarly, in the case of the two disagreement scenarios in Proposition 5, the economic rationale behind both of them can be reconciled by using the converse of Corollary 8: the condition that $U^* < \underline{U}$ clearly holds in either one, implying that the total *net* payoff from an ongoing conflict exceeds the payoff from an agreement.

Implications for a general conflict solution

Our analysis has established that depending on the values taken by parameters Λ and Y, the equilibrium of the distributive game under polarization generates four possible outcome regions grouped into two classes: agreement and negotiation breakdown. A diagram illustrating this framework is shown in Figure 1. The agreement outcomes are the unshaded regions II and IV, where condition C2 (in the former case) or C1 (in the latter case) of Proposition 3 holds and therefore a deal can be reached. The negotiation breakdown outcomes are the shaded regions I and III, where condition C4 (in the former case) or C3 (in the latter case) of Proposition 5
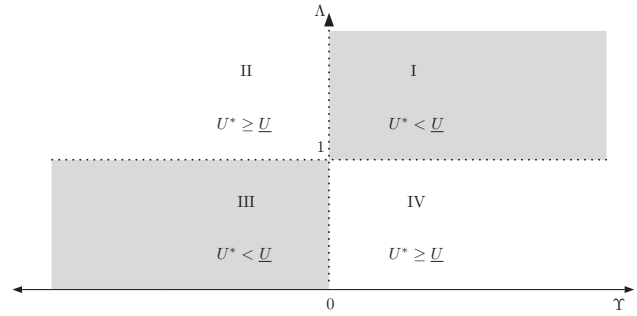

Figure 1. Four-region framework

The two agreement regions are in white while the disagreement (negotiation breakdown) regions are shaded. The dotted lines represent unique agreement equilibria.

holds and bargaining powers have thus become incompatible.

The four-region framework can be applied to the general problem of conflict solution. This approach is a normative analysis that prescribes which class of policies and strategies could be successful in attaining an agreement. Taking the disagreement regions (I and III) as the two possible starting points of a negotiation process that has yet to arrive at an agreement, the analysis then attempts to determine what changes in the Λ and Y parameters are necessary to for a strategy to be successful.

Disagreement in region I. If the starting point is a stalemate in this region, the conflict can converge to either of the two agreement region targets but the combination of policies that will have to be applied will be different for each one.

Policy A: convergence from region I to region IV. To improve the chances of moving the conflict to agreement region IV, the policy adopted should combine two elements: (i) military strategies that maintain a high-intensity war (to keep Y above 0), and (ii) political strategies that reduce polarization (to reduce Λ below 1). This combination is represented by the downward-pointing arrow in Figure 2.

Policy B: convergence from region I to region II. To improve the chances of moving the conflict to agreement region II, the policy adopted should be a mix of: (i) military strategies that mitigate the destructive effects of the conflict, reducing it to a low-intensity war (to reduce Y below 0), and (ii) political strategies that maintain a high global polarization level (to keep Λ above 1). This mix is represented by the leftward-pointing arrow in Figure 2.

Disagreement in region III. If the starting point is a stalemate in this region, the conflict can again converge
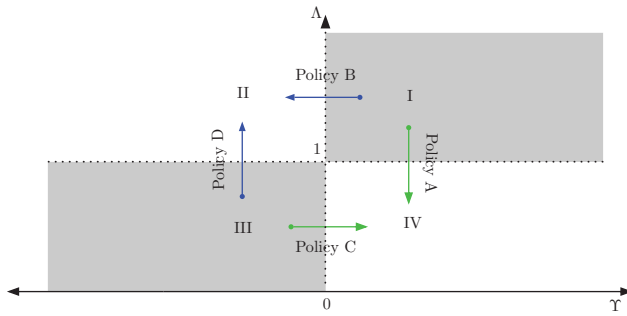
Figure 2. Succesful policies to move the conflict from a stalemate to an agreement region

Rightward and downward-pointing arrows (Policy C and Policy A) represent policies mixing a moderate rhetoric with active military actions. Leftward and upward-pointing arrows (Policy B and Policy D) represent policies combining a belligerent rhetoric with a permissive military strategy.

to either of the two agreement region targets, and here, too, the combination of policies that will have to be applied will be different for each one.

Policy C: convergence from region III to region IV. To improve the chances of moving the conflict to agreement region IV, the policy adopted should combine two elements: (i) military strategies that increase the intensity of the conflict so that it becomes a very destructive war (to increase Y above 0), and (ii) political strategies that maintain a moderate level of polarization (to keep Λ below 1). This policy combination is represented by the rightward-pointing arrow in Figure 2.

Policy D: convergence from region III to region II. To improve the chances of moving the conflict to region II, the policy adopted should be a mix of: (i) military strategies that maintain the conflict at a low intensity level (to keep Y below 0), and (ii) political strategies that strongly polarize the conflict (to raise Λ above 1). This mix is represented by the upward-pointing arrow in Figure 2.

The foregoing suggests that a successful policy must always combine military and political strategies in a *dissociated* manner such that 'words' and 'facts' are in some sense contradictory. Otherwise, with rhetoric and military actions working in the same direction the conflict may simply evolve from one type of disagreement to another. In colloquial terms, and making an analogy between a polarized conflict and a dogfight, we can group and describe the four above policies as follows.

> *Policies A and C:* Move the conflict towards a fight in which the dogs bite more than they bark.
> *Policies B and D:* Move the conflict towards a fight in which the dogs bark more than they bite.

Note that from an ethical viewpoint, policies B and D may be superior in that under them, the costs of the conflict in terms of human casualties may turn out to be lower than under policies A and C. This should be true in general since the latter two policies involve maintaining a high-intensity war whereas the former two entail more passive military strategies that lead the disagreement scenario to a relatively low-intensity armed conflict.

Nevertheless, if policies B and D were, in fact, adopted by the negotiators, they might be difficult to sustain as they would likely strike the parties they represent as contradictory and barely credible. For instance, policy B would imply maintaining a highly belligerent rhetoric in order to sustain a high level of polarization. It seems reasonable to suspect that menacing words by the negotiators in this type of communication strategy would quickly become *empty threats* if peace talks were simultaneously being conducted. Perhaps the only real chance of success for this policy would be to publicly break off the peace process while in reality continuing it in the form of private discussions.

Application to the Colombian conflict

We now apply our four-region framework to the particular problem of solving the Colombian conflict. A descriptive analysis is presented examining why two different plans adopted in the past, previous to the current peace process, failed to result in an agreement and why the current peace process has so far been only partially successful.

Since unlike the previous section we are now studying a real conflict, we must first determine for each of the earlier plans the actual starting region. We then identify the changes that subsequently occurred in each case in the Λ and Y parameters which led the conflict to converge towards one or other of the agreement/disagreement regions.

The previous negotiation initiatives

The two earlier plans pursued by Colombian government in the hope of solving the conflict were (i) the Demilitarized Zone (DZ), and (ii) Democratic Security (DS). In each case, we examine the reasons behind its failure.

Demilitarized Zone. This plan was part of a round of peace talks between the President Pastrana's administration (1998–2002) and the FARC held in the late 1990s. The guerrilla group demanded as a condition the

creation of a demilitarized zone (DZ) without a ceasefire. The government accepted the condition, creating a DZ covering about 42,000 km^2.

At the time of the plan, the conflict had already become a high-intensity war with both sides displaying great destructive power. It had also reached a very high level of polarization, mainly because (i) FARC had intensified the use of very unpopular practices such as kidnapping, extortion and drug trafficking, and (ii) paramilitary groups had entered the conflict and were employing similar tactics (Ortiz, 2002). We thus conjecture that before implementation of DZ, the conflict was in region I.

Our analysis is that the DZ policy failed because it was unable to dissociate rhetoric from facts. The facts were that the strategy brought about a lowering of the intensity of the conflict reflected in a reduction of hostilities on both sides, allowing both parties to increase their gross conflict payoffs V_i. This implied a decrease in the value of parameter Y below 0. As for rhetoric, under DZ the government first and then the FARC toned down their language, which initially led to considerable optimism that a peaceful solution was possible. The result was a decline in the hatred levels, especially on the side of the civil population, which moved the conflict from a high level of polarization to a more moderate level. This meant a decrease in parameter Λ to a value below 1.

Despite this optimism and the agreement by the two sides in mid-1999 on a complicated agenda containing 47 points, none of them were ever discussed. The negotiations focused on achieving a ceasefire and a prisoner exchange, but only on the latter was a narrow and partial agreement arrived at. The process finally collapsed in 2002 due to the lack of progress. In the context of our framework, this breakdown is explained by the joint decrease of both the Y and Λ parameters, which moved the dispute from region I to region III. Thus, according to this analysis the DZ plan merely swapped one type of disagreement for another, a situation represented in Figure 3 by a downward-pointing arrow.

Democratic Security. This policy, adopted by the Colombian government under President Uribe (2002–10), was aimed at ending the conflict by military means.[8] We begin our analysis of this military strategy with a
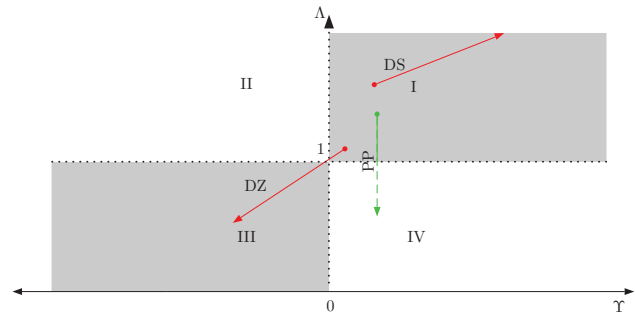

Figure 3. Past and current attempts to reach an agreement in the Colombian conflict

The Downward and upward-sloping arrows represent two failed policies: Demilitarized Zone (downward-sloping arrow) and Democratic Security (upward-sloping arrow). The downward-pointing arrow represents the current Peace Process with an active military strategy.

discussion of what we believe was its starting point. As explained above, in the late 1990s the DZ plan allowed the FARC to increase its military capacity, which was used by the group not only to intensify the violence of its actions but also to extend them to drug trafficking and other practices. Thus, we may argue that by the early 2000s the internal war was again at a high intensity level, in which case the conflict cost returned to a level such that $Y > 0$. The plan's failure then also returned the conflict to a high level of polarization – high enough that $\Lambda > 1$. This conclusion is based on the fact that the civilian population felt the guerrillas had betrayed the good faith shown by the government in the DZ peace process by taking advantage of it to bolster their finances and their military capabilities. The higher degree of violence of the FARC's new practices provoked a significant increase in the level of hatred felt by the civil population. The population used the 2002 election to penalize the government then in power (its delegated negotiator) for the failure of DZ by transferring to the government this higher hatred level. In that election, the centrist coalition that had initiated DZ did not even run and a new right-wing coalition proposing a military strategy to defeat the guerrillas (the DS policy) was voted in with a strong majority.

This political shift is an additional element that may have affected the global polarization level of the conflict in that it probably exacerbated the ideological differences between the two sides, confronting the ultra-left guerrilla group with a more conservative administration. The change can be expressed in our model as an increase in the FARC's λ parameter, which is consistent with a higher hatred level in the group for each unit of the surplus that ends up in the hands of the government.

---

[8] The Uribe administration also passed the Justice and Peace Law (Congress of Columbia, 2005) under which some paramilitary group members and a few guerrilla fighters turned themselves in to the authorities in return for a promise of relatively light sentences, judicial pardons and financial rewards.

Such a situation would reflect a belief on the part of the group's members that a right-wing government would be less likely to redistribute the surplus in the form of land or other resources among the rural poor but rather would divert it to the wealthier sectors of the population.

We therefore conjecture that the failure of the earlier DZ plan led to an increase in both parameters Y and Λ and thus brought the conflict back from region III to region I, the latter then becoming the starting point for the later DS policy.

Regarding the DS itself, we would argue that it failed because it was not able to properly dissociate rhetoric from facts. The facts were that in the first stage of the government's strategy much damage was done to FARC's military capacity, resulting in serious losses for the group both economically and politically (its main leaders ending up either captured or dead). But in the second stage, the guerrillas were able to reorganize militarily and inflict heavy losses both in infrastructure and human life. On balance, we conclude that DS further increased the conflict's military intensity, implying that both parties either held down or lowered their respective gross conflict payoffs V_i. In the context of our model, this means that the Y parameter remained positive and even increased. As for the rhetoric, DS was accompanied by a more aggressive discourse, especially that of the Uribe government which insisted that the conflict would only end when the guerrillas were defeated militarily and refused to consider any further negotiations unless the guerrillas first laid down their arms. This belligerent approach was ratified in some sense when Uribe won a second presidential mandate, thereby exacerbating both the ideological divide separating the government and FARC's leaders and the hatred levels of both Colombian society and the guerrilla group membership. In that sense, there is evidence that at this point ideology continued to play a role in the internal dynamic and cohesion of FARC and the other groups then active in Colombia, constituting another motivation to fight in addition to the individual economic motives (Ugarriza & Craig, 2012). In terms of our model, all of this implies that DS exacerbated the conflict polarization even further, keeping it at a level where parameter Λ was greater than 1.

The joint increase in both Y and Λ meant that the conflict remained in region I, receding even further from a possible solution. In terms of our dogfight analogy, the failure of DS can be described as a policy that strengthened the conflict as *a fight in which the dogs bark and bite*. This situation is represented in Figure 3 by an upward-pointing arrow in region I.

The current peace process

President Santos (2010–18), the successor to Uribe, partially maintained the DS military policy but also initiated a peace process (PP) with FARC in Havana (2012–16). FARC agreed to take part in the process but without surrendering its arms until all of the points to be negotiated had been implemented. This dual approach on the part of the two sides meant in practice the coexistence of two apparently contradictory strategies: (i) an attempt to reach an agreement on a peaceful solution of the conflict, and (ii) a military strategy that, if successful, would condition the PP dynamic, and if not, would allow the two parties to maximize the benefits or minimize the losses of an insoluble armed conflict.

During much of the PP negotiation, the process enjoyed the support of the Colombian people. This positive atmosphere surrounding the negotiations drew strength from the re-election of Santos to a second term (2014–18). In this favourable context, both sides finally arrived at the August 2016 Havana agreement, which resulted in a definitive bilateral ceasefire. Although the accord was ratified by the guerrilla group at a conference of FARC delegates, it was rejected by Colombians in a nationwide referendum held in September 2016.

In the context of the proposed model, our assessment is that contrary to previous peace initiatives, the preliminary success of PP can be explained mainly by the apparently contradictory nature of the dual approach pursued by FARC and the Colombian government during the final few years of the open conflict. This dual approach consisted of the right *dissociative* combination of peace talks ('words') with an active military strategy ('facts') on the part of both parties. Regarding the facts, the military strategy implied keeping the intensity level of the conflict high and total gross conflict payoff $(V_A + V_B)$ therefore low. This in turn meant maintaining Y > 0. As for the rhetoric, the intrinsically conciliatory nature of a peace process forced the negotiators on both sides (the Colombian government and the FARC representatives) to adopt a softer tone, as more belligerent language would have damaged their credibility in the eyes of their respective constituents. This effort in rhetoric was intended to reduce the levels of hatred felt by the two parties represented in the talks (Colombian society and the FARC membership) and thereby lower conflict polarization to a relatively moderate level $\Lambda < 1$.

It would appear that by 2016 the negotiators for the two sides felt they had arrived at the right mix of words and facts that would keep Y positive and Λ down so that the combination could lead to a sustainable agreement. In terms of our model, the PP therefore had the potential

to shift the conflict from the disagreement region I to the agreement region IV. Once again in terms of our two-dogs analogy, the peace process together with the two parties' military strategies was seen at this point as a policy combination capable of moving the conflict from *a fight in which the dogs bark and bite* (region I) to *a fight in which the dogs bite more than they bark* (region IV). This dynamic is represented by the downward-pointing arrow in Figure 3.

As was already noted, however, the preliminary agreement was not ratified by the Colombian population in the September 2016 referendum. Our analysis is that this rejection occurred because the negotiators *underestimated* the hatred level of the median voter and therefore also the conflict polarization level that indeed was manifested in the referendum. In terms of our framework, the efforts in rhetoric were insufficient to reduce the polarization Λ to a level below 1, and thus, the conflict continued in the stalemate region I. This implies that the government as negotiator was willing to give a share of the surplus to FARC that was *larger* than what the majority of referendum voters was prepared to tolerate.

This conclusion is backed by two pieces of evidence. First, in the run-up to the referendum all the polls predicted the Havana agreement would be approved. In terms of our model, this suggests that the government estimated the hatred level λ of the median voter – and hence also the global conflict polarization Λ – to be sufficiently low that voters would support the agreement. But the low overall referendum turnout coupled with higher participation in the most polarized sectors (see the detailed empirical evidence discussed below) suggests that the polarization level reflected in the referendum outcome was higher than what the poll results had indicated.

The second piece of evidence concerns the terms of the preliminary agreement and the benefits they conferred on the FARC. The Havana talks embraced a diverse range of issues grouped into five major items: (i) agricultural development, (ii) political participation, (iii) solution of the drug problem, (iv) justice, truth and reparations, and (v) the end of the conflict.[9] According to the agreement, the FARC would receive significant benefits such as a special status for the guerrilla leaders amounting to a partial amnesty, the right for the leaders to participate in politics and implicit sentence reductions for crimes committed such as kidnapping and drug

trafficking. We may reasonably suspect that the sectors of Colombian society most opposed to the peace process considered these benefits to be excessive concessions by the government, prompting them to turn out for the referendum in greater numbers than those with more moderate views, thus ensuring the agreement was rejected.

The available data on the referendum results broken down geographically are consistent with our conclusions. The 'Yes' vote (i.e. ratifying the agreement) was concentrated in municipalities with the highest rates of conflict-related violence, which were also the most rural ones located furthest from the main urban centres (Herreño & Muñoz, 2016).[10] Interestingly, these same municipalities were those that had voted most heavily to re-elect President Santos in 2014 (Fergusson & Molina, 2016). In terms of our model we interpret these data to mean that the municipalities most likely to ratify the agreement were those that displayed the following two characteristics: (i) higher conflict cost (because of lower value of conflict payoff V), reflecting greater negative human and economic impacts of the guerrilla war and therefore also greater potential benefits of a peace agreement (Herreño & Muñoz, 2016), and (ii) lower hatred levels (lower λ) given that despite the higher direct conflict cost, their stronger support for Santos implied stronger support for his mixed strategy peace process to end the conflict.

As for the 'No' vote, it was concentrated in municipalities where government authority and business activity are strongest (Herreño & Muñoz, 2016). These were the areas where the re-election of Santos was most closely contested. In terms of our model, these data can be interpreted to mean that the municipalities most likely to reject the agreement were those that displayed the following two characteristics: (i) lower conflict cost (due to higher conflict payoff V), reflecting less negative human and economic impacts of the guerrilla war due to their higher degree of urbanization and greater distance from conflict zones, and (ii) higher hatred levels (higher λ) given that despite the lower direct conflict cost, the weaker support for Santos's re-election implied less support for his peace process strategy. The greater levels of hatred were likely due to the fact that these municipalities would have borne a higher share of the

---

[9] See the Online appendix for more details on these five items in the Havana agreement.

[10] The Yes vote won in the 86% of municipalities with the highest multidimensional poverty rates and the 67% with the greatest inequality levels as measured by the Gini coefficient (Herreño & Muñoz, 2016).

tax burden to finance the terms of the agreement (Herreńo & Muńoz, 2016).

The referendum results also suggest that sectors apparently exhibiting higher hatred levels were also those where the mobilization of support for the No vote was greatest. Turnout across the country was lower than the polls had predicted (Herreńo & Muńoz, 2016; Fergusson & Molina, 2016; CERAC, 2017) but the decline in participation was most significant in areas where a high Yes vote was expected (Fergusson & Molina, 2016). Also, according to various analysts the better campaign run by those opposed to the agreement and their greater awareness of the terms of the agreement together explain the No side's better turnout and thus their referendum victory, albeit by a very slim margin (50.23% vs. 49.76%) (García, 2016).

In addition, we may conjecture that the Santos government proved unable to carry out an effective referendum campaign strategy. For example, the president's public statement to the effect that if the agreement was not ratified the conflict would shift from the countryside to the cities could have been seen as a provocation by the largely urban opponents of the agreement, thus exacerbating the polarization level and reinforcing these voters' determination to go to the polls. Also, Santos did little to get the resources of the government bureaucracy and the pro-government political parties behind the effort to promote a high turnout among Yes voters (Fergusson & Molina, 2016; Botero, 2017).

We end our analysis of the PP with an illustration of its dynamic using the following numerical example with our theoretical model. Assume that from the median voter's perspective, the negotiators believed they faced a bargaining process in which the surplus if peace was achieved was $\pi = 1$, the gross conflict payoffs were $V_A = V_B = -0.1$ and the hatred level was $\lambda_A = 0.9$ among the general Colombian population and $\lambda_B = 1$ among the FARC membership. Based on these parameter values, assume further that the preliminary agreement was settled at $(x^*_A, x^*_B) = (\underline{x_A}, \overline{x_B}) = (0.46842, 0.53158)$.[11] This implies that the government believed voters would accept the grant of a share $\overline{x_B} = 0.53158$ to FARC. In fact, however, the majority of those who voted considered this to be too great a concession. This was so because the social sectors most opposed to the peace process were those with the highest turnout rates, bringing the hatred level of the median

voter up above what the government had estimated. Let us say, for example, that this median was $\lambda'_A = 1.1$. Since in this numerical example $Y = 1.2 > 0$, $\underline{x_A}$ is then increasing in λ_A (see footnote 6). Thus, the underestimate of λ_A also led to the government's underestimate of the minimum share the median voter of the referendum was willing to accept from the agreement, that is, there existed an actual $\underline{x'_A} = 0.52857 > \underline{x_A} = x^*_A$. This would explain why the majority of voters did not ratify the preliminary agreement. Alternatively, since $\overline{x_B} = \pi - \underline{x_A}$ is decreasing in λ_A, on the previous analysis the underestimate of λ_A would have led the government to overestimate the maximum share that the median voter would tolerate going to FARC. In other words, there existed an actual $\overline{x_B}' = \pi - \underline{x'_A} = 0.47143 < \overline{x_B} = x^*_B$. On this alternative analysis the preliminary agreement was rejected because it gave a bigger share to FARC than the maximum the majority of referendum voters was willing to concede (i.e. 0.53158 vs. 0.47143).

In terms of our four-region framework, this numerical example suggests that despite the efforts in rhetoric to reduce the polarization level, when the preliminary Havana agreement was reached by negotiators the conflict was not in region IV where $Y = 1.2 > 0$ and $\Lambda = 0.9 < 1$, but rather still in region I where $Y = 1.2 > 0$ and $\Lambda' = 1.1 > 1$.[12] As a result, the agreement could not be ratified and the conflict remained in a stalemate region. This situation is represented in Figure 3 by the dashed segment of the downward-pointing arrow, illustrating the idea that the negotiators (especially the government) underestimated the actual polarization level of the conflict.

Concluding remarks

This article proposed a model of delegated distributive negotiation for resolving conflicts in which the parties to it feel hatred for each other. Using this sentiment-type externality, a measure of conflict polarization was constructed from the penalties imposed by the parties on their respective delegated negotiators when either of them makes concessions to their counterpart. We showed that incorporating polarization adds an

---

[11] This is just one of a continuum of agreements that are possible according to Proposition 3.

[12] The example assumes that negotiators estimated correctly the size of the pie to be divided and the gross conflict payoffs. The actual conflict cost was therefore $Y = 1 - (-0.1) - (-0.1) = 1.2$. However, whereas the negotiators estimated a conflict polarization level of $\Lambda = 0.9 \times 1 = 0.9$, the actual level of polarization manifested in the referendum was $\Lambda' = 1.1 \times 1 = 1.1$.

important dimension to the characterization of bargaining processes and their solution, resulting in a more elaborate approach than one based solely on the conventional cost–benefit trade-off between maintaining or ending a conflict.

This richer analysis predicts that in a polarized dispute, four regions may emerge, two of agreement and two of disagreement, depending non-monotonically on the polarization level and the cost of the conflict. The non-monotonicity property implies that a prerequisite for achieving a negotiated solution is the implementation, ideally by both parties, of contradictory mixed negotiation policies combining rhetoric aimed at modifying hatred and polarization levels with military strategies for modifying the costs of an ongoing conflict. Policies that do not so dissociate rhetoric with military strategy are doomed to failure and will keep the conflict going without a definitive peaceful resolution.

This four-region framework was applied to the case of the long-running internal conflict between the government of Colombia and the FARC guerrilla group. It was shown that past attempts to find a solution to the conflict such as the Demilitarized Zone plan and the Democratic Security policy failed because they were unable to separate rhetoric from military action.

By contrast, the same analysis suggests that the current peace process has so far been partially successful because the agreement reached in Havana was the result of conversations that were properly counterbalanced by active military strategies. However, the negative outcome of the subsequent referendum implies that the negotiators underestimated the hatred level of the median voter. In fact, the significant benefits obtained by the FARC in the Havana agreement were considered by the sectors of Colombian society most contrary to the peace process as unacceptable government concessions, prompting them to turn out for the referendum in larger numbers than the more moderate sectors.

Supplementary material

The Online appendix can be found at http://www.prio.org/jpr/datasets.

Acknowledgments

ORCID iD

Sigifredo Laengle https://orcid.org/0000-0002-8396-4408
Gino Loyola https://orcid.org/0000-0001-9982-4912
David Tobón-Orozco https://orcid.org/0000-0003-4519-5617

References

Anderlini, Luca & Leonardo Felli (2001) Costly bargaining and renegotiation. *Econometrica* 69(2): 377–411.

Botero, Sandra (2017) El plebiscito y los desafíos políticos de consolidar la paz negociada en Colombia [The plebiscite and the political challenges of consolidating Colombia's negotiated peace]. *Revista de Ciencia Política* 37(2): 369–388.

CERAC (2017) Un conflicto largo, de terminación negociada [A long conflict, negotiated termination]. CERAC 30 June: 1.

Congress of Colombia (2005) The Justice and Peace Law (Law 975).

Crawford, Vincent (1982) A theory of disagreement in bargaining. *Econometrica* 50(3): 607–637.

Esteban, Joan-Maria & Debraj Ray (1994) On the measurement of polarization. *Econometrica* 87: 379–415.

Esteban, Joan-Maria & Debraj Ray (1999) Conflict and distribution. *Journal of Economic Theory* 87(2): 379–415.

Esteban, Joan-Maria & Debraj Ray (2008) Polarization, fractionalization and conflict. *Journal of Peace Research* 45(2): 163–182.

Esteban, Joan-Maria & Debraj Ray (2011) Linking conflict to inequality and polarization. *American Economic Review* 101(4): 1345–1374.

Esteban, Joan-Maria & Gerald Schneider (2008) Polarization and conflict: Theoretical and empirical issues. *Journal of Peace Research* 45(2): 131–141.

Fergusson, Leopoldo & Carlos Molina (2016) Un vistazo a los resultados del plebiscito [A view to plebiscite results]. *La Silla Vacía* 6 December: 1.

Fernandez, Raquel & Jacob Glazer (1989) Striking for a bargain between two completely informed agents. *American Economic Review* 81(1): 240–252.

García, Miguel (2016) Entendiendo el 'No' al acuerdo de paz en Colombia [Understanding the 'No' to the Colombian peace agreement]. *Revista UNO* 27: 31–32.

Gorbaneff, Yury & Flavio Jacome (2000) El conflicto armado en Colombia: Una aproximación desde la teoría de juegos [The Colombian armed conflict: A game theory approach]. *Planeación y Desarrollo* 3–4: 193–208.

Haller, Hans & Steinar Holden (1990) A letter to the editor on wage bargaining. *Journal of Economic Theory* 52(1): 232–236.

Herreño, Juan & Juan Muñoz (2016) País divergente: radiografía del plebiscito [Divergent country: Plebiscite x-ray]. *El Pulso* 5 December (https://www.pulzo.com/nacion/resultados-plebiscito/PP134639).

Hirshleifer, Jack (1991) The technology of conflict as an economic activity. *American Economic Review* 81(2): 130–134.

Kirchsteiger, Georg (1994) The role of envy in ultimatum games. *Journal of Economic Behavior & Organization* 25(3): 373–389.

Montalvo, José & Marta Reynal-Querol (2005) Ethnic polarization, potential conflict, and civil wars. *American Economic Review* 95(3): 796–816.

Muthoo, Abbinay (1996) A bargaining model based on the commitment tactic. *Journal of Economic Theory* 69(1): 134–152.

Myerson, Roger (1991) *Game Theory: Analysis of Conflict.* Cambridge, MA: Harvard University Press.

Nash, John (1953) Two-person cooperative games. *Econometrica* 21(1): 128–140.

Ortiz, Román (2002) Insurgent strategies in the post-Cold War: The case of the revolutionary armed forces of Colombia. *Studies in Conflict and Terrorism* 25(2): 127–143.

Reynal-Querol, Marta (2002) Ethnicity, political systems, and civil wars. *Journal of Conflict Resolution* 46(1): 29–54.

Salazar, Boris & Pilar Castillo (2001) La hora de los dinosaurios: conflicto y depredación en Colombia [Dinosaur time: Conflict and predation in Colombia]. CIDSE 1, Universidad del Valle.

Skaperdas, Stergios (1992) Cooperation, conflict, and power in the absence of property rights. *American Economic Review* 82(4): 720–739.

Ugarriza, Juan & Matthew Craig (2012) The relevance of ideology to contemporary armed conflicts: A quantitative analysis of former combatants in Colombia. *Journal of Conflict Resolution* 57(3): 445–477.

Varian, Hal (1974) Equity, envy, and efficiency. *Journal of Economic Theory* 9(1): 63–91.

Zambrano, Andres & Hernando Zuleta (2017) Goal and strategies of an insurgent group: Violent and non-violent actions. *Peace Economics, Peace Science and Public Policy* 23(2).

Zuleta, Hernando; Juanita Villaveces & Veneta Andonova (2013) Conflict and negotiation in Colombia: Are pre-donations useful? *Journal of Socio-Economics* 47: 105–117.

GINO LOYOLA, b. 1971, PhD in Economics (Carlos III University of Madrid, 2008); Associate Professor, University of Chile (2008–); current main interest: bargaining and polarized conflicts.

SIGIFREDO LAENGLE, b. 1962, Dr. rer. pol. (University of Konstanz, 2000); Associate Professor, University of Chile (2006–); current main interest: optimization, information theory and decision theory.

DAVID TOBÓN-OROZCO, b. 1972, MA in Economics (University of Chile, 1999); Associate Professor, University of Antioquia (2000–); current main interest: applied microeconomics.