Divide and Rule: An Origin of Polarization and Conflict*

Simon Alder† Yikai Wang‡

December 31, 2018
(link to latest version)

Abstract

We propose a theory of civil conflict where political elites strategically initiate conflicts in order to polarize society and sustain their own power. We provide a micro-foundation for this divide-and-rule strategy by modeling polarization as a lack of trust among groups in the population. Trust is shaped by economic interactions between members of different groups who observe past interactions and update their belief about the relationship with the other group. High trust in turn increases the expected gains from trade and therefore the common interest of people to trade without being taxed by the elite. This leads to a revolutionary threat to which the elite responds by starting a conflict to interrupt trade and prevent trust from increasing further.

JEL-Classification: D74, D83, F15, H30, N40
Keywords: Polarization, divide and rule, conflict, ethnic violence, autocracy, trade, trust

*We would like to thank Simeon D. Alder, Josef Falkinger, Dominic Rohner, Viktor Tsyrennikov, Fabrizio Zilibotti, Josef Zweimüller, and participants at the Annual Meeting of the Society for Economic Dynamics, the University of Oslo, and the University of Zurich for valuable comments. Caroline Chouard and Sebastian Ottinger provided excellent research assistance. Financial support from the European Research Council (ERC Advanced Grant IPCDP-229883) is gratefully acknowledged.

†Department of Economics, University of North Carolina at Chapel Hill, salder@email.unc.edu.
‡Department of Economics, University of Oslo, yikai.wang@econ.uio.no.

1 Introduction

Conflict causes enormous costs, and yet we often observe wide support in the population for political elites that initiate conflicts. Several studies have shown theoretically and empirically that polarization is an important determinant of the incidence of conflict (see for example Montalvo and Reynal-Querol, 2005; Esteban and Ray, 2011; Esteban, Mayoral, and Ray, 2012). These studies provide important evidence on the causes of conflict by studying exogenous aspects of polarization. In this paper, we argue that there may be an important endogenous element and that it is useful to study the potential origins of polarization. We endogenize polarization by modeling the role of political elites and how they strategically create conflict and polarization in order to sustain their own power.

For the vast majority of the population, the consequences of conflict can be catastrophic. Why is it that some societies are so highly polarized that conflict erupts on a large scale? Predetermined differences are unlikely to be the sole explanation and the same groups have often lived in peace for long periods of time. This is particularly clear in the case of ethnic conflict, which is the most common form of civil conflict[1]. For example, the genocide in Rwanda in 1994 was based on ethnic labels, but there is no evidence of mass political violence between Hutu and Tutsi before the revolution that started in 1959 (Gourevitch, 1998) and the ethnic labels did not even have a political meaning until the mid 19th century (Pottier, 2002). In contrast, there are numerous examples in Rwanda and elsewhere showing that conflict is a *strategic choice* by a political elite that created polarization. Fearon and Laitin (2000: 846) emphasize this in their review of conflict studies: "If there is a dominant or most common narrative in the texts under review, it is that large-scale ethnic violence is provoked by elites seeking to gain, maintain, or increase their hold on political power." By initiating conflict along ethnic lines, elites can deepen ethnic divisions and thus increase polarization (Horowitz, 2000). Such polarization based on ethnicity is an important and well-documented case, but our argument applies more broadly to heterogeneous groups within a country. The key is that polarization

[1]Fearon and Laitin (2003) code 51 percent of conflicts in their sample as primarily ethnic and 18 percent as mixed. Esteban and Ray (2008) emphasize that groups in conflict are often horizontally differentiated along ethnic or religious lines, as opposed to vertical differentiation by economic class.

strengthens the position of the elite and allows it to exploit power. This strategy is well known as "divide-and-rule".[2]

We propose a model of divide-and-rule where the polarization between groups is endogenous and a result of strategic choices by the political elite. Polarization is modeled as a lack of trust between groups. High levels of trust threaten the autocratic elite because it may be overthrown if trust among the different groups in society is high and the common interest of the people is large enough to bear the cost of a revolution. The elite therefore has an incentive to manipulate trust and create polarization to avoid the revolutionary threat.

A key question for such a divide-and-rule argument is how an elite can affect polarization among groups of people. Fearon and Laitin (2000: 853) describe this challenge as follows: "The puzzle for such theoretical arguments is to explain how elites can convince their followers to adopt false beliefs and take actions that the followers would not want to take if they understood what the leaders were up to." Our contribution is to provide a micro-foundation for how elites can polarize society when people are rational. This is achieved by embedding a model of trade and trust (Rohner, Thoenig, and Zilibotti, 2013a) into a political economy framework that allows for conflict and revolution (Besley and Persson, 2011).

The model society has two groups of equal size, one of which initially being the incumbent group. Within each group, there is a small political elite that derives rents from being in power. We start from a situation of autocracy where the elite of the incumbent group sets policies. The incumbent elite faces a threat from the other group (transition of power) and from the people of its own group (revolution). A revolution leads to a regime switch to democracy where people of both groups trade with each other without being expropriated by an elite. The threat of revolution therefore depends on the expected gains from trade with the other group.

Similar to Rohner, Thoenig, and Zilibotti (2013a), we model the expected outcome of trade as a function of trust between trade partners. More precisely, the outcome of trade is stochastic and the likelihood of a positive trade outcome depends on how mutually beneficial trade is (e.g. the complementarity between the trade partners), which is unknown. Agents hold a prior belief about how benefi-

[2]Posner, Spier, and Vermeule (2010) discuss divide-and-rule strategies from a theoretical perspective. See also Acemoglu, Robinson, and Verdier (2004).

cial trade is. A belief that trade is beneficial implies a high level of trust. Trust can increase through Bayesian updating of this belief after positive trade outcomes occur. However, trade can only take place during peaceful times, such that trust cannot emerge while there is war. The ruling elite can therefore prevent trust from emerging by starting a war. This enables the elite to limit the threat of revolution that originates in the common interest of people when trust is high.

The trade-offs faced by the people and by the elite are as follows. A currently ruling elite taxes people in both groups, but redistributes part of the revenues as transfers to the people of its own group. When trust is very low and therefore expected gains from trade small, then transfers are relatively more important for the people of the incumbent group than the gains from trading with the other group. They therefore prefer their own elite to stay in power. The probability that the currently ruling group stays in power is higher when it fights a war against the other group than when it is in peace. Elite and people of the incumbent group therefore both support war when trust is very low. When trust increases somewhat, the people start to prefer peace because they want to reap the gains from trade with the other group. But the elite still prefers war over peace because it is afraid of losing power to the other group. The people are less concerned than the elite about such a transition of power because they know that they may still have part of the trade surplus even if the other group takes over. In contrast, the elite loses everything permanently when there is a transition of power to the other group. This setting therefore generates that the ruling autocratic elite stays at war for longer than is optimal for the incumbent people, because the elite has more at stake from being in power.

At intermediate levels of trust, both elite and people of the incumbent group support peace and trade with the other group. The reason why the elite starts being in favor of trade as trust increases is that it can tax the gains from trade, which starts to outweigh the risk of losing power when trust increases. Both elite and incumbent people also prefer to maintain autocracy (as opposed to democracy and free trade among both groups), because it allows them to exploit the other group by taxing their endowments and gains from trade. The interest of elite and people of the incumbent group are therefore aligned again for these intermediate levels of trust.

At high levels of trust, the interests of the incumbent elite and the incumbent people diverge again: the income that the people get from the expected trade surplus becomes larger than the transfers that their own elite can credibly promise them. When trust is above a certain threshold, the people start a revolution and establish a peaceful democracy with the other group in order to reap the full benefits from trade without being taxed by an elite. The elite on the other hand loses all its rents when there is a revolution and cannot regain power.

In our dynamic model, we show that when trust is in an intermediate range and approaches the threshold where the revolutionary threat becomes binding, then the elite will start a war in order to stop trade and prevent trust from reaching the threshold. The incumbent people would prefer peace and trade, but since they cannot learn about the benefits from trade with the other group anymore, they will not start a revolution when the threshold is not reached. Despite rational expectations of the people, the elite can manipulate trust between different groups.

In summary, the key implication of the model is that the elite applies a divide-and-rule strategy to maintain its own power: when the threat of revolution is high, it starts a conflict between the groups, which harms trust, polarizes society, and limits the common interest of people. This strategy is against the interest of the people, since they prefer trade and may be better off under democracy where they are not exploited by an elite.

1.1 Related Literature

This paper is related to several strands in the existing literature. First, we relate to a large literature on civil conflict (Blattman and Miguel, 2010, provide a survey). Fearon and Laitin (2003) estimate that there were about 127 major civil wars between 1945 and 1999 with more than 16 million fatalities overall. Our analysis is particularly related to the strand in the civil war literature focusing on conflict between ethnic groups. Esteban and Ray (2011), Esteban, Mayoral, and Ray (2012), and Rohner (2011) show that the polarization of societies along ethnic lines is associated with high degrees of conflict. Esteban and Ray (2008) argue that coalitions along ethnic lines are more likely to emerge than along classes. Caselli and Coleman (2013) point out that certain ethnic characteristics that are easily observable (such

as skin color) allow to distinguish between winners and losers of a conflict and this constraint on intergroup mobility makes starting a conflict along these lines more profitable. Bhattacharya, Deb, and Kundu (2015) also model intergroup mobility and generate a non-monotonic relationship with conflict. Besley and Persson (2011) provide a framework where repression and civil war can depend on polarization between groups. We contribute to this literature by endogenizing ethnic polarization. In our framework, a ruling elite can strategically affect polarization between people in order to sustain its own power. Such patterns of "divide-and-rule" have been described also outside of the economic literature.[3] Related papers have argued that a political elite can expropriate its supporting group because of their fear that otherwise an even less favorable elite would take over (Figueiredo and Weingast, 1997; Padro i Miquel, 2007)

The findings on the salience of ethnic conflict are complemented by a set of theoretical papers on divide-and-rule strategies. Posner, Spier, and Vermeule (2010) analyze divide-and-rule strategies in two simple game theoretic models, the Prisoner's Dilemma and the Stag Hunt Game. They consider how an outsider (the elite) that is affected by the cooperation of the agents can influence outcomes.[4] They state that two conditions are important for such strategies: "(1) A unitary actor bargains with or competes against a set of multiple actors. (2) The unitary actor follows an intentional strategy of exploiting problems of coordination or collective action among the multiple actors." They consider different ways to apply a divide-and-rule strategy: destroying communication channels, payment of bribes, imposition of penalties, generating distrust, limiting interaction, and mixing agents with heterogeneous interests. Our model is best described by the category "generating distrust". However, Posner, Spier, and Vermeule (2010) implement this by the elite being able to directly manipulate agents beliefs about the other agents' private

[3]See for example the review of case studies by Fearon and Laitin (2000). Similarly, Horowitz (2000) writes "By appealing to electorates in ethnic terms, by making ethnic demands on government, and by bolstering the influence of ethnically chauvinist elements within each group, parties that begin by merely mirroring ethnic divisions help to deepen and extend them. Hence the oft-heard remark in such states that the politicians have created ethnic conflict."

[4]Posner, Spier, and Vermeule (2010) use the term "divide-and-conquer", which is equivalent to "divide-and-rule". The Oxford Dictionary (2014) defines both terms as "[t]he policy of maintaining control over one's subordinates or opponents by encouraging dissent between them, thereby preventing them from uniting in opposition."

payoffs. In contrast, we provide a setting where the elite cannot directly manipulate beliefs, but it can limit the learning process between agents by starting conflict. In this sense our setting also relates to the category "limiting interaction", although their use of this strategy refers to limiting the time horizon of repeated interactions.

Acemoglu, Robinson, and Verdier (2004) consider a divide-and-rule strategy that falls into the category described by Posner, Spier, and Vermeule (2010) as "payment of bribes". In Acemoglu, Robinson, and Verdier (2004), elites can exploit people because they can prevent them from cooperating in a revolution. The kleptocratic elite achieve this by bribing one of two groups in society in order to induce them to reject offers by the other group to cooperate in a revolution. They show that such kleptocratic regimes are more likely to arise if the regime faces fractionalized opponents instead of large players that can solve the coordination problem internally and thus put a constraint on the behavior of the ruler.

We model polarization as a lack of common interest, which in turn depends on trust between trade partners. This is based on Rohner, Thoenig, and Zilibotti (2013a) where trust and cooperation is shaped by Bayesian updating during interactions with the trade partner. Rohner, Thoenig, and Zilibotti (2013a) show that even accidental conflicts can be persistent and generate cycles of recurring conflict due to the destruction of trust. Such cycles are also a feature of Acemoglu, Ticchi, and Vindigni (2010), although without the link to trade. Rohner, Thoenig, and Zilibotti (2013b) provide empirical evidence on the effect of conflict on trust in Uganda. There is also a large literature on the measurement of trust and its correlates (see for example Glaeser et all, 2000; Alesina and La Ferrara, 2002). Zak and Knack (2001) provide a theory that can generate low-trust poverty traps. A related literature has analyzed social learning (for example Banerjee, 1992; Acemoglu and Wolitzky, 2014). Our contribution to this literature is to show that a conflict can be started strategically by the currently ruling elite in order to affect people's beliefs and to sustain its own power.

The channel through which in our framework polarization can be manipulated by a political elite is trade. We therefore relate to the literature on the relationship between conflict and trade – without the link to trust. The "liberal peace" argument postulates that increased economic interdependence through trade between countries reduces the likelihood of conflict. However, recent contributions find that

this relationship is more complex. Barbieri (1996) finds that strong economic interdependence is positively associated with militarized conflict. Martin, Mayer, and Thoenig (2008) show theoretically and empirically that increases in multilateral trade increases the risk of conflict, especially between neighboring states.[5] In our framework, an increase in trade may fail to prevent conflict for a different reason: a forward looking autocratic elite fears that high levels of economic interaction will augment the common interest of people because this increases the threat that the elite may be overthrown by the people.

A large literature studies social and political polarization in the United States (see for example Fiorina and Abrams, 2008, and Layman et al., 2006, for reviews). There is agreement that polarization between political parties has been increasing, but the findings regarding polarization in the public are mixed (see also the discussion in section 2.2). Mason (2015) argues that these diverging findings can be explained by an increase in social polarization instead of issue polarization. Social polarization means that members of different groups perceive themselves different from other groups based on social identities. We share with this literature the focus on polarization and how it can affect policy making, but we differ by allowing polarization to be endogenous and determined by the strategic actions of political elites. Bail et al. (2018) study whether the exposure to political views from the opposing party reduces polarization and find that it does not, and may even increase polarization. Our framework is consistent with this finding, because polarization is shaped not by the interaction between groups itself, but by the outcome of the interaction. Individuals update their belief about the benefits from cooperation after observing past outcomes, and this shapes trust and polarization.

The rest of the paper is organized as follows. In section 2, we discuss the anecdotal evidence on the divide-and-rule strategy. Section 3 presents a simple benchmark model to illustrate how the tradeoff between peace and war depends on trust. In section 4, the benchmark model is extended to a dynamic setting to study how trust evolves over time and how it can be manipulated by the elite. Finally, we conclude in section 5.

[5]The liberal peace argument has recently also been challenged by other authors, for example Goldsmith (2013) or Gowa and Hicks (2014).

2 Divide-And-Rule Strategies

Our model provides a micro-foundation for why and how an elite strategically polarizes society in order to sustain its own power. Such patterns have frequently been described in the literature (Fearon and Laitin, 2000). The evidence suggests that provoking conflict is a way of polarizing society. In this section, we discuss two types of polarization. Ethnic polarization, described in section 2.1, has been shown to be associated with conflict and there many examples of autocratic rulers who have strategically polarized society along ethnic lines in order to strengthen their power, thus following a divide-and-rule strategy. Social polarization describes the perceived differences between groups within a society based on social identity and is discussed in section 2.2. There is a growing literature documenting and analyzing social polarization in the U.S. (Mason, 2018). This form of polarization based on social identity is also closely related to our notion of polarization as a lack of trust and the literature on the U.S. shows that the model can also be applied to countries with democratic institutions.

2.1 Ethnic Polarization

Fearon and Laitin (2000) conclude from their review of studies of ethnic conflict, that conflict is often provoked by elites who seek power. In some cases, the vicious cycle of conflict and polarization goes as far as leading to genocide. Rwanda and former Yugoslavia are both examples where elites have used a divide-and-rule strategy by initiating conflict between ethnic groups with the result that polarization between the groups became so strong that large-scale genocide was possible.

Rwanda The history of Rwanda is an example of how ethnicity can be "constructed" by elites for their own benefit.[6] The ethnic cleavages between Hutu and Tutsi that led to the genocide in 1994 were not due to predetermined differences between these groups and the ethnic groups in Rwanda have not always been in

[6]See Fearon and Laitin (2000) for a comprehensive discussion of the "construction" of ethnic differences and the implications for violent conflict. Rogall (2014) provides evidence for the role of centrally planned militia interventions during the genocide in 1994.

violent conflict with each other.[7] Instead, the conflict between Hutu and Tutsi in the 1990s "had its origin in modern struggles for power and wealth" (Pottier, 2002).

Evidence on how the terms Hutu and Tutsi were initially used prior to colonization is scant. But it appears that before 1860 there was substantial social mobility and "ethnicity was not a principal organizing factor" (Pottier, 2002). From the mid 19th century on, it is known that the Tutsi king Rwabugiri started to polarize the Rwandan people with discriminating rules concerning the ownership of cattle (Newbury, 1988). Families rich in cattle were regarded as Tutsi while poorer families were labeled Hutu. Therefore, "[...] wealth, not race, was the basis of the ethnic distinction between Hutu and Tutsi" (Pottier, 2002: 14).

In her study of the impact on the region of Kinyaga, Newbury (1978) states that "social stratification" was among the most important transformation due to the new rule. With colonization, ethnic identity gained further relevance. The European colonizers used the Tutsi administration in order to control the country and even helped it in expanding its region of influence.Belgian colonizers actively divided society further. In 1933-34, they introduced identity cards that labeled individuals as either "Hutu", "Tutsi", or "Twa" (Gourevitch, 1998). This labeling drastically reduced the mobility between Hutu and Tutsi, which previously was relatively high.[8] The Belgium support of the Tutsi elite ended only shortly before independence when it started supporting the Hutu majority (Pottier, 2002).

According to Prunier (1995), the ethnic identity was a key aspect in the mobilization of Hutu peasants in the genocide that started in 1994. One channel through which elites could affect polarization was trust. Prior to the start of the genocide, Hutu extremists purposely fostered conflict in order to harm the interaction between moderate Hutu and Tutsi and to reduce trust. Their success with this strategy of limiting the relevance of moderate Hutus strengthened their position within the

[7] Hutu (Bantu) are the majority group in Rwanda with 84% of the population. Tutsi (Hamitic) account for 15% and Twa (Pygmy) for 1% (CIA World Factbook, 2014). There has been no systematic political violence between Hutu and Tutsi before the start of the Rwandan Revolution in 1959 (Gourevitch, 1998).

[8] Gourevitch (1998: 57) explains how ethnicity gained importance with the identity cards as follows: "The identity cards made it virtually impossible for Hutus to become Tutsis, and permitted the Belgians to perfect the administration of an apartheid system rooted in the myth of Tutsi superiority. [...] Whatever Hutu and Tutsi identity may have stood for in the precolonial state no longer mattered; the Belgians had made "ethnicity" the defining feature of Rwandan existence."

Hutu group. This allowed them to mobilize the masses for the genocide. Fearon and Laitin (2000) describe how moderate Hutus negotiated a cease-fire with Tutsi groups but Hutu extremists then initiated violent conflict that led to the collapse of the agreement and destroyed trust: "Unable to assign blame for the failure of the cease-fire with certainty, Hutu moderates increased their estimate that the RPF could not be trusted in political negotiations, exactly what the extremists had sought in their violent attacks" (Fearon and Laitin 2000: 865).

The genocide of Tutsi by Hutu militiamen, police, and army in 1994 killed more than 800,000 people. The conflict, including counter-killings undertaken by Tutsi extremists and the resulting large refugee flows, destabilized the region and contributed to the two Congo Wars, the deadliest conflicts since the Second World War.

Yugoslavia Although predetermined ethnic differences are sometimes used as an explanation for the eruption of violence in former Yugoslavia in the 1990s, it actually provides yet another example of how political elites *construct* ethnicity in order to divide people and sustain their own power. Referring to the Balkan conflict, Fearon and Laitin (2000: 867) state that violence was used strategically: "[. . . to] foment outrage among their own moderates, ethnic leaders will provoke inter-ethnic violence."

For example, Woodward (1995) describes how Slobodan Milŏsević polarized politics by emphasizing ethnic identity and accusing liberal parties in Serbia of being "too lenient on Albanians in Kosovo" Woodward (1995: 90). Slobodan Milŏsević strategically sought to polarize in order to gain power within his own group and by provoking fear of other groups, he could convince his own people that cooperation with them was not an option. The last US ambassador to Yugoslavia from 1989 to 1992, Warren Zimmermann also holds the view that the nationalism that emerged in Yugoslavia during this period was the product of elite's manipulation by way of triggering ethnic violence: "The breakup of Yugoslavia is a classic example of nationalism from the top down – a manipulated nationalism in a region where peace has historically prevailed more than war and in which a quarter of the population were in mixed marriages. The manipulators condoned and even provoked local ethnic violence in order to engender animosities that could then be magnified by the

press, leading to further violence." (Zimmermann 1995: 12). The region experienced a series of ethnic wars in the 1990s, including the genocide of Muslims in Srebrenica.

The descriptions above show that elites can strategically start ethnic conflict in order to affect trust and to divide societies. They may apply such a policy of divide-and-rule to strengthen their power. The incentives to do so depend on the rents from being in power and the gains from peaceful trade. The model in section 3 allows us to analyze this tradeoff and it highlights the dynamics of polarization and conflict. In particular, we will see that a divide-and-rule strategy by the elite can lead to a vicious cycles of conflict and polarization. The polarizing elites may not have wanted to end up in genocide, but the vicious cycle of conflict and polarization in combination with a series of negative stochastic shocks can have the consequence that the divide-and-rule strategy leads to a low-trust war trap. Before turning turning to the model, we will discuss another form of polarization that is not based on ethnicity, but social identity.

2.2 Social and Political Polarization

A large literature in political science has analyzed mass polarization in the public as well as polarization among the political elite, focusing on democratic settings and in particular on the US.[9] While there is agreement that polarization between parties and in the political elites has been increasing, the findings regarding polarization in the public are mixed. For example, Fiorina and Abrams (2008) do not find evidence of increasing mass polarization in the U.S., but Abramowitz and Saunders (2008) find that there has been a substantial increase since the 1970s.

Mason (2015) argues that the discrepancy between these two conclusions can be explained by distinguishing 'issue polarization' from 'social polarization'. There is evidence of a strong increase in social polarization, which refers to the extent to which members of different groups perceive themselves different from other groups based on social identities and is characterized by partisan bias. Issue polarization is based on disagreement on policy issues and has not changed to the same extent (Fiorina and Abrams, 2008). Mason (2015) therefore suggests that groups such as liberals and conservatives in the U.S. are becoming increasingly hostile to each other,

[9] See Fiorina and Abrams (2008) and Layman et al. (2006) for reviews.

and this trend is stronger than what would be predicted by their disagreement on policy issues.

According to Mason (2015), the reason for the increasing trend in social polarization is an underlying trend in partisan sorting, which results in groups that are more homogeneous within, but more hostile towards other groups.[10] This trend in partisan sorting affects social polarization more than issue polarization. Mason (2015) draws a pessimistic conclusion from these trends in social polarization and predicts that it harms a nation's ability to find political solutions: "It may therefore be disturbing to imagine a nation of people driven powerfully by team spirit, but less powerfully by a logical connection of issues to action. These results, however, demonstrate that as our political identities fall increasingly into alignment, and our partisanship consequently strengthens, the outcome is a nation that may agree on many things, but is bitterly divided nonetheless."(Mason, 2015: p. 142).

Our notion of polarization as a lack of trust between groups maps closely to the literature on social polarization that has found empirical support in the U.S. data. One crucial difference in our approach is that we investigate the origins of polarization by making it an endogenous outcome of agents' strategic actions. While social polarization as discussed above is driven by exogenous trends in partisan sorting, our model offers an explanation for why and how polarization changes over time. Specifically, we argue that polarization can be the result of a strategic choice by the elite, which uses a divide and rule strategy to strengthen its own power.

Another similarity between our framework and Mason (2015) is the emphasis on interactions between the groups. Highly polarized groups not only dislike each other, but also interact less with members of the other groups. This reinforces partisan isolation (Mason, 2018). In our model, agents engage in trade interactions with the other group and update their belief about the trustworthiness of the other group based on the trade outcome. Hence, in the absence of these interactions, trust cannot be formed. The low trust in turn reduces the expected gains from trade with the other group and can lead the members of the incumbent group to prefer conflict with the other group rather than trade. This is again similar to the

[10]See also Iyengar and Westwood (2015) on the evidence of increasing polarization in terms of social identity. Mason (2018) discusses the evidence from social psychology that has shown how individuals interact differently with members of the same group compared to other groups, even when the groups are artificial and arbitrary.

increased importance of winning against the other group that the political science literature has documented in the U.S. (Mason, 2018). A key feature of our model is that we do not need to assume that elites can directly manipulate people's believes or that they are irrational. Instead, the evolution of trust is driven by the learning process of rational agents and thus depends on the interactions between groups.

Why do political elites take actions that polarize society to the extent that its groups do not cooperate anymore to address important issues for the society, or that conflict erupts? Our theory not only predicts that elites may do so strategically to increase their own power, but also that this strategy depends crucially on the existing level of polarization. In particular, when the political leader of an incumbent group cannot rely on bipartisan support due to a high level of polarization, then she may seek to increase polarization further in order to gain more support within the own group. This interaction between underlying trends in polarization, politicians' actions, and their effect on polarization is an important aspect of political dynamics in democratic countries as well.

In particular since the 2016 election in the U.S., there has been an increased awareness of the consequences of polarization. This leads to the question what could potentially decrease polarization. One view in the literature is that an increased exposure to the views of the other group increases the understanding for their positions and reduces polarization. However, a recent study by Bail et al. (2018) found that randomly exposing Democratic and Republican voters to tweets by politicians from the other party does not decrease polarization, and may in fact increase it. While there are several reasons that could explain what the authors call 'backfire effect' from the perspective of our model it is clear that contact or exposure to the other group's views is not sufficient to reduce polarization, but that there must be an expected gain from cooperating with the other group. In our theory – and in other political economy frameworks such as Besley and Persson (2011) – polarization is a lack of a common interest, which in our setting is a lack of trust that cooperating with the other group would lead to gains from trade. An exposure through Twitter feeds may lead to a better understanding of the argument of the other group, but it is unlikely to change the expected benefits from cooperating with them.

3 The Benchmark Model

In this section, we build a benchmark model to illustrate why a political elite may want to influence polarization between groups in society, how they achieve this, and what the consequences are for conflict. Polarization is modeled as the opposite of trust between two groups. High polarization implies low trust and therefore less expected benefit from cooperation between these two groups. A high enough level of trust can give the two groups of people an incentive to revolt against the elite, so the elite may not like high trust and strategically creates a conflict to prevent trust from increasing to a level where a revolution may occur. The elite creates this conflict to affect the current level of trust, but this policy may lead to future increases in polarization and long-lasting conflicts. In Section 4, we study several extensions of the benchmark model and we discuss more general implications, not only for ethnic polarization and conflicts, but also for political polarization and polarization between two countries.

3.1 Environment

In the model, there are infinitely many periods, two political systems (autocracy and democracy), and two ethnic groups (A and B). First, we describe the environment in autocracy where an elite sets policy, which is the focus of this paper.

Autocracy In autocracy, in each period, one group is the incumbent, while the other is the opponent. Initially, group A is the incumbent. [11] The incumbent group consists of two types of agents: an elite and the people. The opponent group only contains the people in the current period. However, if in some later period, the current opponent group becomes the incumbent group, an elite from that group is born and the elite from the current incumbent group dies. The people in the incumbent and the opponent groups are denoted as I and O, respectively, while the elite in the incumbent group is simply denoted as E. The instantaneous utility function is linear in income. The elite discounts future incomes with the discount

[11] This setup is based on Besley and Persson (2011) and like them, we refer to the the two groups as ethnic groups, but the model can be interpreted more generally and used to discuss groups differing in other dimensions, for example, different political ideology.

factor β and maximization of the elite's lifetime utility is equivalent to maximization of its discounted sum of expected incomes. The people are hand-in-mouth and they only care about their current period income. In other words, they are myopic and their discount factor is 0.

There are two forms of production. One takes place within each group and can be thought of as home production. The second form of production requires cooperation of both groups and labeled as trade in this model. Outputs of incumbent and opponent people's home productions are y_I and y_O, respectively. If trade takes place, then each group gets an additional output of y_T, which is assumed to be the same for both groups, without loss of generality. The output y_T is a stochastic variable with two states. It can be "high": $y_T = y_H$, or "low": $y_T = y_L < y_H$. The probability of getting the high output depends on the type of relation between the two ethnic groups, which can be either "good" or "bad". A "good" type means that their cooperation is efficient and likely to generate a high output. More specifically, the probability of getting a high trade output is q_H, while the probability of getting a low trade output is $1 - q_H$. If the relation is of the "bad" type, implying that the cooperation is not as efficient, then the output is low with probability q_L and high with probability $1 - q_L$. We assume $q_H > 1 - q_L$, implying that if the relation is good, then the probability of getting a high trade output is higher than if the relation is bad. The interpretation of the two types of the relation is the following. The type of the relation, or equivalently, the likelihood of getting a high output from trade, depends on how complementary the characteristics of the two groups are, and on the communication capital that facilitates interactions. First, if two groups have complementary skills, for example, one group is good at producing intermediate goods that the other group needs, trade and cooperation between them are more likely to generate high outputs. Second, if the communication capital - e.g. roads connecting the two groups, or understandings of each other's language and culture - is high, then the gain from trade and cooperation between the two groups is more likely to be high. The type of relation is fixed in the benchmark model and allowed to change stochastically in an extension.

The agents don't know the type of relation for sure, but hold a common belief about the probability of the relation being good, denoted by p, which is updated given the whole history of realized trade outcomes. In this model, p can be inter-

preted as trust: the higher the trust (believing that with a higher probability the relation is good), the larger the expected trade surplus. Because both groups get the same gains from trade, high trust implies a larger common interest for the two groups. Therefore trust is interpreted as the opposite of polarization, since the latter implies conflicting interests between groups[12] and the belief that cooperation with the other group is not very beneficial.[13]

The events in autocracy are the following. At the beginning of each period, given trust p, I people choose whether to revolt against the ruling elite. If a revolution occurs, I people pay a revolution cost d^R and then the regime switches to democracy. Revolution also implies a cost of d_E^R for the elite[14] If there is no revolution in this period, then the elite of the incumbent group stays in power and makes the following decision: whether to have a conflict with the opponent group or stay in peace. If a conflict happens, then there cannot be any trade anymore. The conflict also leads to a further cost d^C for I people.[15] If peace is chosen, trade can happen.

After production has taken place, a political turnover may occur and the political power may switch from the incumbent group to the opponent group. If a turnover occurs, the previous opponent group becomes the new incumbent group and its elite takes over the government. With probability π^s, where $s \in \{C, P\}$ represents whether the two groups are in conflict or in peace, the previous incumbent group stays as the incumbent, while with probability $1-\pi^s$, the political turnover happens. Then the elite in the final incumbent group in this period taxes people - both the incumbent and opponent people - at the tax rate τ. The elite transfers fraction θ of total tax revenue to the incumbent people, and the remaining fraction $1-\theta$ is kept by the elite as its rent. Similar to Beasley and Persson (2011), we assume that the

[12]Besley and Persson (2011) also refer to low polarization as a higher common interest, for example when the two groups agree on what public goods to provide.

[13]In this model, higher trust generates higher expected output in trade, while the output itself, given the true type of relation, is not directly affected by the trust. A richer model can allow trust directly increasing output by inducing more cooperative behaviors, as in Rohner et. al. (2013a). The general implications are similar, so we focus on the simple setup.

[14]This setting implies that the incumbent people's choice of revolution is sufficient and necessary for democratization. It can be interpreted as assuming that when the incumbent people support revolution, the opponent people also support revolution. This assumption is without loss of generality, because the incumbent people benefit more in the autocracy than the opponent people, as we will see in the analysis below.

[15]We can also assume that there is a cost for O people, but it does not change the results so we simply assume it away.

tax rate τ is exogenously determined by the state capacity.[16] Similarly, transfer as a fraction of total tax income (θ) is also assumed to be exogenously determined.[17]

Democracy In democracy, the preference of the people and the two forms of production are the same as in autocracy, but there is no elite that runs the government and taxes people. In this case, the game is simple: the economy is in a Laissez-Faire competitive equilibrium and stays in it forever. In every period, the outputs from home productions and trade add up to people's incomes. There is no government tax or transfer in favor of any group. In other words, we assume away tyranny of the majority in democracy.[18] We can see that the definition of democracy in this model corresponds to full democracy in reality. In the real world, there are some partial democracies or even democracies where elites are still very influential in government policies and try to manipulate the political system, and in this model they would correspond to autocracy.

The timing of events starting from autocracy is described in detail below and also shown in Figure 3 in the Appendix.

1. At the beginning of each period, the level of trust p is given. I people decide to revolt or not.
    1. If revolution happens, the political system switches to democracy. Then, both groups of people live in the Laissez-Faire equilibrium forever.
    2. If there is no revolution, then autocracy survives and the game moves to step 2.
2. If autocracy survives, the elite chooses conflict or peace.
    1. If peace is chosen, I people can trade with O people.

[16] Tax policies are not the focus of this paper, so we simplify the tax decisions and focus on the political choices of the government. In Beasley and Persson (2011), tax rate τ is determined by the state capacity, which is predetermined and cannot be changed in the short-run.

[17] Beasley and Persson (2011) assume that θ is subject to a lower bound due to checks and balances. Notice that transfers happen after the revolution decision by the people. This implies that a promise by the elite to transfer more than the minimum amount given by checks and balances is not credible. Therefore, a promise of transfers cannot be used by the elite to prevent revolution. This is equivalent to simply assuming that θ is given.

[18] We rule out the case that people from one group, for example the group with a larger population, choose government policies and tax the other group.

2. If conflict is chosen, there is no trade.

3. Output is realized: home production yields y_I and y_O. If trade happens, nature draws the trade output from $\{y_H, y_L\}$, depending on the type of the relation between the two groups.

4. Political turnover: in autocracy, with probability $\pi^s, s \in \{C, P\}$, the existing incumbent group remains as the incumbent group, and with probability $1-\pi^s$, the other group becomes the new incumbent.

5. The current elite taxes the opponent people and makes transfers.[19]

6. Trust p' is updated. The game moves to the next period.

3.2 The Equilibrium

The equilibrium concept is sub-game perfect Nash equilibrium. We solve the model by backward induction. First, we calculate the payoffs in each period, i.e., the net incomes of the agents – the elite and the incumbent people – in all the states of the world, i.e., in conflict, peace, and revolution. Then we can discuss the choices of agents. We omit subscript t for time when there is no confusion.

3.2.1 I People

The expected output from trade given trust p is:

$$\begin{aligned} E[y_T] &= p(q_H y_H + (1-q_H) y_L) + (1-p)(q_L y_L + (1-q_L) y_H) \\ &= p(q_H + q_L - 1)(y_H - y_L) + (q_L y_L + (1-q_L) y_H). \end{aligned}$$

Recall that $q_H > 1-q_L$, implying $q_H + q_L - 1 > 0$, so $E[y_T]$ is positively related to p: the higher trust, the higher expected income from trade. To simplify the expression,

[19] Incumbent and opponent groups refer here to their status after the potential political turnover in the previous step.

we can normalize $E[y_T]$ to p by setting

$$\begin{aligned} y_H &= \frac{q_L}{q_H + q_L - 1}, \\ y_L &= -\frac{1 - q_L}{q_H + q_L - 1}, \end{aligned}$$

and then we get the following expression:

$$E[y_T] = p.$$

The positive relationship between the trade surplus and the belief of the other group is preserved, so this normalization is without loss of generality and does not change the results of the model.[20]

The net expected income of I people equals the sum of home production output, trade surplus and the transfer, minus the tax. I people get positive transfers if the political turnover does not happen and they thus remain incumbent. If a political turnover occurs, the previous I people become the new opponent group and get 0 in transfers. Then the expected incomes of the I people in all states of the world are the following:

- Peace: $y_I^P = (1-\tau)(y_I + p) + \pi^P \theta\tau(y_I + y_O + 2p)$.
- Conflict: $y_I^C = (1-\tau)y_I + \pi^C \theta\tau(y_I + y_O - d^C)$.
- Revolution: $y_I^R = y_I + p - d^R$.

Notice that here y_I^s denote the expected income given the political choices of I people and the elite, before the productions takes place. In other words, those are the expected incomes at the end of step 2 or in the beginning of step 3 in the timing of events discussed above. Remember that the people are myopic and care only about their own income. So knowing I people's current period income, we can characterize their preferences and choices.

First, we compare I people's income in conflict and peace. The potential tradeoff is the following: in peace, there is an additional output from trade, but the prob-

[20] This normalization is equivalent to multiplying y_I and y_O by $(q_H + q_L - 1)(y_H - y_L)$ and adding $q_L y_L + (1 - q_L) y_H$ to them if we don't set y_H and y_L.

ability of staying as the incumbent may be lower, if $\pi^P < \pi^C$. In this case, if and only if the expected gain from trade, or equivalently, trust p, is large enough, peace dominates conflict for I people. This argument can be formally shown as follows:

$$\begin{aligned} y_I^P &\geq y_I^C \Rightarrow \\ (1-\tau)(y_I+p)+\pi^P\theta\tau(y_I+y_O+2p) &\geq (1-\tau)\,y_I+\pi^C\theta\tau\left(y_I+y_O-d^C\right) \Rightarrow \\ p &\geq \frac{\left(\pi^C-\pi^P\right)\theta\tau(y_I+y_O)-\pi^C\theta\tau d^C}{1-\tau+2\pi^P\theta\tau} \\ &\doteq p^C. \end{aligned}$$

We denote the cutoff as p^C, and if the cutoff calculated above is smaller than 0 we can set $p^C = 0$. If $p < p^C$, I people strictly prefer conflict instead of peace because the expected gain from trade is relatively low. This tradeoff would disappear if $\pi^C \leq \pi^P$, i.e., if peace not only brought additional trade output, but also increased the incumbent's probability of staying in power, and thus I people strictly preferred peace no matter how low the trust is. The case with$\pi^P < \pi^C$ is more interesting and also more realistic: the incumbent controls the government and therefore is in and advantageous position when a conflict happens, while the opponent groups has more hope of taking over the government using peaceful approaches. We therefore make the following assumption:

Assumption 1. $\pi^P < \pi^C$.

Similarly, we can compare y_I^C and y_I^R:

$$\begin{aligned} y_I^R &\geq y_I^P \Rightarrow \\ y_I+p-d^R &\geq (1-\tau)(y_I+p)+\pi^P\theta\tau(y_I+y_O+2p) \Rightarrow \\ p &\geq \frac{\pi^P\theta\tau y_O-\left(1-\pi^P\theta\right)\tau y_I+d^R}{\tau(1-2\pi^P\theta)} \\ &\doteq p^R, \end{aligned}$$

if $2\pi^P\theta < 1$, or equivalently, $\theta < 1/2\pi^P$. Here p^R denotes the cutoff that makes revolution more attractive for I people and we can set $p^R = 0$ if the cutoff calculated above is smaller than 0. The condition $\theta < 1/2\pi^P$ implies that the expected transfer to I people is not larger than the tax that they pay, so in autocracy I people only

obtain part of the trade output ($2\pi^P\theta$) while in democracy they can claim all the trade surplus. In this case, if the trust is high, i.e., $p > p^R$, I people prefer revolution and democracy over the peaceful autocracy where they are still exploited by the elite. This condition is reasonable and we assume it throughout this paper.

Assumption 2. $\theta < \frac{1}{2\pi^P}$.

Notice that I people can choose to revolt, but they cannot choose between peace and conflict, which is decided by the elite. Hence, the analysis above about I people's preference between revolution and peace in fact has implications on their action and the equilibrium outcome. For example, if $p > p^R$, then $y_I^R > y_I^P$, implying that I people prefer democracy over peaceful autocracy. XYZ SA CAN WE STATE THIS DIFFERENTLY? This suggests that if trust is high, then the elite is in danger of being overthrown, if I people expect the elite will choose peace in autocracy or the income for I people will be even lower if the elite chooses conflict. This result implies that the elite may not want trust to be too high, even though a higher trust is always better for the society and implies higher aggregate output. This preference of I people has important implications for the elite's choice to start a conflict and also for the dynamics of trust. We summarize this in the following Lemma.

Lemma 1. *If $p > p^R$, I people prefer revolution over autocracy with peace. They choose to revolt and establish democracy, if they expect that the elite will choose peace in autocracy.*

3.2.2 The Elite

If revolution occurs, then the elite pays the cost d_E^R. If there is no revolution, the incumbent elite from the beginning of the period can stay in power with probability π^s and then its income equals tax minus transfer. This income also depends on its choice between peace and conflict. If it chooses conflict, it loses the tax revenue from the trade surplus and also part of the tax on home production, because conflict reduces y_I by d^C, but it benefits from preventing the people from learning about the type of the relation and updating the trust. This benefit can be important for the elite if it is in danger of being overthrown by the people if they form a high level of trust. If the elite chooses peace, then it is in power with probability π^P and can tax

the trade surplus. The elite's expected incomes in each period, given the states of the world, are summarized as follows:

- Revolution: $y_E^R = -d_E^R$.
- Peace: $y_E^P = \pi^P(1-\theta)\tau(y_I + y_O + 2p)$.
- Conflict: $y_E^C = \pi^C(1-\theta)\tau(y_I + y_O - d^C)$.

Different from the people, the elite cares also about the future, so we need to write down the elite's lifetime utility, which is equivalent to its discounted expected lifetime income. The elite's lifetime utility also depends on the state of the world. The first case is that revolution happens. Then in the current period the elite's income is $y_E^R = -d_E^R$ and from next period on it gets 0 in democracy. So the lifetime income is $V^R = -d_E^R$, which is in fact independent of p if revolution happens. In the second case, revolution does not happen and the elite chooses peace, such that the elite's lifetime income includes its current period expected income and the future incomes if it stays in power. The expected lifetime income can then be written recursively, as follows:

$$V^P(p) = y_E^P + \beta\pi^P \mathbb{E}V(p'),$$

where p' denotes the level of trust in the next period, which depends on the trade output realization, and $V(p')$ is the continuation value given next period's trust. Notice that the elite still stays in the government with probability π^P and the political turnover happens with probability $1-\pi^P$, where the elite looses all future incomes.[21] After the trade output realizes, agents update their belief on whether the relation is good or bad, according to the corresponding conditional probabilities. Remember that if the type of the relation is good, then the output is high with probability q_H, and if the type is bad, then the output is low with probability q_L. The belief is then updated according to Bayes' rule:

$$p' = \begin{cases} \frac{pq_H}{pq_H + (1-p)(1-q_L)} & \text{if } y_T = y_H, \\ \frac{p(1-q_H)}{p(1-q_H) + (1-p)q_L} & \text{if } y_T = y_L. \end{cases} \tag{1}$$

[21] It can be that a few periods after the political turnover, another political turnover happens, and the current incumbent group regains the political power. However, it is the newly born incumbent elite and not the current incumbent elite that sits in the government, so the current elite expects getting no future incomes if a political turnover happens.

The expectation term $\mathbb{E}V(p')$ includes different cases with different types of relation and trade outcomes. By the law of iterated expectations, we can simplify the expression and make it conditional only on the realizations of the trade outcome, i,e.,

$$\begin{aligned}\mathbb{E}V(p') = &\left(pq_H + (1-p)(q-q_L)\right)V(p'|y_H) \\ &+ \left(p(1-q_H) + (1-p)q_L\right)V(p'|y_L),\end{aligned}$$

where $p'|y_H$ and $p'|y_L$ are the updated beliefs given trade outcomes expressed above in equation (1). Finally, if revolution does not happen and the elite chooses to initiate a conflict against the opponent group, then the value function is:

$$V_E^C(p) = y_E^C + \beta\pi^C V(p).$$

In conflict, there is no trade and the belief is not updated, i.e., $p' = p$.

Knowing the elite's lifetime income in these three cases, we can then characterize the elite's optimal choice between conflict and peace, and determine the value function in the beginning of each period, denoted as $V(p)$:

$$V(p) = \begin{cases} \max\left\{V^C(p), V^P(p)\right\}, & \text{if } p \leq p^R, \\ V^R, & \text{if } p > p^R, \end{cases}$$

3.2.3 The Equilibrium with $\beta = 0$

The elite's choice is key for characterizing the equilibrium of the economy. We first solve the elite's problem and the equilibrium under the condition $\beta = 0$. In this case, the elite only cares about the current period income like the myopic people and it is easy to solve the elite's problem. This simple case is informative for the static trade-offs and delivers an intuitions that is also relevant in the more general case with $\beta > 0$. The dynamic trade-offs are not present in this case and will be discussed later.

When trust is low, then the expected output from trade is low, such that the elite's expected income (tax revenue) in peace is also low. This is the same mechanism as for I people, because both the elite and I people's incomes partly come

from trade. Hence, similar to the discussion on I people, we can also characterize a cutoff for the elite's choice of peace v.s. conflict:

$$\begin{aligned} y_E^P &\geq y_E^C \Rightarrow \\ \pi^P(1-\theta)\tau(y_I+y_O+2p) &\geq \pi^C(1-\theta)\tau(y_I+y_O-d^C) \Rightarrow \\ p &\geq \frac{\left(\pi^C-\pi^P\right)(y_I+y_O)-\pi^C d^C}{2\pi^P} \\ &\doteq p_E^C. \end{aligned}$$

Given that $\pi^C > \pi^P$, the elite's tradeoff between conflict and peace is similar to the one for I people: peace brings in more (taxable) income from trade, but increases the probability of a political turnover. Consequently, if trust is high enough, then the elite finds peace more beneficial than conflict. Otherwise it prefers conflict, because it implies a higher probability of staying in power. We can show that this cutoff p_E^C is higher than the corresponding cutoff for I people, i.e., p^C, implying that the elite is more aggressive in the sense that it prefers conflict over peace even for intermediate levels of trust when the people already prefer peace. This property can be formalized as follows:

$$\begin{aligned} p^C &= \frac{\left(\pi^C-\pi^P\right)\theta\tau(y_I+y_O)-\pi^C\theta\tau d^C}{1-\tau+2\pi^P\theta\tau} \\ &= \frac{\left(\pi^C-\pi^P\right)(y_I+y_O)-\pi^C d^C}{\frac{1-\tau}{\theta\tau}+2\pi^P} \\ &< \frac{\left(\pi^C-\pi^P\right)(y_I+y_O)-\pi^C d^C}{2\pi^P} \\ &= p_E^C, \end{aligned}$$

because $\frac{1-\tau}{\theta} > 0$. The intuition behind this result is that the income from trade is less important for the elite than for the I people. The trade income matters for the elite only because it enters the tax revenue, while for I people, the trade income is not only part of the tax and transfer, but also a direct source of income: after paying the tax (and receiving transfers), they also keep $1-\pi$ fraction of the trade output, corresponding the term $1-\pi$ in the denominator of the expression of p^C. So I people are more in favor of peace than the elite is. This is one of the two

reasons for the elite to be more aggressive in conflict and less interested in peace. This first reason is often discussed in the literature. For example, Rohner, Thoenig, and Zilibotti (2013a) use such an argument to generate a random war that is out of the control of people. Jackson and Morelli (2007) argue that if the political process is captured by a biased political elite, then the war occurs against the interest of people. Notice that though the elite and the people differ in when to start this type of conflict quantitatively, they actually agree on starting this type of conflict qualitatively: when the trust is slow and gain from peace is small, they prefer conflict over peace.

There is a second reason for the elite to prefer conflict over peace and a fundamental difference between the elite and I people.We also often see conflicts and polarization occur between two groups who could have benefited much from peaceful trade and cooperation. In the example of Rwanda discussed above, Hutu and Tutsi did not always use to have a political meaning and the there was mobility between the two groups. Peaceful interactions between the groups could have generated valuable output, which could be beneficial for the government as well. So why did the government decide to polarize them and create a strict distinction between them and even initiate conflicts? Similar questions may also be asked in other contexts that are not necessarily about ethnic groups, e.g., the political polarization between the left and the right wing in the current U.S., and the anti-Japan movements and boycott of Japanese products in China after 2000s. The reason also lies in the biased interest of the political elite, but with a fundamental difference: the elite may lose the political power from increasing trust which is beneficial for the aggregate output and the people, if the high trust leads to revolution. So the elite may want to initial a conflict when the trust is high, in order to prevent it from increasing further and resulting in revolution, which happens if $p > p^R$ and then the elite gets no income but pays a cost: $y_E^R = -d_E^R$. In this version of the model with a myopic elite, i.e., $\beta = 0$, we can see the incentives for the elite to prevent trust from being too high. Intuitively, starting a conflict can stop people from seeing trade outcomes and updating their belief, i.e., there is no possibility for trust to grow, while it does not decrease either. This keeps the elite in a safe place without risking a revolution. This intuition can be formalized in the case that $\beta > 0$, which is discussed below, based on the solution from this case with $\beta = 0$. We first summarize the equilibrium

outcomes with $\beta = 0$ in the following proposition:

Proposition 1. *$p > p^R$: revolution. Trust is high, I people revolt and establish democracy.*

$p \leq p^R$ and $p \geq p_E^C$: peace. Trust is not, high so I people do not revolt, while it is high enough for the elite to choose peace and trade.

If $p \leq p^R$and $p < p_E^C$: conflict. The expected gain from trade is low so the elite chooses conflict.

Notice that without loss of generality, we assume that the agents choose peace when they are indifferent between peace and other choices. The results in this proposition have a clear intuition: if trust is high enough, then the expected return from trade is high and I people is willing to pay the cost of revolution to establish democracy so that their incomes from trade are not taxed by the elite. If trust is lower than the revolution threshold, but still higher than p_E^C, then the trade output is still relatively high, such that the elite finds the expected income in peace higher than in conflict, and therefore chooses peace. If the trust is low enough, i.e., $p < p_E^C$, then the elite does not expect to benefit much from trade, so it chooses conflict to reduce the probability of political turnover.

Moreover, if we know which of the two cutoffs, i.e., p^R and p_E^C is higher, then we can characterize additional properties of the equilibrium. The ranking of p_E^C and p^R depends on the parameters, including the cost of revolution for I people, i.e., d^R. If this cost is sufficiently high, then the revolution cutoff p^R can be higher than the conflict cutoff p_E^C. More precisely, $p_E^C < p^R$ if and only if $d^R > \underline{d}^R$, where the expressions for $\underline{d}^R$ is given in the Appendix. In this case, the equilibrium can be characterized as in the following corollary:

Corollary 1. *If $d^R > \underline{d}^R$, the space of trust can be split into three regions, corresponding to three political outcomes.*

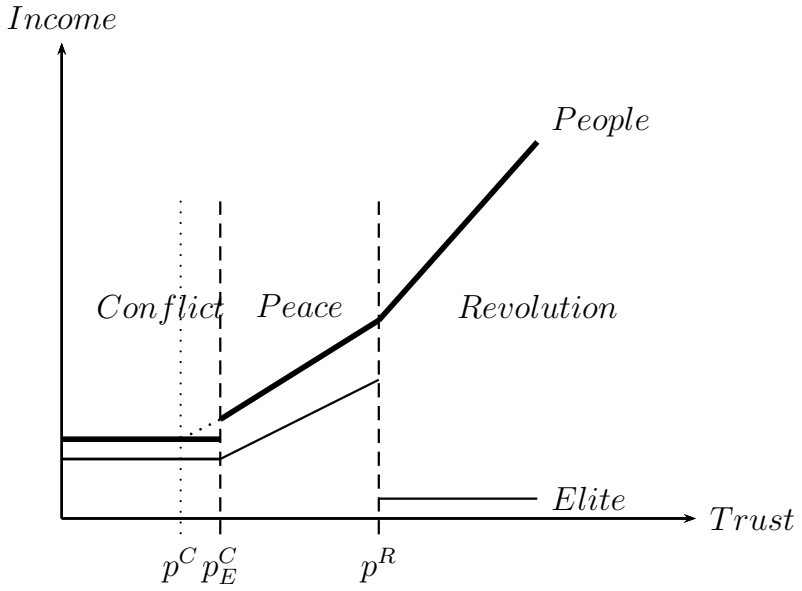
High trust: $p > p^R$. Revolution.

Intermediate level of trust: $p \in \left[p_E^C, p^R\right]$. Autocracy with peace.

Low trust: $p < p_E^C$. Autocracy with conflict.

This corollary is a natural result of Proposition 1. It can also be intuitively illustrated in Figure 1.

Figure 1: Trust and Income



The figure shows different equilibrium outcomes and incomes for different trust levels. The thin solid lines refers to the income of the elite, while the thick solid line represents the income of I people. The dashed lines separate different states of the world: revolution, peace and conflict. The dotted line marks the threshold level of trust above which I people prefer peace to conflict.

This figure shows that if trust moves from the low to the intermediate and then to the high level, the equilibrium moves from conflict to peace and then to revolution. When trust is low and the equilibrium is conflict, then the incomes of the elite and I people are flat lines, because there is no trade output and then expected output does not depend on trust. When trust is in the intermediate range and the equilibrium is peace, then the incomes of the elite and I people both increase with trust, because the expected output from trade increases with trust. Notice that at the junction of conflict and peace, i.e., $p = p_E^C$, the elite's expected income changes continuously while I people's expected income jumps up. The reason is that the decision on peace or conflict is in the hands of the elite but not I people. The elite chooses to switch to peace when its expected income in peace just surpasses the expected income in conflict, while I people already prefer peace to conflict at an earlier stage when trust is lower, i.e., $p^C < p_E^C$. Finally, when trust is high and revolution happens, I people's income increases with trust, while the elite's income is a flat line, because after revolution and democratization, all the trade surplus belongs to the people and nothing to the elite. Moreover, at the junction of revolution and peace, I people's income is continuous while the elite's income is not continuous but drops dramatically, because revolution is subject to the decision of I people and not the elite.

If the condition $d^R > \underline{d}^R$ for Corollary 1 is not satisfied, then $p^R \leq p_E^C$, and the second case with the intermediate range of trust does not exist anymore. The space of trust can be only split into two regions corresponding to conflict and revolution. The remaining properties of the equilibrium are similar, so we omit the discussion to focus on the more interesting case discussed above.

The analysis above with $\beta = 0$ highlights two important trade-offs. The first is static and faced by both I people and the elite: compared to conflict, peace brings in more income from trade, but decreases the probability of staying as the incumbent. So they only prefer peace when trust and expected return from trade are high enough. The second tradeoff is only for the elite: a higher level of trust increases the expected income of the elite in peace, but it may also dramatically decrease the income of the elite if it is too high and leads to revolution. So if the elite chooses peace in this period and receives a high income, it is possible that the trust increases, and in the next period revolution occurs, in which case the elite pays

a large cost. This dynamic tradeoff suggests that the elite has an incentives to take into account – and perhaps influence – the dynamics of trust. When $\beta = 0$, we can see this dynamic tradeoff, but we cannot really discuss the dynamic consideration of the elite, because it cares only about income in the current period. In the beginning of each period, the trust level is given, hence the elite cannot change the trust within this period: if the trust level is higher than the revolution threshold, then the elite cannot do anything but let the revolution happen. Since the elite is myopic, it chooses conflict or peace only according to the current period income while ignoring future incomes. So the next period trust evolves given the static decision made by the elite and it may go above the revolution threshold. If $\beta > 0$ and the elite is not myopic, it will take into account this dynamic tradeoff and it can influence the future trust by choosing peace or conflict in the current period. Next, we study the case with $\beta > 0$, and discuss how the elite prevents trust from increasing too much, with consequences for polarization and conflict.

3.2.4 The Equilibrium with $\beta > 0$

Given $\beta > 0$, the elite cares not only about the current period income, but also about the future incomes. Hence, the elite takes into account how current period policies affect the dynamics of trust and the future equilibria. The crucial dynamic consideration by the elite is to avoid the revolution, which happens when the trust is high enough - above the threshold p_E^R and implies that the elite loses all future incomes and has to pay a direct cost d_E^R. How does the elite prevent trust from increasing too much? It can start a conflict to stop trade and prevent I people from observing trade outcomes and updating their belief. It balances the current period gain from trade against the risk of revolution. If the cost of the revolution is large enough, e.g., the direct cost of revolution d_E^R is large enough, the elite may sacrifice the current period gain from trade and stop the peaceful interactions, when the trust is high and close to the threshold of revolution, in order to avoid the possibility that trust increases above the threshold of revolution when the trade outcome is good. We first state this result formally in the following proposition, and then analyze how this policy affects the dynamics of trust, polarization and conflicts.

Proposition 2. *If the cost of revolution is high enough, then the elite chooses con-*

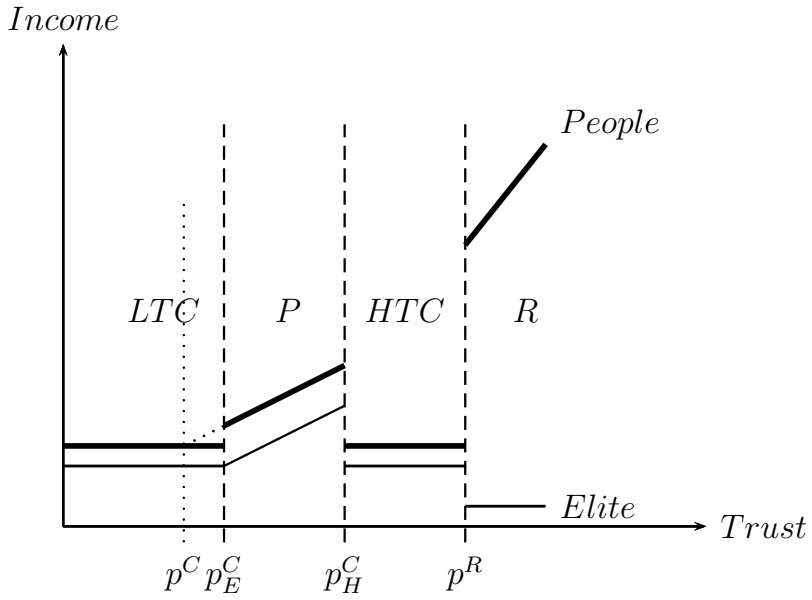
flicts when trust is high and close enough to the threshold of revolution. Formally, there exists a $\bar{\beta}$ such that for all $\beta < \bar{\beta}$, the following result is true: if $d_E^R > \underline{d}_E^R$, then there exists a threshold $p_E^R < p^R$ such that for all $p > p_E^R$, the elite chooses conflict instead of peace, while I people prefer peace.

The proof of the proposition and the expressions of $\bar{\beta}$ and $\underline{d}_E^R$ are in the Appendix. Notice that to make the analytical proof possible, we restrict β to be small, however, the tradeoff applies to any level of β, so it can be the case that this result applies to a large β. We should not interpret the restriction on β as the this result only holding when β is very small, but interpret this restriction as a useful step for the analytical proof.

The intuition behind this result is that if the cost of revolution is high enough, then the elite always wants to avoid the possibility of revolution. The elite wants to keep trust in the "safe" region and when there is a possibility that the next period trust goes above the revolution threshold, the elite always stops trade and keeps trust at the current level.[22] More specifically, when p is close enough to p^R, if the trade output realization is high, the updated trust in the next period will be above p^R, and I people will revolt. Then the optimal choice for the elite is to use conflict to stop people from trading and possibly developing higher trust in the next period. The dynamic mechanism behind this "conflict with high trust" is quite different from the static one for the "conflict with low trust", which is described above in the case with $\beta = 0$. In the low trust case, the tradeoff is in the current period income, and conflict is in some sense efficient for the incumbent, since the expected trade surplus in peace is low. I people and the elite both agree that when the trust is low, conflict is better than peace, though they have some disagreements on the exact threshold level of starting conflict - p^C v.s. p_E^C. In the high trust case, the elite's dynamic interest is completely in the opposite direction of I people's interest and the aggregate efficiency: I people and the output always benefit from a higher level of trust, but the elite's income is not monotonically related to trust. The elite's one period expected income, as we can see from Figure 1, increases with trust initially but dramatically drops when trust surpasses the revolution threshold p^R. Anticipating this potential loss with a too high trust in the future, the elite's optimal

[22] In the extension with the stochastic type of relation discussed later, trust does not stay at the same level after conflicts.

Figure 2: Conflict with High Trust



The figure shows different equilibrium outcomes and incomes for different trust levels. The thin solid lines refers to the income of the elite, while the thick solid line represents the income of I people. The dashed lines separate different states of the world: revolution (R), high trust conflict (HTC), peace (P), and low trust conflict (LTC). The dotted line marks the threshold level of trust above which I people prefer peace to low trust conflict.

choice in this period can be starting a conflict to prevent the possibility of further increase of trust.

Figure 2 shows the different trust regions. The "conflict with high trust" is an interesting result, because conflict is endogenously created by the elite for its own interest, against the efficiency of the society. The previous literature mainly focuses on exogenous causes of conflicts. For example, given the exogenous differences between the two groups (e.g., Esteban and Ray 2011), and the realization of trade outcomes (e.g., in Rohner et.al. 2013a), conflicts happen because the two groups expect little gain from cooperation and trade. These types of conflicts correspond to the low trust conflicts in this model. The high trust conflict discussed in this paper is new to the literature. This type of conflicts and polarization, i.e., conflicts that stop cooperation between groups who are not exogenously so different and have high

potential gains from cooperation, are often observed in the real world. For example, the initial segregation between Hutu and Tutsi was initially introduced by Rwandan rulers in the 19th century and then strengthened by colonizers, although there had been many peaceful interactions between these two groups, including marriages and changing of ethnicity. The interactions stopped not because they were expected to be inefficient, but because the cooperation of the two groups may have been dangerous for the political elite: high trust can lead to a strong common interest among the two groups and trigger a successful revolution against the political elite. The colonizers adopted the divide and rule strategy and reduced the interactions and cooperation between Hutu and Tutsi when trust was still relatively high and polarization was low. The colonizers used the Tutsi in the administration to rule the country and they assigned identity cards that made ethnicity a central feature, whereas it used to be possible to change ethnicity. The increasing polarization and decline of trust between these two groups were the consequences, not the cause, of the decline of interactions and cooperation. Other examples include the current increasing political polarization in the U.S., and the anti-Japan sentiments in China. In both cases, there is no reason to believe that the return to cooperation between the two groups is particularly low. Our model provides a different perspective – it can be the high potential return from the interactions of the people that leads to the stop of interactions and the initiation of conflicts, because the cooperation of the groups may harm the interest of the polarizing elite. The elite's interest is different from the interest of the overall society and of the people. While a higher level of trust and more cooperation generate a higher aggregate output, they also change the allocation of income against the interest of the elite. This is why the elite wants to reduce interactions, start conflicts, and potentially increase polarization, when there is high potential gain from the cooperation of the two groups.

In addition to understanding the causes of conflicts, there are a number of other interesting questions. What are the consequences of conflict when trust is high? Why, after the colonizers adopted the divide-and-rule strategy, did severe polarization and conflicts happen between Tutsi and Hutu? In the next section, we study these questions using an extension of the model with a stochastic type of relation between the two groups. We discuss the consequences of the high trust conflicts -

how they may lead to further polarization and persistent conflicts with low trust.[23]

Notice that the "conflict with low trust" still exist when $\beta > 0$, which is formally stated in the following proposition and proved in the Appendix.

Proposition 3. *If the cost of conflict is not too high, then the elite chooses conflict when trust is low enough. Formally speaking, there exists a $\bar{\beta}$ such that for all $\beta < \bar{\beta}$, the following result is true: if $d^C < \frac{\pi^C - \pi^P}{\pi^C}(y_I + y_O)$, then there exist a $p_E^C > 0$, such that for all $p < p_E^C$, the elite chooses conflict over peace.*

The main logic behind this result is the same as in the case when $\beta = 0$. While there are dynamic considerations when $\beta > 0$, they are in line with the static trade-offs between conflict and peace, which are sufficient for this result. Just like before, when the cost of conflict is low and the potential gain from trade is low, conflict happens.

4 Extension and Discussion

In this extension, we introduce stochastic types for the relation between the two groups, which allows for interesting dynamics after conflicts happen and further discussions on the consequences of conflicts, especially the conflicts at a high trust level.

4.1 Stochastic Types of the Relation between Two Groups

We assume that there is a probability that the type of the relation between the two groups changes at the end of each period. Remember that the type of the relation affects the probability of getting a high trade output and the belief of the type being good is the trust. The changes of types can be driven by cultural shifts, which can be related to social structure, population mixture, and so on (see, for

[23]Here in the benchmark model, we can study why conflicts happen when trust is high and how the elite use conflict to prevent revolution. However, the consequences of this policy, i.e., what happen after the high trust conflict, are not so interesting: since there are no shocks to the type of the relation and no additional shocks to trust when trade and trade outcome shocks are shut down, trust stays constant at the level when the elite starts the high trust conflict. Then the high trust conflict becomes an absorbing state and the society stays in that state forever. It is similar when the low trust conflict happens.

example, Rohner, Thoenig, and Zilibotti, 2013a). The changes can also be affected by the state of the world. For example, if two groups are in conflict, the capital for more efficient interactions depreciates - roads are destroyed during the war, or knowledge of each other's language and culture decline when cultural interactions and exchanges stop - and then the type of the relation may turn bad. In the model, the type of the relation changes with certain probabilities: a good relation may turn into a bad one with the probability ψ^s, while a bad relation may become a good one with probability ϕ^s, given the state of the world in conflict or peace, i.e., $s \in \{C, P\}$. In other words, the stochastic process follows a two-state first-order stochastic Markov process with the following transition matrix from period $t-1$ to t :

	t	G	B
$t-1$			
G		$1-\psi^s$	ψ^s
B		ϕ^s	$1-\phi^s$

where G and B denote the types of the relation being "good" and "bad", respectively. Agents are aware of the possibility that the relation can change. This implies that trust - the belief of agents on the probability of the relation being good - can change even if there is no trade.

Now we can show how trust evolves over time. Right after the trade outcome realizes, similar to before, all agents update their belief according to Bayes' rule:

$$p^+ = \begin{cases} \frac{pq_H}{pq_H+(1-p)(1-q_L)} & \text{if } y_T = y_H, \\ \frac{p(1-q_H)}{p(1-q_H)+(1-p)q_L} & \text{if } y_T = y_L, \end{cases}$$

where p^+ is the belief after observing the trade outcome and before the type change may happen at the end of the period. If the society is in conflict and no trade takes place, then there is no new information about the type of the relation, so the belief stays the same, i.e., $p^+ = p$. At the end of the period, the stochastic shock to the type of the relation realizes. After that, the belief that the type is "good" becomes $p' = p^+(1-\psi^s) + (1-p^+)\phi^s = p^+(1-\psi^s-\phi^s) + \phi^s$, and this is also the level of trust at the beginning of the next period. The evolution of trust across two periods

is summarized below:

$$p' = \begin{cases} \left(1-\psi^C-\phi^C\right)p+\phi^C & \text{if } s=C, \\ \left(1-\psi^P-\phi^P\right)\frac{pq_H}{pq_H+(1-p)(1-q_L)}+\phi^P & \text{if } s=P, y_T=y_H, \\ \left(1-\psi^P-\phi^P\right)\frac{p(1-q_H)}{p(1-q_H)+(1-p)q_L}+\phi^P & \text{if } s=P, y_T=y_L. \end{cases} \tag{2}$$

How does conflict affect trust, i.e., the belief that the relation is good? There are two channels. First, as in the benchmark model, conflict stops agents from learning about the relation from the trade outcome. Second, in this stochastic relation setup, conflict directly changes the belief if $\psi^C \neq \psi^P$ or $\phi^C \neq \phi^P$. One possible case is that $\psi^C > \psi^P$ and $\phi^C < \phi^P$: in conflict, the relation turns from good to bad with a higher probability compared to the case in peace, and turns from bad to good with a lower probability. This can happen if conflict directly harms the communication capital and thus the relation. For example, during the conflict, roads may be destroyed, and both groups stop sending people to learn and understand the other group's language and culture. Then the frictions in the interactions become larger and the potential net gain from cooperation and interaction declines. This difference in the transition probability of conflict and peace implies that the belief of the relation being good, i.e., the trust, is directly harmed by the conflict.

If conflict does directly harm trust, then after a high trust conflict is started, trust may gradually decline and eventually become low enough such that revolution is no longer an immediate threat. This is the key deviation from the benchmark. First, we show in the Appendix that with the stochastic relation, there can still be cutoffs for high trust and low trust conflicts. As before, we still denote the cutoff for the high trust conflict as p_E^R, i.e., if $p > p_E^R$ then the elite uses conflict to avoid future revolution. After the conflict happens, trust starts to converge towards $\frac{\phi^C}{\phi^C+\psi^C}$ with the auto-correlation $\left(1-\psi^C-\phi^C\right)$, as can be seen in equation (2). If ψ^C is large enough, such that $\frac{\phi^C}{\phi^C+\psi^C} < p_E^R$, then after a certain number of periods, trust becomes lower than the cutoff for high trust conflict, so the elite wants to end the high trust conflict. If by then trust is higher than the cutoff for low trust conflict, peace is recovered, trade starts, and the belief gets updated again.[24] In contrast,

[24]It can also be that in the period that trust drops below the threshold of high trust conflict, it also goes below the threshold of low trust conflict, such that the society immediately jumps from

the low trust war may be permanent, especially if ψ^C is large enough such that $\frac{\phi^C}{\phi^C+\psi^C} < p_E^C$. After the low trust conflict starts, p converges to $\frac{\phi^C}{\phi^C+\psi^C}$ and is always below p_E^C, implying that the elite always chooses to stay in conflict in all future periods. If $\frac{\phi^C}{\phi^C+\psi^C} \geq p_E^C$, then after some periods, p grows back to a level above p_E^C, and peace resumes, but it may take long if ψ^C is relatively large. Hence, the low trust conflict can last very long or even forever. These results on the duration of the high trust and the low trust conflict can be formally stated in the following proposition.

Proposition 4. *Conflict with high trust is temporary, while conflict with low trust can last long or even be permanent if conflict is sufficiently harmful for the relation between the two groups. Formally, if ψ^C is sufficiently large, then after a high trust conflict occurs, there exist a $\bar{t}$ such that after at most $\bar{t}$ periods, the elite stops the high trust conflict. Moreover, for any integer $\underline{t}$, there exists a sufficiently large ψ^C such that the low trust conflict lasts at least $\underline{t}$ periods. Here $\underline{t}$ can be generalized to $+\infty$.*

The proof can be in the Appendix. The intuition is that if conflict reduces the probability of the relation being good significantly, then during the high trust conflict, the trust level must decline and eventually reach such a low level that the elite does not have to worry about revolution in the next period; while during the low trust conflict, the trust level stays low and makes conflict more attractive for the elite than peace for a long time, or even permanently.

Notice that there is an interesting link between the high trust conflict and the low trust conflict: the high trust conflict, though temporary, lowers the trust and harms the relation, so it may lead to a persistent or permanent low trust conflict, especially if the trade outcomes are bad during the peace periods that follow the high trust conflict. The high trust conflict increases the polarization of the society and the possibility of persistent low trust conflicts through two channels. First, it directly harms the relation and may change the relation to the bad type. During the high

the high trust conflict to the low trust conflict, skipping the peace stage. This possibility exists only in the discrete time model when ψ^C is very large, e.g., close to 1. If we change to continuous time, or simply set one period to be short enough, then this jump will not happen and there are always some periods when trust is lower than p_E^R, but still higher than p_E^C given that $p_E^R > p_E^C$. For this reason, we only consider the case where, when the high trust conflict ends, peace resumes.

trust conflict, the probability of switching from the good relation to the bad type - ψ^{C-} can be higher than the counterpart in peace - ψ^{P}. Hence, it is possible that before the conflict, the type is good, but after the conflict, the type changes to the bad. If this happens, later trade outcomes are more likely to be low and then trust is more likely to decline to a level below the low trust conflict threshold, resulting in increasing polarization and eventually persistent conflicts with low trust. Second, even if the real type does not change during the conflict, the belief on the type also changes after the high trust conflict ends: the trust level decreases from above to below p_E^C, therefore below the level when the conflict starts. Given a sequence of trade outcomes, a lower trust implies a higher probability of encountering a trust level lower than p_E^C. We formally establish these results about the impacts of the elite's policy on polarization and conflicts in the following proposition.

Proposition 5. *The elite's divide-and-rule strategy, i.e., using temporary high trust conflict to prevent revolution, increases polarization and the probability of long-lasting or even permanent low trust conflicts. Formally, if the high trust conflict starts in period* t_0 *and ends in period* t_1*, then* $p_{t_0} > p_{t_1}$ *and* $\mathbb{P}\left(p_t < p_E^C | s_{t_0} = C, t > t_1\right) > \mathbb{P}\left(p_t < p_E^C | s_{t_0} = P, t > t_0\right)$.

The proof is in the Appendix. These results show that when the elite uses conflict to prevent revolution, then the channel that increases trust is shut down. Then, inevitably, trust declines and polarization increases. Though the elite may not want the polarization to increase so much that the society ends in long-lasting low trust conflicts, which can be detrimental for both the people and the elite, this process is not under the control of the elite, but it depends on the stochastic outcomes of trade and cooperation between the two groups. There is a danger that after the initial increase of polarization due to the elite's divide-and-rule strategy, polarization keeps increasing and eventually ends up at a very high level that leads to long-lasting conflicts. Through the lens of this model, we can interpret the example of Tutsi and Hutu as follows. The king and later on the colonizers might not have liked Hutu and Tutsi to develop too much trust, which could allow them to cooperate in the revolution against the colonizers. Consequently, the colonizers decided to polarize the population by dividing them into two fixed ethnic groups and reducing their interactions. This policy seemed to work well in terms of allowing

the colonizers to maintain their political power and exploit the country. However, the consequences afterwards may be much more detrimental than what would have been in their interest initially: very high polarization, long-lasting domestic wars and genocide. Once polarization was started and the Pandora's box was open, bad outcomes could follow: negative shocks to the outcomes of interactions eroded the trust between Hutu and Tutsi, the ethnic polarization kept increasing, and the conflicts became different and out of the control of later governments. The colonizers may not have wanted to directly create the intense conflicts with low trust, but its initial polarization policy could be the first step towards the uncontrollable polarization and conflicts.

5 Conclusion

We propose a theory of divide-and-rule where political elites strategically initiate conflicts between ethnic groups in order to polarize society and thus sustain their own power. We model polarization as a lack of trust that is shaped by trade interactions between agents of the different ethnic groups. The elite can prevent trust from emerging by starting a conflict that interrupts trade. The elite follows this strategy when there is a threat of revolution. The model also generates that an elite is more likely to seek war than the people. This is the case because the elite has the double fears of losing power to the other group and to be overthrown by the own group if the common interest between groups becomes too large.

We document that our model is consistent with a number of cases of ethnic violence under autocratic rule and also with political polarization in a democratic setting. Cases such as the incidents in Rwanda and Yugoslavia show that polarization was not simply exogenously given, but to some extent *constructed* by power-seeking elites. Violence often had the goal of destroying trust and creating instability. This allowed elites to exercise their power more freely. An implication of these observations is that treating conflicts as "inherently ethnic" may be misleading, because the role of these identities are shaped by political actions. Our framework of trust and polarization is also applicable in a democratic setting, where the political elite may benefit from increased polarization if it strengthens its power within the own group.

Our model shows the elite's role in creating polarization and it therefore enables us to discuss possible counter-measures. Generally speaking, attempts to prevent conflict should pay considerable attention to the role of the elite. Promising ways to achieve this would be to reduce elites' incentive or ability to reduce trust. This is especially important in cases where the elite may attempt to do so because they are threatened by a *high* level of trust. In such cases, the interests of the elite diverge from those of society overall, since the elite values being in power more than possible gains from productive interactions.

References

Abramowitz, A.I. & Saunders, K.L. (2008). Is Polarization a Myth? *The Journal of Politics*, 70(2), 542-555.

Acemoglu, D., Ticchi, D., & Vindigni, A. (2010). Persistence of civil wars. *Journal of the European Economic Association*, 8(2-3), 664-676.

Acemoglu, D., Robinson, J. A., & Verdier, T. (2004). Kleptocracy and divide-and-Rule: A model of personal rule. *Journal of the European Economic Association*, 2(2-3), 162-192.

Acemoglu, D., & Wolitzky, A. (2014). Cycles of Conflict: An Economic Model. *American Economic Review*, 104 (4), 1350-1367.

Alesina, A. & La Ferrara, E. (2002). Who trusts others?. *Journal of Public Economics*, 85(2), 207-234.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3), 797-817.

Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P., Chen, H., Hunzaker, M.F., Lee, J., Mann, M., Merhout, F. & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216-9221.

Barbieri, K. (1996). Economic interdependence: A path to peace or a source of interstate conflict? *Journal of Peace Research*, 33(1), 29-49.

Besley, T., & Persson, T. (2011). *Pillars of prosperity: The political economics of development clusters*. Princeton: Princeton University Press.

Bhattacharya, S., Deb, J. & Kundu, T. (2015). Mobility and conflict. *American Economic Journal: Microeconomics*, 7(1), 281-319.

Blattman, C., & Miguel, E. (2010). Civil war. *Journal of Economic Literature*, 3-57.

Caselli, F., & Coleman, W. J. (2013). On the theory of ethnic conflict. *Journal of the European Economic Association*, 11(1), 161-192.

CIA (2014). The world factbook 2014. *Central Intelligence Agency, Washington, DC*.

De Figueiredo Jr, R. J., & Weingast, B. R. (1997). Rationality of fear: political opportunism and ethnic conflict. *Military Intervention in Civil Wars*.

Esteban, J., Mayoral, L., & Ray, D. (2012). Ethnicity and conflict: An empirical study. *The American Economic Review*, 102(4), 1310-1342.

Esteban, J., & Ray, D. (2008). On the salience of ethnic conflict. *The American Economic Review*, 98(5), 2185-2202.

Esteban, J., & Ray, D. (2011). Linking conflict to inequality and polarization. *The American Economic Review*, 101(4), 1345-1374.

Fearon, J. D., & Laitin, D. D. (2000). Violence and the social construction of ethnic identity. *International Organization*, 54(4), 845-877.

Fearon, J. D., & Laitin, D. D. (2003). Ethnicity, insurgency, and civil war. *American political science review*, 97(1), 75-90.

Fiorina, M.P., & Abrams, S.J. (2008). Political Polarization in the American Public. *Annual Review of Political Science*, 11, 563-588.

Glaeser, E.L., Laibson, D.I., Scheinkman, J.A. & Soutter, C.L. (2000). Measuring trust. *The Quarterly Journal of Economics*, 115(3), 811-846

Goldsmith, B. E. (2013). International Trade and The Onset and Escalation of Interstate Conflict: More to Fight About, or More Reasons Not to Fight?. *Defence and Peace Economics*, 24(6), 555-578.

Gourevitch, P. (1998). *We wish to inform you that tomorrow we will be killed with our families: Stories from Rwanda*. London: Macmillan.

Gowa, J., & Hicks, R. (2012). Conflict and Commerce: New Data about the Great War.

Horowitz, D. L. (2000). *Ethnic groups in conflict*. Berkeley: University of California Press.

Iyengar, S., & Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, 59(3), 690-707.

Jackson, M. O., & Morelli, M. (2007). Political bias and war. *The American Economic Review*, 1353-1373.

Layman, G. C., Carsey, T. M., & Horowitz, J. M. (2006). Party polarization in American politics: Characteristics, causes, and consequences. *Annual Review of Political Science*, 9, 83-110.

Martin, P., Mayer, T., & Thoenig, M. (2008). Make trade not war? *The Review of Economic Studies*, 75(3), 865-900.

Mason, L. (2015). 'I Disrespectfully Agree': The Differential Effects of Partisan Sorting on Social and Issue Polarization. *American Journal of Political Science*, 59(1), 2015, 128-145.

Mason, L. (2018). Losing Common Ground: Social Sorting and Polarization. *The Forum*, 16(1), 47-66.

Montalvo, J. G., & Reynal-Querol, M. (2005). Ethnic polarization, potential conflict, and civil wars. American economic review, 95(3), 796-816.

Newbury, M. C. (1978). Ethnicity in Rwanda: the case of Kinyaga. *Africa: Journal of the International African Institute*, 48(1), 17-29.

Newbury, M.C. (1988). *The Cohesion of Oppression: clientship and ethnicity in Rwanda, 1860-1960*. New York: Columbia University Press.

Newbury, M. C. (1998). Ethnicity and the Politics of History in Rwanda. *Africa Today*, 7-24.

Padro-i-Miquel, G. (2007). The control of politicians in divided societies: the politics of fear. *The Review of Economic Studies*, 74(4), 1259-1274.

Posner, E. A., Spier, K. E., & Vermeule, A. (2010). Divide and conquer. *Journal of Legal Analysis*, 2(2), 417-471.

Pottier, J. (2002). *Re-imaging Rwanda*. Cambridge: Cambridge University Press

Prunier, G. (2009). *Africa's World War: Congo, the Rwandan genocide, and the making of a continental catastrophe*. Oxford: Oxford University Press.

Prunier, G. (1995). *The Rwanda crisis: History of a genocide*. New York: Columbia University Press.

Rogall, T. (2014). Mobilizing the Masses for Genocide. Mimeo.

Rohner, D. (2011). Reputation, group structure and social tensions. *Journal of Development Economics*, 96(2), 188-199.

Rohner, D., Thoenig, M., & Zilibotti, F. (2013a). War Signals: A Theory of Trade, Trust, and Conflict. *The Review of Economic Studies*, 80(3), 1114-1147.

Rohner, D., Thoenig, M., & Zilibotti, F. (2013b). Seeds of distrust: conflict in Uganda. Journal of Economic Growth, 18(3), 217-252.

Tabellini, G. (2008). The scope of cooperation: Values and incentives. *The Quarterly Journal of Economics*, 123(3), 905-950.

The Oxford Dictionary of Proverbs (2014). "Divide and rule". *Oxford Reference Online*. Oxford University Press. Harvard University Library.

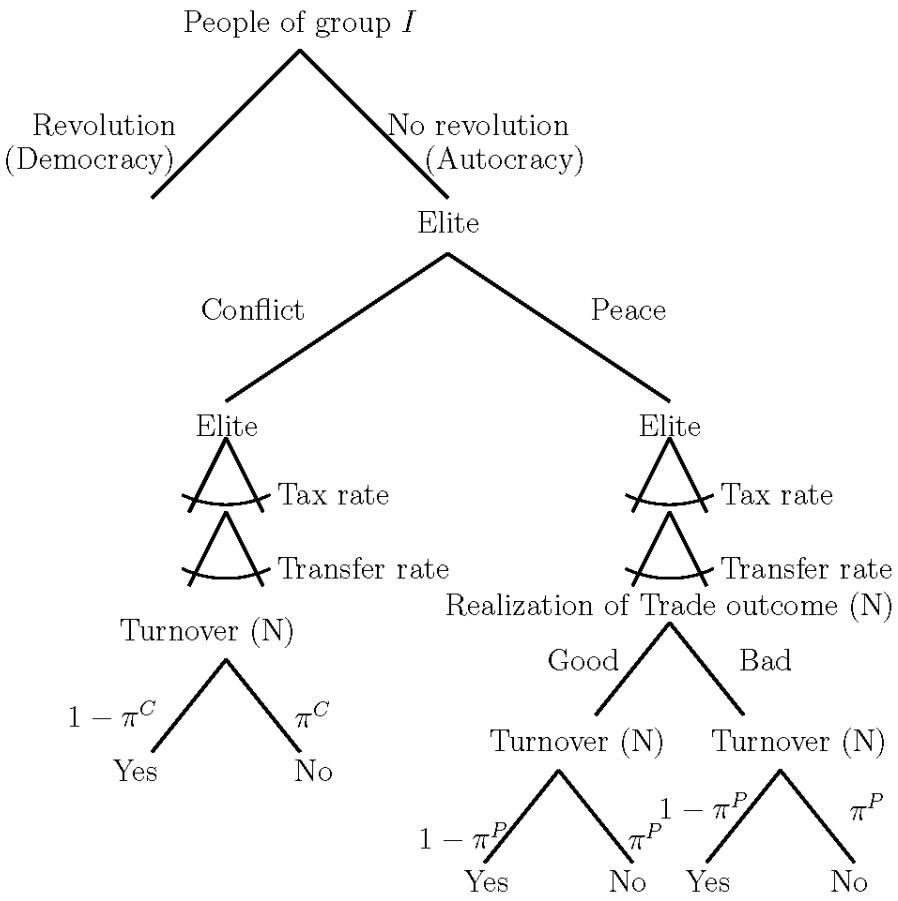
Woodward, S. L. (1995). *Balkan tragedy: Chaos and dissolution after the Cold War*. Washinton, D.C.: Brookings Institution Press.

Zak, P.J. & Knack, S. (2001). Trust and growth. *The Economic Journal*, 111(470), 295-321.

Zimmerman, W. (1995). Last Ambassador-A Memoir of the Collapse of Yugoslavia, *Foreign Affairs*, 74, 2.

A Figures

Figure 3: Game Tree



The figure shows the timing of the baseline model. (N) denotes nature.

B Proofs of Propositions

In this appendix, we provide proofs for the propositions and some lemmata that are useful.

B.1 Proof of Proposition 1

Here we provide the proof for proposition 1, which states the equilibrium outcomes conditional on sufficiently large revolution cost.

Proof. If the cost of revolution for incumbent people is high, then they prefer revolution to peace only when trust is very high, potentially higher than the cut-off value of war and peace. This is to say,

$$\begin{aligned}
p^R &> p_E^C \iff \\
\frac{\pi^P\theta\tau y_O - \left(1-\pi^P\theta\right)\tau y_I + d^R}{1-\pi^P\theta} &> \frac{\left(\pi^C-\pi^P\right)\theta\tau(y_I+y_O) - \pi^C\theta\tau d^C}{1-\tau+2\pi^P\theta\tau} \iff \\
d^R &> \frac{\left(1-\pi^P\theta\right)\left(\left(\pi^C-\pi^P\right)\theta\tau(y_I+y_O) - \pi^C\theta\tau d^C\right)}{1-\tau+2\pi^P\theta\tau} \\
&\quad -\pi^P\theta\tau y_O + \left(1-\pi^P\theta\right)\tau y_I \\
&\doteq \underline{d^R}.
\end{aligned}$$

If the cost of revolution on the elite is high enough, then the elite doesn't want revolution at all, since it is always dominated by either war or peace. Let us consider the case that revolution is always dominated by war, as the following:

$$\begin{aligned}
y_E^R &< y_E^C \iff \\
-d_E^R &< \pi^C(1-\theta)\tau(y_I+y_O-d^C) \iff \\
d_E^R &> -\pi^C(1-\theta)\tau(y_I+y_O-d^C) \\
&\doteq \underline{d_E^R}.
\end{aligned}$$

□

B.2 Proof of Proposition 3

Lemma 2. *The value function of the elite's lifetime income $V(p)$ is bounded above and below by some $\bar{V}$ and $\underline{V}$, respectively.*

Proof. To find the solution for the value function, we first characterize important properties of the value function. First, it is bounded. If the expected trade surplus is very close to its lower bound 0, then the elite can always at least start a war and gain from the tax extraction, so we get a lower bound of V:

$$\begin{aligned} V &\geq y_E^C + \beta\pi^C y_E^C + \beta^2\pi^C y_E^C + ... \\ &= \frac{1}{1-\beta\pi^C} y_E^C \\ &\doteq \underline{V}. \end{aligned}$$

Similarly, if in each period, the elite expects to get at most $\pi^C(1-\theta)\tau(y_I+y_O+2) \doteq \bar{y}_E$ when the expected trade surplus is as high as 1 for each group, and the probability of the elite staying in power is at most π^C, so we can get a upper bound of V as follows:

$$\begin{aligned} V &\leq \bar{y}_E + \beta\pi^C \bar{y}_E + \beta^2\pi^C \bar{y}_E + ... \\ &= \frac{1}{1-\beta\pi^C} \bar{y}_E \\ &\doteq \bar{V}. \end{aligned}$$

□

With this lemma, we can first prove and discuss proposition 3 which shows that war occurs when trust is low.

Proof. War dominates peace if and only if:

$$\begin{aligned} V^P(p) &< V^C(p) \iff \\ & \pi^P(1-\theta)\tau(y_I+y_O+2p)+\beta\pi^P E^P\left[V\left(p'\right)\right] \\ &< \pi^C(1-\theta)\tau(y_I+y_O-d^C)+\beta\pi^C E^C\left[V\left(p'\right)\right] \iff \\ p &< \frac{\left(\pi^C-\pi^P\right)(1-\theta)\tau(y_I+y_O)+\beta\pi^C E^C\left[V\left(p'\right)\right]-\beta\pi^P E^P\left[V\left(p'\right)\right]-\pi^C(1-\theta)\tau d^C}{2\pi^P(1-\theta)\tau} \\ &\doteq \underline{p}(p). \end{aligned}$$

This means that if p is sufficiently small, the gain from trade is smaller than the benefit of war $\underline{p}(p)$, which contains three parts: (1) higher probability of staying in power and getting the current period income: $\left(\pi^C-\pi^P\right)(1-\theta)\tau(y_I+y_O)$; (2) difference in continuation value in the future between war and peace: $\beta\pi^C E^C\left[V\left(p'\right)\right]-\beta\pi^P E^P\left[V\left(p'\right)\right]$; and (3) the cost of war $\pi^C(1-\theta)\tau d^C$. If the cost the war d^C is small, $\underline{p}(p)$ is large, and if meanwhile the trust is low, we have $p<\underline{p}(p)$, implying that the elite chooses war instead of peace. Consider the case that $d^C<\frac{\left(\pi^C-\pi^P\right)}{\pi^C}(y_I+y_O)\doteq\bar{d^C}$. If the trust is at the minimal level, i.e., $p=0$, we have $\underline{p}(0)>p=0$, and the elite prefers war. We can verify this in the following three steps. First, $E^C\left[V\left(p'\right)\right]=E^P\left[V\left(p'\right)\right]=V(\phi)$, as $p=0$ implies $p^+=0$ in both war and peace, and then $p'=\phi$ due to the possibility of type change. Second, $V(\phi)>0$. We know that $\phi\leq\frac{\phi}{\phi+\psi}\leq p^R$, and given $p=\phi$, one possible choice for the elite is to keep having war forever and there will be no threat of revolution, as p will converges to $\frac{\phi}{\phi+\psi}\leq p^R$. In his case, every period, the elite's expected income is $y_E^C=\pi^C(1-\theta)\tau(y_I+y_O-d^C)>0$. The optimal choice for the elite gives higher life-time income than permanent war, so $V(\phi)>0$. This is in fact true for all $p\leq p^R$. Third,

$$\begin{aligned} p(0) &= \frac{\left(\pi^C-\pi^P\right)(1-\theta)\tau(y_I+y_O)-\pi^C(1-\theta)\tau d^C}{2\pi^P(1-\theta)\tau}+\frac{\beta\pi^C\left[V(\phi)\right]-\beta\pi^P\left[V(\phi)\right]}{2\pi^P(1-\theta)\tau} \\ &> 0+0 \\ &= 0. \end{aligned}$$

This shows that when the trust is as low as 0, the elite prefers war. Moreover, if

$V(p)$ is continuous at $p = \phi$, $\underline{p}(p)$ is continuous at $p = 0$. Then given $0 < \underline{p}(0)$, there exists a neighborhood of 0, denoted as $[0, p_E^C)$, for any $p \in [0, p_E^C)$, we have $p < \underline{p}(p)$. In other words, in this low trust region, war is started because there is little to gain from trade. □

B.3 Proof of Proposition 2

Proposition 2 shows that "surprisingly", war also occurs when trust is too high, because the elite is afraid of even higher trust leading to revolution. We provide the proof and the discussion below.

Proof. Consider the situation $p = p^R$. If peace is chosen, with probability $q^R = p^R q_H + \left(1 - p^R\right)(1 - q_L)$, the trade outcome is good and $p' > p^R$, which triggers a revolution in the next period. Then

$$V^P\left(p^R\right) \leq \pi^P(1-\theta)\tau(y_I + y_O + 2p^R) + \beta\pi^P\left(-q^R d_E^R + \left(1 - q^R\right)\bar{V}\right).$$

A sufficient condition for $V^P\left(p^R\right) < V^C\left(p^R\right)$ is:

$$\begin{aligned}
& \pi^P(1-\theta)\tau(y_I + y_O + 2p^R) + \beta\pi^P\left(-q^R d_E^R + \left(1 - q^R\right)\bar{V}\right) \\
< \; & \pi^C(1-\theta)\tau(y_I + y_O - d^C) + \beta\pi^C\underline{V} \iff \\
d_E^R \; > \; & \frac{\pi^P(1-\theta)\tau(y_I + y_O + 2p^R) - \pi^C(1-\theta)\tau(y_I + y_O - d^C) + \beta\pi^P\left(1 - q^R\right)\bar{V} - \beta\pi^C\underline{V}}{\beta\pi^P q^R} \\
\doteq \; & \underline{d_E^R}.
\end{aligned}$$

We can see that given d_E^R is sufficiently large, the elite chooses to go to war when the trust is as high as p^R. In fact, we can see that if p increases from below, before it reaches p^R, war is already preferred by the elite to peace. This is because if trust is slightly smaller than p^R, the threat of revolution in peace – the probability of having a high enough trust that leads to revolution – is only slightly smaller and the cost of peace for the elite is still large. In other words, there exists a neighborhood of p^R, denoted as $(p_E^R, p^R]$, for all p in this neighborhood, we have

$$\begin{aligned}
d_E^R \; > \; & \frac{\pi^P(1-\theta)\tau(y_I + y_O + 2p) - \pi^C(1-\theta)\tau(y_I + y_O - d^C) + \beta\pi^P\left(1 - q^R\right)\bar{V} - \beta\pi^C\underline{V}}{\beta\pi^P q^R} \\
\implies \; & V^P\left(p^R\right) < V^C\left(p^R\right).
\end{aligned}$$

War is preferred by the elite when $p > p_E^R$. □