Ethnic Conflicts with Informed Agents: A Cheap Talk Game with Multiple Audiences*

Pathikrit Basu[a], Souvik Dutta[b] & Suraj Shekhar[c]

a - California Institute of Technology

b - Indian Institute of Management Bangalore

c - University of Cape Town

July 31, 2018

Abstract

We consider a society on the brink of ethnic conflict due to misinformation. An 'informed agent' is a player who has private information which may prevent conflict. We analyze whether the informed agent can achieve peace by communicating privately with the players. The issue is that if the informed agent is known to be biased towards her own

*a - Pathikrit Basu: California Institute of Technology. Address - 202, Annenberg Building, Pasadena, CA - 91125, USA. E-mail - pathkrtb@caltech.edu, Phone number - +1814778012. b - Souvik Dutta: Indian Institute of Management, Bangalore. Address - E001, Indian Institute of Management, Bangalore, India. E-mail - souvik@iimb.ernet.in, Phone number - +919741533886. c - Suraj Shekhar (Corresponding Author): African Institute of Financial Markets and Risk Management (AIFMRM), University of Cape Town. Address - 622 Leslie Commerce building, Upper Campus, University of Cape Town, Rondebosch, Cape Town - 7701, South Africa. E-mail - suraj.shekhar22@gmail.com, Phone number - +27788813926. We are deeply grateful to Kalyan Chatterjee, Sona Golder, Edward Green, Kala Krishna and three anonymous referees for helpful suggestions. We also want to thank Ryan Fang, Anirban Mitra, Bruno Salcedo, Tetsuya Hoshino, Guillem Roig, Parimal Bag, Debraj Ray, Vijay Krishna, Syed Nageeb Ali, Ashutosh Varshney, Co-Pierre Georg, Sourav Bhattacharya, participants at the summer meeting (2014) of the Econometric Society at the University of Minnesota, participants at the 10^{th} Annual Conference on Economic Growth and Development at the Indian Statistical Institute (New Delhi), participants at the 33^{rd} Australasian Economic Theory Workshop at Deakin University, seminar participants at IIM Bangalore, participants at the inaugural Society for Economics Research conference at the Indian School of Business (Hyderabad) and seminar participants at The Pennsylvania State University for their comments. An early version of this paper was circulated under the title *Urban Ethnic Conflicts* and more recently under the title *Ethnic Conflicts, Rumours and an Informed Agent*. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ethnicity, she is unable to communicate credibly with the other ethnicity. Despite this, we show that peace can be achieved in equilibrium. Our result explains how organizations trying to prevent conflict by dispelling false rumours and fake news could be effective *even* if they are perceived to be biased towards a specific group. (JEL - D74, D83, P16, D82)

Keywords - Ethnic Conflicts, Cheap Talk, Multiple Audiences, Private Signals, Payoff Externalities

Introduction

We live in a world where 'fake news', rumours and biased rhetoric are far too common. This can have disastrous consequences. Often, an ethnic conflict[1] is preceded by several small events and false news stories which stoke the fire between ethnic groups. There have been many incidents which show that this kind of misinformation can cause conflicts to erupt. For example, in Kenya's Tana delta about 170 people lost their lives to conflict between 2012 and 2013, many of which were precipitated by rumours such as - a Pokomo (a primarily agrarian people) health worker tried to poison rather than vaccinate Orma (cattle-herding nomadic people) children[2]. In other instances, rumours spread via whatsapp caused lynchings and conflict[3]. Could these conflicts have been prevented if the correct information was disseminated? While this appears to be a simple solution, the implementation is problematic as we run into the question of how to convey the information *credibly*.

In particular, suppose there exists an 'informed agent' who has information which could prevent the ethnic conflict if everyone knew it. If the informed agent belongs to one of the ethnicities, the other ethnicity may take a jaundiced view of any information conveyed to them by her which suggests them to remain peaceful. They might think that the informed agent simply wants her own ethnicity to win the conflict and in order to gain advantage she is attempting to restrict the number of the other ethnicity players who fight. Thus, if the informed agent is

[1] All conflicts based on ascriptive group identities (race, language, religion, tribe, or caste) can be called ethnic (Horowitz (1985)).

[2] See http://buildingpeaceforum.com/2015/02/una-hakika-preventing-rumors-and-violence-in-kenyas-tana-delta/

[3] See https://www.nytimes.com/2017/06/21/opinion/whatsapp-crowds-and-power-in-india.html.

known to be biased towards one ethnicity, she will not be able to communicate credibly with the other ethnicity. One might think that therefore, an informed agent will be ineffective in preventing conflict once misinformation has ignited tensions.

The contribution of this paper is to analyze the role of informed agents in preventing conflict. We show that even if the informed agent is known to be biased towards one ethnicity, peace can still be an equilibrium outcome. This is key to understanding the success of projects like the 'Una Hakika' project in Kenya's Tana delta[4]. Upon hearing potentially conflict inducing information, people can text the same to an organization who will verify its veracity and respond with their findings. This can reduce the occurrence of conflict by giving people correct information *before* they react to news[5]. Another example of a similar organization is 'Hoax Slayer' (see Facebook.com/SMHoaxSlayer and http://smhoaxslayer.com/), an Indian website and Facebook page which debunks fake viral stories on social media, and Hoaxmap.org, which aims to collate and refute rumours about offences allegedly committed by migrants in Germany. However, these organizations faces the same problems we raise here. If one group thinks that the members of the organization are biased towards the other group, can these initiatives be effective? We discuss the lessons from our analysis for such organizations in more detail in section 5.

In our model, there are two ethnicities, each with a continuum of members. There are two states of the world - one in which peace is possible, and another in which conflict is inevitable. The society is on the brink of conflict since the prior belief about the state of the world is that the latter state is more likely. The informed agent knows the state of the world but the players themselves do not. Before the players decide on whether they want to fight or not fight, the informed agent can send private cheap talk messages containing information about the state of the world to each and every player of both ethnicities. This is akin to members of the Una Hakika project sending messages to people to dispel or confirm a rumour. The informed agent's preferences are such that she prefers peace if it is possible but in the state in which it is not, she

[4]See https://www.unahakika.org/.

[5]For details look at - *Using Cell Phones To Stop False Rumors, Before They Lead To Ethnic Violence*- http://www.fastcoexist.com/3029321/using-cell-phones-to-stop-false-rumors-before-they-lead-to-ethnic-violence, April 29, 2014.

wants her own ethnicity to win. Thus, she is biased towards her own ethnicity. If the true state of the world is the one in which peace is possible *and* a large enough fraction of both ethnicity players choose to not fight, conflict can be averted.

We show that the biased informed agent cannot communicate effectively with the opposite ethnicity. This is because she always has the incentives to convince them to not fight, which renders her messages uninformative in equilibrium. Despite this limitation, peace is an equilibrium outcome. There are two key ideas here. One, though the informed agent sends uninformative signals to the opposite ethnicity, she is able to communicate credibly with her own ethnicity and this is common knowledge. Thus, the presence of an informed agent allows the players of the opposite ethnicity to evaluate their action choices with the knowledge that the players of the informed agent's ethnicity will condition their play on the true state. Without the informed agent, neither group can condition their action on the state. Two, the informed agent is only partially biased towards her own ethnicity - in one state the informed agent wants the outcome (winning the conflict) which only benefits her ethnicity, but in the other state the informed agent's preferred outcome (peace) is one that favours both ethnicities. These two ideas allow us to define an equilibrium in which players of the opposite ethnicity realize that the lack of information does not allow them to launch a coordinated attack while the other ethnicity is fully coordinated. This reduces their chance of winning the conflict (and therefore their payoff from fighting), and they find it optimal to not fight, and hope that the state is good (where the informed agent implements peace).

Theoretically, our paper contributes to the literature on cheap talk games with multiple audiences with the novel addition of payoff externalities along with private signals. Allowing for private signals distinguishes our paper from those concerned with cheap talk games and public signals like Levy and Razin (2004), Baliga and Sjöström (2012). Private signals are important if the informed agent communicates in person or via whatsapp/text messages[6]. Furthermore, unlike private signals, a public signal may allow the informed agent to communicate effectively with the opposite ethnicity because the informed agent cannot lie to the opposite side without

[6]See https://www.unahakika.org/ (informed agent communicates via messages) and https://www.nytimes.com/2017/06/21/opinion/whatsapp-crowds-and-power-in-india.html (information conveyed via whatsapp messages cause conflict).

also lying to her own ethnicity. Payoff externalities are key to any analysis of conflict since one side's actions may have severe repercussions for the other side. Thus, this paper differs from those which allow for private signals but do not have payoff externalities like Farrell and Gibbons (1989), Goltsman and Pavlov (2011). Our paper differs from those in the mediation literature like Kydd (2003) and Kydd (2006) because our mediator's preferences are dependent upon her information (state dependent preferences), and because in our model, the mediator can achieve peace *without* being truthful to one ethnicity.

1 Literature

Our paper is primarily related to the literature on cheap talk games with multiple audiences, and the literature on mediation. Our model and results also have a flavour of global games and models of strategic information disclosure.

With respect to the literature on cheap talk games with multiple audiences, two features make our paper novel - we allow for the possibility of private communication with both audiences (ethnicities) *and* we have payoff externalities in our model as well. In the literature on cheap talk games with multiple audiences and public signals, the papers most closely related to ours are Baliga and Sjöström (2012) and Levy and Razin (2004)[7]. These papers have payoff externalities between the two audiences but they only allow for public signals. Private signals are important if the informed agent communicates in person or via whatsapp/text messages[8]. Also, the informed agent may not have access to the media to announce information publicly. This will be particularly true if vested interests want the conflict to happen. Furthermore, unlike private signals, a public signal may allow the informed agent to communicate effectively with the opposite ethnicity because the informed agent cannot lie to the opposite side without also lying to her own ethnicity. For example, Levy and Razin (2004) show that in a democracy, though the leader has an incentive to misrepresent her information to the rival country and reveal the

[7]There is also a literature on leadership in which the leader uses a public signal to coordinate on a desired equilibrium. See Ahlquist and Levi (2011) for a survey of this work.

[8]See https://www.unahakika.org/ (informed agent communicates via messages) and https://www.nytimes.com/2017/06/21/opinion/whatsapp-crowds-and-power-in-india.html (information conveyed via whatsapp messages cause conflict).

correct information to the home audience, she will not be able to do so because public signals are observed by the rival country as well. Allowing for a private signal takes away this power to send credible signals. Despite this limitation, we show that it may be enough that the informed agent can communicate effectively with her own ethnicity to obtain peace in equilibrium.

Two related papers in the cheap talk games with multiple audiences literature which allow for some degree of private communication are Farrell and Gibbons (1989) and Goltsman and Pavlov (2011). Farrell and Gibbons (1989) consider a cheap talk environment with one sender and two receivers. However, while we consider a game where the informed agent (sender) communicates with *both* ethnicities (receivers) simultaneously and privately, the games considered in Farrell-Gibbons either include private communication with a single receiver or public communication with both. Additionally, the action chosen by one receiver does not influence the utility of the other receiver i.e there are no payoff externalities. This is in contrast to our model where coordination within and across ethnicities is important (payoff externalities). Goltsman and Pavlov (2011) is closely related to Farrell-Gibbons (1989) and they allow for private communication with *both* receivers. However, they also do not have payoff externalities.

In the literature on mediation, Kydd (2006) and Kydd (2003) emphasizes the fact that an effective mediator must find it incentive compatible to reveal her information truthfully. In the case of Kydd (2006), this occurs when the mediator is relatively moderate (if the mediator strongly prefers peace or is biased in favour of one of the disputants then she will have incentives to lie). On the other hand, truth telling is optimal only for the biased mediator in Kydd (2003)[9]. The idea here is that a side will only believe a mediator who share preferences with it. Our paper differs from these in two main ways. One, the mediator's preferences in Kydd (2003) and Kydd (2006) are independent of their private information whereas the informed agent in our model has state dependent preferences. Furthermore, we show that the informed agent can achieve peace despite being biased. This is clearly in contrast to Kydd (2006) and different from Kydd (2003) because in our model, the informed agent is able to achieve peace without being truthful to one of the parties. This is possible because the informed agent is partially biased towards one party - in one state the informed agent wants the outcome which only favours

[9]Also see Regan (2002).

that party, but in the other state the informed agent's preferred outcome is one that favours both parties.

In the information disclosure literature, Rauchhaus (2006) shows that a third party mediator can effectively avoid war if the mediator possesses private information about one of the disputant's capabilities. In the paper, the mediator's preferences are independent of her information whereas in our paper, the informed agent's preferences are state dependent. Furthermore, in Rauchhaus (2006) the mediator can send a signal to only one of the agents, whereas in our model the informed agent can send a signal to each and every member of both the groups[10]. Kamien, Tauman and Zamir (1990) also study the value of information in conflict. However, the informed agent there is unbiased, and their focus is on the set of outcomes the informed agent can implement as a unique Nash equilibrium. In contrast, we ask if the presence of the informed agent makes peace an equilibrium outcome. Furthermore, in our paper the informed agent is biased so she cannot commit to a disclosure strategy (unlike in Kamien, Tauman and Zamir (1990)), and this also creates the problem of conveying her information credibly. Other papers in a similar vein are Egorov and Sonin (2014) and Shadmehr and Bernhardt (2015).

On the global games front, Chen, Lu and Suen (2016) assess the impact of circulation of rumours on regime change by studying a coordination game under a global game structure with both public and private signals. In our paper, the informed agent can strategically send private signals to all players. In contrast, in Chen, Lu and Suen (2016), everyone receives *exogenous* private signals about the rumour. Few players have additional information about the state, and the ones that do are likely to use this information to their advantage. Thus, the assumption that the signals are exogenous seems untenable to us. Tyson and Smith (2017) study a two-sided coordination problem in a global games environment. They highlight how public information influences coordination within a group, and between groups. In Tyson and Smith (2017), the groups have different first best outcomes, and only one group may have their preferred outcome in equilibrium whereas in our model, both ethnicities prefer peace over other

[10]In a previous version of our paper, we show that if the probability of the informed agent existing is low, and the informed agent can send signals to only one ethnicity, then conflict cannot be avoided. This is because, in that case, not receiving a signal is most likely because the informed agent does not exist. In this case the ethnicity which does not receive any signal always fights.

outcomes and therefore there are inter-group coordinating incentives. Also, Tyson and Smith (2017) have both groups receiving exogenously generated signals. In our model, the strategic informed agent has the ability to send a different signal to each and every player. This plays an important role since it allows the informed agent to send less informative signals to the opposite ethnicity and even misinform some players of her own ethnicity.

Finally, in the literature on ethnic conflicts, much has been written about its causes. Miguel, Satyanath and Sergenti (2004) relate the occurrence of ethnic conflicts to economic shocks using data from 41 african countries[11], while Reynal-Querol (2002) suggests that religious divisions are more important than language divisions and natural resources to explain ethnic conflicts. Esteban and Ray (2008) point out that ethnic conflict may be more likely to occur than class conflict where there is within-group economic inequality. Esteban and Ray (2011) use a theoretical model to show how within-group heterogeneity in radicalism and income help in precipitating ethnic conflicts. In contrast, our paper studies the role of information in preventing conflicts which can be precipitated by misinformation. We abstract away from discussing the underlying cause of conflict itself.

2 Model

There are a continuum of players and each player can belong to one of two ethnicities - $\{E_1, E_2\}$. Each ethnicity has the same mass of players. This is a simplifying assumption. Our results would go through even when the ethnicities are not symmetric in size. The ethnicity of each player is common knowledge. Additionally, every player can be one of two types - *Good* (G) or *Bad* (B). The two types differ in terms of the actions available to them. Players can decide to fight (f) or not fight (nf). The G type player is strategic and can choose either action. This type of player can be interpreted as a citizen who will only fight if he perceives it to be the best course of action in the game. B type players on the other hand, are behavioural and always fight. One can think of the B type players as players who are motivated by elements

[11] Also see Bohlken and Sergenti (2010).

outside the game to always fight[12].

Let $c \in (0,1)$ be an exogenously given threshold. Denote A_i as the fraction of players from ethnicity E_i who choose to fight. A conflict will occur if and only if at least one group has $A_i > c$. Conditional on the conflict happening, probability of winning for any group i is given by $A_i/(A_i + A_j)$. The threshold c is common knowledge. Thus, we assume that ethnic conflict only occurs if a significant fraction of at least one ethnic group chooses to fight.

At the beginning of the game, players are uncertain about the distribution of types in the world. Let $n^y{}_l$ be the fraction of y ethnicity players who are l type. For simplicity, we assume that there are only two possible type distributions[13]. With probability ω the type distribution is such that $(n^{E_1}{}_G, n^{E_2}{}_G) = (q,q)$, and with probability $(1-\omega)$ the type distribution is such that $(n^{E_1}{}_G, n^{E_2}{}_G) = (r,r)$, where $(1-q) < c < (1-r)$. Thus, if (r,r) is the true distribution of G types, then the number of bad types alone is so high that conflict must happen (because bad types always choose to fight and their fraction is above the threshold required for conflict). On the other hand, if (q,q) is the true distribution of types then conflict may not happen if a large enough fraction of the G types choose not to fight. If the true distribution of types is (q,q) then we will refer to it as the good state of the world and if the true distribution of types is (r,r), we will call it the bad state of the world. Uncertainty about the state of the world represents the apprehensions and the fears in the minds of people when instigating rumours are floating about. Mathematical details about the type space and the prior distribution of types can be found in the appendix. Here on, unless otherwise stated, everything is described for only the G type player. This is because the B type player is behavioural with fixed actions.

The payoffs to any player i of type G depends on a) his action, b) whether or not conflict takes place and c) whether he was part of the winning or losing side if conflict did take place. The payoffs are summarized precisely in Table 1. $\alpha, \beta, \gamma, \delta, \varepsilon > 0$ in table 1. *CW* refers to the event where conflict happens and own ethnicity wins, *CL* - conflict happens and own ethnicity loses, and *NC* means no conflict occurs. The entire payoff matrix is common knowledge

[12]For example, these could be individuals who are acting under the influence of political parties, or people with a vested interest in conflict. A player's type is private knowledge of the player.

[13]This assumption is not crucial to our results. In particular we could have assumed positive weights on a multitude of distribution states and as long as conflict is inevitable in some states and not in others, our claims will go through.

amongst all players. We assume that $\varepsilon < \alpha + \beta$. This is to ensure that payoff from fighting and winning is better than payoff from fighting and losing.

Essentially, only two aspects of the payoff table are important for our results. One, peace time payoffs $(\alpha + \delta)$ are higher than the best conflict payoff (α) for the players. There are many costs associated with a conflict like the loss of lives, collateral damage, an atmosphere of uncertainty, apprehension and animosity. Therefore, we feel that that our assumption that war is never more desirable than peace is not unjustified. Two, the payoffs are such that it always pays to fight when conflict is inevitable. This can arise naturally in a society where players who don't fight for their ethnicities are subsequently ostracized/punished by their own communities. Thus, such ex-post social costs to not fighting may outweigh any private costs to fighting, especially since this cost may have to be suffered by not just the people who did not fight but also by their families (for possibly many generations). Other authors (example Egorov and Sonin (2014)) have justified this sort of assumption by a 'warm glow' a player might experience by participating along with his community in a fight against an enemy.

The other payoffs are chosen just for simplicity. For example, in case of a conflict, it is *not* crucial that the player receives the same payoff from not fighting regardless of whether his ethnicity won or not. If a player chooses to fight and conflict does not happen, we assume that that player's payoff is negative. This may be interpreted as the cost of getting arrested for unruly behaviour in public.

Table 1: Payoffs

	CW	CL	NC
f	α	$-\beta+\varepsilon$	$-\gamma$
nf	$-\beta$	$-\beta$	$\alpha+\delta$

Next, we describe the informed agent and the actions available to her. Often, there exist players who have additional information about the state of the world. For example, think of ethnic conflicts induced by rumours. An informed agent could be one who knows if the rumour is actually true or false. The veracity of the rumour is likely to be correlated with the state of the world. We model this by assuming that there exists an 'informed agent' (whom we

denote by b) who is perfectly[14] informed about the state of the world. The informed agent chooses which signal to send to each and every player in the environment. Given a player i in the population, she can take one of two actions: 1. Send a letter with the signal Q. This action is denoted by (LQ), 2. Send a letter with the signal R. This action is denoted by (LR). Thus, the informed agent b has the ability to send private cheap talk messages to all players. The contents of the letter serve as a signal of the state of the world. As a real life analogy, in the 'Una Hakika' example given in the introduction, the informed agent could be seen as the organization which tries to learn the truth about the rumour, and then send private messages to people to dispel/confirm the rumour.

Like all other players, let b have an ethnicity from the set $\{E_1, E_2\}$. However, we assume that b is outside the population and does not herself fight or not-fight in the conflict. This is just for simplicity of calculations. The same results will go through if b is thought to be a player in the population, albeit with messier maths. Without loss of generality, we will assume that b has ethnicity E_1. This is common knowledge.[15] Since b is not part of the population, she only gets the payoffs from outcomes. If conflict happens and $b's$ ethnicity (E_1) wins, b gets α. If conflict happens and $b's$ ethnicity loses, she gets $(-\beta)$. If conflict does not happen, b gets $\alpha + \delta$. Thus, b gets maximum payoff if conflict does not happen. However, if conflict does happen then she would like her own ethnicity to win. In the appendix, we consider an alternate utility function for the informed agent where we study the case of an extremist informed agent who prefers conflict if her ethnicity wins over a peaceful outcome - see section A.4.

The timeline of events is as follows. At time 0, players have priors on the true distribution of types i.e. about the state of the world. The informed agent sends a private signal to each and every player and then the players decide their action simultaneously.

We focus only on strategies of the informed agent that are symmetric within ethnicity. This does not mean that all people of the same ethnicity will receive the same message in

[14]Note that the informed agent doesn't need to know the state perfectly for our results to go through, she just needs to have sufficiently good information.

[15]This can be justified because letter sending is supposed to represent a meeting process. People know (specially if the player b is one of the players in the population) or can guess the ethnicity of others by observing the name, clothes and 'look' of the person. In the 'Una Hakika' example, people may know or have beliefs about the ethnic composition of the organization.

equilibrium. This is because we allow b to play mixed strategies. Thus, we allow the informed agent to send different signals to different members of the same ethnicity as well. The strategy of the informed agent is a function of the ethnicity of the receiving player and the true state. For simplicity, we will denote letters sent to the opposite ethnicity (opposite from b, i.e. E_2 ethnicity) as LQ^d, LR^d where superscript 'd' stands for different and same ethnicity as LQ^s, LR^s (superscript 's' is for same). Since b has been assumed to be of ethnicity E_1 if she exists, her strategy can be described by the following functions:

$$f_b : \{E_1\} \times \{(q,q),(r,r)\} \to \Delta\{LQ^s, LR^s\}$$
$$f_b : \{E_2\} \times \{(q,q),(r,r)\} \to \Delta\{LQ^d, LR^d\}$$

We will assume that players play symmetric (within ethnicity) strategies only. Thus, players who are of the same type and who are at the same information set will play the same strategies. Strategy for any player i of the population is a function from his information set to the action set $\Delta\{f, nf\}$. Let g^{E_i} denote the strategy of a player of ethnicity E_i. Then:

$$g^{E_1} : \{LQ^s, LR^s\} \to \Delta\{f, nf\}$$
$$g^{E_2} : \{LQ^d, LR^d\} \to \Delta\{f, nf\}$$

Players update beliefs in a Bayesian manner and they choose actions which are optimal given beliefs. Thus, our equilibrium concept is Perfect Bayesian Equilibrium.

3 No Informed Agent

First, suppose that the informed agent does not exist. The key result here is that if players are sufficiently pessimistic i.e. they believe that it is the bad state of the world with high probability, then conflict is inevitable. Note that any strategy profile where all agents always choose the action f, constitutes an equilibrium[16]. We will call this the *all-fight* equilibrium.

[16]Our payoff matrix is such that fighting is the best response if conflict is inevitable (as will be the case when everyone chooses to fight).

We show that there is a threshold for the belief (which we denote as ω^*) about the good state of the world, such that if the prior belief about the state being good is below this threshold then *all-fight* is the unique equilibrium. On the other hand, if the prior is above the threshold, peace is an equilibrium outcome in the good state. We state the lemma formally below and leave the proof for the appendix.

Lemma 1. *There exists an ω^* such that $\forall\ \omega < \omega^*$, there exists a unique equilibrium in which all players choose to fight and, $\forall\ \omega \geq \omega^*$ there is an equilibrium in which all G type players choose to not fight, thereby ensuring peace if the state is good. This equilibrium is the highest payoff equilibrium for the G type players if $\alpha + \varepsilon + 2\gamma < \beta$.*

The formal proof is rather simple so we relegate it to the appendix (see section A.2). This result is not a surprise. Conflict is inevitable in the bad state of the world and the best response to conflict is to fight. Thus, if the players place sufficient weight on the bad state of the world, *all fight* is the natural equilibrium outcome. When players are more optimistic ($\omega \geq \omega^*$), they can coordinate and achieve the high peace equilibrium payoff.

In this paper, we want to understand the role of an informed agent in preventing conflict once tensions are high. To this end, we will take pessimistic beliefs ($\omega < \omega^*$) as given for the rest of this paper, so that without the informed agent conflict is inevitable. The challenge for us is to determine the conditions needed so that the informed agent's cheap talk messages are able to achieve peace, despite the fact that the informed agent is known to be biased towards her own ethnicity.

4 Strategic Informed Agent

4.1 Equilibrium

In this section, we investigate the nature of equilibria if the informed agent exists. When the belief about the state of the world is pessimistic to begin with ($\omega < \omega^*$), we want to determine if there are equilibria in which there is a positive probability of conflict being averted.

4.1.1 Truth telling

First, we argue there cannot exist a symmetric (within ethnicities) strategy profile (apart from playing fight to all signals) where the informed agent fully reveals her private information. This is because she will always have an incentive to deviate and lie to the opposite ethnicity. We express this observation as a proposition below:

Proposition 1. *There does not exist any symmetric equilibrium (different from all-fight) where the informed agent's strategy is truth telling i.e. the informed agent's strategy is:*

$$f_b(E_1, Q) = LQ^s$$
$$f_b(E_2, R) = LR^d$$
$$f_b(E_1, Q) = LQ^s$$
$$f_b(E_2, R) = LR^d$$

Proof. Suppose not i.e. let's suppose that truth telling by the informed agent can be an equilibrium strategy in an equilibrium which is not *all fight*. Consider an agent of the opposite ethnicity $i \in E_2$. If he receives the message LR^d, then he knows that the state of the world is (r,r) and hence conflict will occur with probability one. This is because b's signal is perfectly informative and she always reveals her information truthfully. Hence, everyone from the opposite ethnicity chooses to fight when the message LR^d is received.

If i receives LQ^d, then it cannot be the case that the action nf (not fight) is played with positive probability. If this were the case, then, when the informed agent knows that the state is bad, she would deviate to the message LQ^d to maximize the probability of her ethnicity winning by reducing the number of opposite ethnicity players who fight (if she sends the signal LR^d, every player fights). Hence, the opposite ethnicity players always fights making conflict inevitable. The same ethnicity players, knowing that conflict cannot be avoided, would also always choose to fight. This is a contradiction to the assumption that the equilibrium being played was different from the all-fight equilibrium. □

The intuition here is simple. Since the informed agent is biased towards her own ethnicity,

she has the incentive to lie to the opposite ethnicity in the bad state to prevent some of them from fighting, and give her own ethnicity a higher probability of winning the conflict. Knowing this, the opposite ethnicity players should not believe the informed agent's signals in equilibrium.

4.1.2 Non-truth telling

The previous subsection highlights the problem faced by the biased informed agent. Since it is known that she is biased towards one ethnicity, the players of the other ethnicity realize that she has incentives to lie to them and this makes effective communication with them very difficult. This reduces the informed agent's power to change the outcome of the game when the belief about the state is pessimistic (below ω^*). However, in this subsection we will show that despite this limitation, we can obtain peace as an equilibrium outcome. This result is relevant for understanding the effectiveness of projects like 'Una Hakika'. A deeper discussion of the implications of our results for organizations like Una Hakika and Hoax Slayer is presented in section 5. The intuition for the conditions needed for a peaceful equilibria follows.

Suppose the players of the opposite ethnicity believe that very few of their own ethnicity players are going to fight in equilibrium. Then, the returns from fighting are low because they are likely to lose in case of a conflict[17]. On the other hand, they could play 'not fight', in which case if the state happens to be the good state of the world, they can get the high peace equilibrium payoff. The opposite ethnicity players reason that peace may prevail in the good state because a) The informed agent has the incentive to try and avoid conflict (to get the high peace time payoff) when the state is good b) the informed agent can communicate effectively with her own ethnicity and stop them from fighting. This can ensure peace as an equilibrium outcome if the state is actually good. However, for the above play to be optimal, the players of the opposite ethnicity must place sufficient weight on the event that the state is good. If their prior beliefs are too pessimistic (ω is much lower than ω^*) then it will always be optimal for them to fight since they realize that when the state is bad, the informed agent cannot avoid conflict. Thus, we must have a lower bound on the prior belief of the players about the good

[17]Remember that the probability of winning a conflict is increasing in the fraction of own ethnicity players who choose to fight.

state (q,q) i.e. we must have lower bounds on ω (so ω can be below ω^*, but not too far below). When very few players of their own ethnicity fight, it will be optimal for the opposite ethnicity players to not fight (thereby making their own beliefs about very few of their own ethnicity players fighting correct), and bet on the possibility that the state is good so that peace may prevail. We present the equilibrium strategies formally in proposition 2.

The equilibrium strategy for the informed agent is to send uninformative signals about the state to the opposite ethnicity players. However, the presence of the informed agent allows the opposite ethnicity players to know what actions the players of the same ethnicity will play in different states. We go on to show that this is enough to get peace as an equilibrium outcome in the good state of the world[18]. The informed agent sends informative signals about the state to her own type.

Proposition 2. *There exists $\underline{\omega}$, $\bar{\omega}$ such that if $\omega \in (\underline{\omega}, min\{\bar{\omega}, \omega^*\})$, then there exists a perfect Bayesian equilibrium in the class of strategies described below for a unique p_d.*

b's strategy : (1)

$f_b(E_1,(r,r)) = LR^s$

$f_b(E_1,(q,q)) = LQ^s$

$f_b(E_2,(r,r)) = q_b^R LR^d + (1-q_b^R)LQ^d$

$f_b(E_2,(q,q)) = q_b^Q LR^d + (1-q_b^Q)LQ^d$

Player's strategies

E_1 *ethnicity/Same ethnicity*

$g^{E_1}(LQ^s) = nf$

$g^{E_1}(LR^s) = f$

E_2 *ethnicity/Opposite ethnicity*

$g^{E_2}(LQ^d) = p_d f + (1-p_d)nf$

$g^{E_2}(LR^d) = p_d f + (1-p_d)nf$

where $0 < p_d \leq z$, $q_b^R = q_b^Q \in [0,1]$, & *where z is such that* $zq + (1-q) = c$

Proof. Consider a player of the same ethnicity : $i \in E_1$. Since player i is of the same ethnicity as b, receiving the message LQ^s perfectly reveals to him that the true state of the world is (q,q).

[18] In effect, peace is like a correlated equilibrium outcome.

In this case, given the strategies of others, he knows that all G types from E_1 ethnicity will choose to not fight and a proportion p_d of G types from E_2 ethnicity will choose to fight, but this is not enough to start a conflict. Thus, peace will be the outcome and i's optimal strategy is to play nf. So, the agent's response to LQ^s is optimal. Similarly, the same ethnicity player knows that a signal LR^S implies that the state must be bad and so conflict is inevitable. In this case, the payoff matrix tells us that it is optimal for the player to fight.

Let us now discuss the optimality of b's strategy. Consider first the case that the state is (r,r) i.e conflict cannot be avoided. Her optimal response is then to maximise the probability of her ethnicity winning, which is achieved by persuading all from her own ethnicity to fight (she does this by sending them all the signal LR^s) and dissuading as large a proportion of the opposite ethnicity from fighting as possible. Given the strategy of the opposite ethnicity players, this can be achieved by randomly sending each player either LQ^d or LR^d. Now, suppose that the state is (q,q). The informed agent would prefer that conflict be averted. She can enforce no conflict by adhering to the strategy prescriptions (a fraction $p_d q + (1-q)$ from the opposite ethnicity fight and a fraction $1-q$ of the same ethnicity players fight, but both fractions are less than c[19]).

We have so far shown optimality of the strategies for E_1 ethnicity players and b. We now show optimality for agents of the opposite ethnicity (E_2). In particular, it will be important to find conditions under which the randomization p_d is optimal. Define the function:

$g : [0,z] \to \mathbb{R}$ *such that*

$$g(p) = \omega(-\gamma) + (1-\omega)[\frac{p+(1-r)}{p+r+2(1-r)}(\alpha) + \frac{r+(1-r)}{p+r+2(1-r)}(-\beta+\varepsilon)] - [\omega(\alpha+\delta) + (1-\omega)(-\beta)]$$

Thus, the function g shows the difference in payoffs for the opposite ethnicity players from playing f and nf, when a fraction p of the good type players in the opposite ethnicity players are going to play fight, and the same ethnicity players and the informed agent follow the

[19] $p_d q + (1-q) < c$ because $p_d < z$.

equilibrium strategy. It is easy to show that:

$$\omega > \underline{\omega}(= \frac{(1-r)(\alpha+\beta)+\varepsilon}{(1-r)(\alpha+\beta)+\varepsilon+(\alpha+\delta+\gamma)(1+1-r)}) \Rightarrow g(0) < 0$$
$$\omega < \bar{\omega}(= \frac{(\alpha+\beta)(z+1-r)+\varepsilon}{(\alpha+\beta)(z+1-r)+\varepsilon+(\alpha+\delta+\gamma)(z+1-r+1)}) \Rightarrow g(z) \geq 0$$

Pick $\omega \in (\underline{\omega}, min\{\bar{\omega}, \omega^*\})$ so that the above is satisfied[20]. Thus, we have that $g(0) < 0$ and $g(z) \geq 0$. Also, since g is strictly increasing and continuous, by the intermediate value theorem, there exists a unique $p_d \in (0, z]$ such that $g(p_d) = 0$. □

This class of strategies has the following desirable property: the informed agent can successfully avoid conflict when the state of the world is good. If she knows that the state of the world is bad, she is able to prevent some of the opposite ethnicity from engaging in conflict, thereby providing her own ethnicity with an advantage. Also, note that when the belief about the good state of the world is below ω^*, then all fight is the unique equilibrium in the game without an informed agent (lemma 1). However, the presence of the informed agent changes this result, and allows for a peaceful outcome even with such pessimistic beliefs (in the range $(\underline{\omega}, min\{\bar{\omega}, \omega^*\})$).

Proposition 2 demonstrates that *even if* the informed agent is considered biased towards one ethnicity, her presence can allow for peaceful outcomes because the presence of the informed agent allows the opposite ethnicity players to realize that the same ethnicity players will play for peace in the good state and for conflict in the bad state. This is because the informed agent has the incentive to truthfully reveal the state to her own ethnicity. Thus, while the opposite ethnicity players get no information about the state in equilibrium, they are able to condition their actions based on the knowledge of the state dependent play of their rivals. When there is no informed agent, the action choice of their rivals is not state-dependent.

In the proof, the upper and lower bounds on the belief of players (ω) are chosen so that, given the equilibrium strategies of the informed agent and the same ethnicity players, the opposite ethnicity players want to play not-fight when none of their own good types want to fight and they want to play fight when a fraction z or above of their own good type players want

[20] $\varepsilon < \alpha+\beta \Rightarrow \underline{\omega} < min\{\bar{\omega}, \omega^*\}$. Thus, there exists an ω which satisfies $\underline{\omega} < \omega < min\{\bar{\omega}, \omega^*\}$.

to fight. This is possible due to the increasing (decreasing) returns from fighting when more players of own (other) ethnicity choose to fight and, this gives us an intermediate point p_d such that if a fraction p_d of the opposite ethnicity players are playing fight then the opposite ethnicity players are indifferent between playing fight and not-fight.

Finally, we would like to point out that it is not necessarily the case that the informed agent sends completely uninformative signals to members of the opposite ethnicity. In the appendix we show that there is an equilibrium in which the players of the opposite ethnicity play pure strategy not fight (see section A.3). In this case, we show that the informed agent is indifferent between giving no information and a very small amount of information which will keep 'not-fight' optimal for the opposite ethnicity. We call this the 'barely informative' equilibrium.

5 Lessons for the Una Hakika Example

In the introduction and elsewhere in the paper we have spoken about the 'Una Hakika' project in Kenya's Tana delta, where, upon hearing potentially conflict inducing information, people can text the same to the organization who will verify its veracity and respond with their findings. This can reduce the occurrence of conflict by giving people correct information before they react to news. We have also mentioned other similar organizations like 'Hoax Slayer', an Indian website and Facebook page which debunks fake viral stories on social media, and Hoaxmap.org, which collates and refutes false rumours about offences allegedly committed by migrants in Germany. These organizations face the same problem we raise here. If one ethnic group thinks that the members of the organization are biased towards the other group, can these initiatives be effective?

Repeated game theorists may argue that there are two reasons why these organizations may not face any issues in conveying information. One, in a long run repeated game, the opposite ethnicity players can punish the informed agent for lying in any period with a simple 'trigger strategy' wherein they play fight in all periods after they discover that the informed agent gave them incorrect information. If the probability of a good state is high enough in any period, it may be optimal for the biased informed agent to reveal the true signal in every period when

faced with such a trigger strategy. Two, if the organization is actually not biased, then this information will be learnt over time by the players as the organization builds a reputation for unbiased reporting via truthful messages, and in the long run (when their reputation is high enough), the organization may face no problems in conveying information credibly.

However, what of the short run? The damage that a few ethnic conflicts can do is enough to warrant a study on how to prevent such conflicts from occurring in the short run as well. Consider the Una Hakika example. When the project was first initiated in Kenya, people did not know if this organization will remain in the area for a long time. This may have lead people to believe that the organization may not have the kind of incentives needed for a long run truth telling equilibria as described above. Additionally, in the short run, the organization will not have the time to build a reputation for being unbiased.

Despite these difficulties in conveying information credibly in the short run, in a survey conducted in 2015 (about 2 years into the project), people were asked - How much has Una Hakika helped prevent the spread of rumours? The average score given (on a subjective scale) was 8.53 out of 10, where 10 was the best score possible. This was in spite of the fact that the organization did not score as high on the question - How neutral and impartial do you feel Una Hakika is? They got a score of 7.43[21] of 10 on this question[22]. Thus, the organization was largely effective in preventing conflict even while some uncertainty remained about their impartiality.

What explains this early success of the project? Our paper points to one explanation. We show that even if the organization is perceived to biased towards one ethnicity, it can achieve peace. This is possible because even though the players of the opposite ethnicity may disbelieve any information given to them directly, they can condition their actions on the knowledge that the informed agent has the incentives to reveal the correct information at least to the ethnicity towards which it is biased.

[21]The survey seems to suggest that several people thought of the organization as unbiased while a small group thought of them as biased. The difficulty of effective communication under these beliefs is not easy to interpret since these are subjective answers. However, our result is important because we show that *even if* the organization was thought of as completely biased, we could still get peaceful outcomes in equilibrium.

[22]See https://thesentinelproject.org/2015/04/28/una-hakika-users-vote-on-value-of-misinformation-management/

Finally, in proposition 2, we show that the presence of an informed agent can make peaceful outcomes possible even in the range of beliefs where it was impossible to achieve peace without an informed agent (below ω^*). However, we note that there is bound to the biased informed agent's effectiveness. She can only prevent conflict if the belief about the good state is not too far below ω^* (not below $\underline{\omega}$). This is because if the belief about the bad state is high enough, then the biased informed agent will be helpless in preventing conflict. Thus, while our result demonstrates the usefulness of informed agents, it also points out an upper bound to the effectiveness of biased informed agents. This is the reason why organizations like Una Hakika try very hard to establish a reputation for being unbiased[23]. It is easy to show[24] that an unbiased informed agent can be effective even if the beliefs are extremely pessimistic, because an unbiased informed agent does not face the issue of conveying information credibly.

6 Conclusion

Rumours, 'fake news' and propaganda can lead to misinformation induced conflicts. Governments and private organizations[25] try to prevent charged environments from boiling over by providing the correct information. However, since one ethnic group may perceive them to be biased towards their rivals, they face the issue of conveying information *credibly*.

If effective communication with one ethnicity is not possible, then this may lead one to believe that a biased informed source has no hope of preventing conflict once misinformation has ignited tensions in a society. Our paper shows that this is not the case. In a simple model, we demonstrate that there exists an equilibrium with a peaceful outcome even when it is common knowledge that the informed agent is biased towards one ethnicity. Furthermore, our paper demonstrates that there could be multiple equilibria (apart from the peaceful equilibrium, there exists another equilibrium where conflict occurs). The presence of multiple equilibria is another way to explain why some areas remain peaceful while other (similar) areas are destroyed by ethnic conflicts. Thus, it is possible that informed agents are able to prevent conflict in some

[23]"... community trust is an essential facet to the program's success" - https://thesentinelproject.org/2015/04/28/una-hakika-users-vote-on-value-of-misinformation-management/.

[24]Proof available on request.

[25]Like the 'Una Hakika' project, 'Hoaxslayer' website, Hoaxmap.org.

instances, but are ineffective in others.

This paper is a step towards understanding the role of informed players in preventing conflicts. There can be very interesting extensions of this paper. One can look at a repeated environment where a new signal may arrive every period and one informed agent receives it. It will be useful to understand the dynamics in such an environment, specially if the informed agent has reputation concerns. We could also look at an environment where the informed agent can choose the portfolio of the people she meets i.e. given her capacity constraint, she can choose exactly what fraction of the players she meets are from either community. In such an environment, what is the optimal portfolio choice and the equilibrium strategy? Finally, our paper is silent on when the peaceful equilibria will be selected over the all-fight one. This presents yet another research avenue. There are many such important and interesting questions which we hope to investigate in the future.

References

Ahlquist, John S and Margaret Levi. 2011. "Leadership: What it means, what it does, and what we want to know about it." *Annual Review of Political Science* 14:1–24.

Baliga, Sandeep and Tomas Sjöström. 2012. "The strategy of manipulating conflict." *The American Economic Review* 102(6):2897–2922.

Berenschot, Ward. 2011. "The spatial distribution of riots: Patronage and the instigation of communal violence in Gujarat, India." *World Development* 39(2):221–230.

Bhavnani, Ravi, Michael G Findley and James H Kuklinski. 2009. "Rumor dynamics in ethnic violence." *The Journal of Politics* 71(03):876–892.

Bohlken, Anjali Thomas and Ernest John Sergenti. 2010. "Economic growth and ethnic violence: An empirical investigation of Hindu-Muslim riots in India." *Journal of Peace research* 47(5):589–600.

Brass, Paul R. 2011. *The production of Hindu-Muslim violence in contemporary India*. University of Washington Press.

Carlsson, Hans and Eric Van Damme. 1993. "GLOBAL GAMES AND EQUILIBRIUM SELECTION'." *Econometrica* 61(5):989–1018.

Chen, Heng, Yang K Lu and Wing Suen. 2016. "The Power of Whispers: A Theory of Rumor, Communication, and Revolution." *International economic review* 57(1):89–116.

Crawford, Vincent P and Joel Sobel. 1982. "Strategic information transmission." *Econometrica: Journal of the Econometric Society* pp. 1431–1451.

Diamond, Douglas W. 1985. "Optimal release of information by firms." *The Journal of Finance* 40(4):1071–1094.

Dutta, Souvik. 2014. "Ethnic Conflict and Civic Engagememt." *Unpublished Manuscript, Indian Institute of Management Bangalore* .

Dye, Ronald A. 1985. "Disclosure of nonproprietary information." *Journal of accounting research* pp. 123–145.

Edmond, Chris. 2013. "Information Manipulation, Coordination, and Regime Change." *The Review of economic studies* 80(4):1422–1458.

Egorov, Georgy and Konstantin Sonin. 2014. Incumbency advantage in non-democracies. Technical report National Bureau of Economic Research.

Enikolopov, Ruben, Maria Petrova and Ekaterina Zhuravskaya. 2011. "Media and Political Persuasion: Evidence from Russia." *The American Economic Review* pp. 3253–3285.

Esteban, Joan and Debraj Ray. 2008. "On the Salience of Ethnic Conflict." *American Economic Review* 98(5):2185–2202.

Esteban, Joan and Debraj Ray. 2011. "A model of ethnic conflict." *Journal of the European Economic Association* 9(3):496–521.

Farrell, Joseph and Robert Gibbons. 1989. "Cheap talk with two audiences." *The American Economic Review* 79(5):1214–1223.

Fearon, James D and David D Laitin. 1996. "Explaining interethnic cooperation." *American political science review* 90(04):715–735.

Goltsman, Maria and Gregory Pavlov. 2011. "How to talk to multiple audiences." *Games and Economic Behavior* 72(1):100–122.

Grossman, Sanford J. 1981. "The informational role of warranties and private disclosure about product quality." *Journal of law and economics* pp. 461–483.

Hörner, Johannes, Massimo Morelli and Francesco Squintani. 2015. "Mediation and peace." *The Review of Economic Studies* 82(4):1483–1501.

Horowitz, Donald L. 1985. *Ethnic groups in conflict*. Univ of California Press.

Judd, Kenneth L. 1985. "The law of large numbers with a continuum of iid random variables." *Journal of Economic theory* 35(1):19–25.

Jung, Woon-Oh and Young K Kwon. 1988. "Disclosure when the Market is Unsure of Information Endowment of Managers." *Journal of Accounting Research* 26(1):146–153.

Kamenica, Emir and Matthew Gentzkow. 2011. "Bayesian Persuasion." *American Economic Review* 101(6):2590–2615.

Kamien, Morton I, Yair Tauman and Shmuel Zamir. 1990. "On the value of information in a strategic conflict." *Games and Economic Behavior* 2(2):129–153.

Kosfeld, Michael. 2005. "Rumours and markets." *Journal of Mathematical Economics* 41(6):646–664.

Kydd, Andrew. 2003. "Which side are you on? Bias, credibility, and mediation." *American Journal of Political Science* 47(4):597–611.

Kydd, Andrew H. 2006. "When can mediators build trust?" *American Political Science Review* 100(3):449–462.

Larson, Jennifer M. 2013. "Interethnic Conflict, Incendiary Rumors, and the Networks that Help or Hurt.".

Levy, Gilat and Ronny Razin. 2004. "It takes two: an explanation for the democratic peace." *Journal of the European economic Association* 2(1):1–29.

Martin-Shields, Charles and Elizabeth Stones. 2014. "Smart phones and social bonds: Communication technology and inter-ethnic cooperation in Kenya." *Journal of Peacebuilding & Development* 9(3):50–64.

Miguel, Edward, Shanker Satyanath and Ernest Sergenti. 2004. "Economic shocks and civil conflict: An instrumental variables approach." *Journal of political Economy* 112(4):725–753.

Milgrom, Paul R. 1981. "Good news and bad news: Representation theorems and applications." *The Bell Journal of Economics* pp. 380–391.

Mitra, Anirban and Debraj Ray. 2014. "Implications of an economic theory of conflict: Hindu Muslim violence in India." *Journal of Political Economy* pp. 719–765.

Morris, Stephen and Hyun Song Shin. 1998. "Unique equilibrium in a model of self-fulfilling currency attacks." *American Economic Review* pp. 587–597.

Morris, Stephen and Hyun Song Shin. 2001. "Global games: theory and applications.".

Morris, Stephen and Hyun Song Shin. 2002. "Social value of public information." *The American Economic Review* 92(5):1521–1534.

Petrova, Maria. 2008. "A Formal Theory of Public Opinion in Conflicts." *Available at SSRN 1149367* .

Rauchhaus, Robert W. 2006. "Asymmetric information, mediation, and conflict management." *World Politics* 58(02):207–241.

Regan, Patrick M. 2002. "Third-party interventions and the duration of intrastate conflicts." *Journal of Conflict Resolution* 46(1):55–73.

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

Shadmehr, Mehdi and Dan Bernhardt. 2015. "State censorship." *American Economic Journal: Microeconomics* 7(2):280–307.

Tyson, Scott A and Alastair Smith. 2017. "Dual-Layered Coordination and Political Instability: Repression, Co-optation, and the Role of Information." *The Journal of Politics* 80(1).

Vaitla, Srinivas. 2011. Preventing ethnic violence with local capacities: lessons from civil society in India PhD thesis Rutgers University-Graduate School-Newark.

Varshney, Ashutosh. 2003. *Ethnic conflict and civic life: Hindus and Muslims in India*. Yale University Press.

Varshney, Ashutosh and Steven Wilkinson. 2006. "Varshney-Wilkinson Dataset on Hindu-Muslim Violence in India 1950-1995 Version 2.".

Yanagizawa-Drott, David. 2014. "Propaganda and conflict: Evidence from the Rwandan genocide." *The Quarterly Journal of Economics* 129(4):1947–1994.

A Appendix

A.1 Type space and prior

Denote as $T = \{G,B\}^N$ the set of all type profiles. Define $T_q = \{t : \mu(\{i \in E_1 : t_i = G\}) = \mu(\{i \in E_2 : t_i = G\}) = q\}$ and similarly define T_r. We endow T with the appropriate sigma algebra such that the sets of the form T_q and T_r are measurable and we assume that the prior $p \in \Delta(T)$ has the following properties: a) $p(T_q \cup T_r) = 1$, b) $\forall i \in N$, $p(T_q|t_i = G) = \omega(< \omega^*)$, c) $p(t_i = G|T_s) = s \; \forall \; i \in N$ and $\forall s \in \{q,r\}$. The construction of such priors has been discussed in Judd (1985). We may do so here by separately performing Judd's construction for T_q and T_r and then naturally extend the measure to the union $T_q \cup T_r$. The first condition says that the type distribution is either (q,q) or (r,r). The second condition says that when an agent learns

that he is of type G, his belief about (q,q) is ω which is less that ω^*. Third, conditional on T_s, the probability of each player being a good type is s.

A.2 Results: No Informed Agent

Proof of lemma 1

Proof. First, we want to show that if ω is high enough then it will be optimal for the G players to not fight, given that other G players are playing nf. Consider the strategy profile where all good type players (irrespective of ethnicity) play nf. An arbitrary G player will make the following calculations

Payoff from playing $f = \omega(-\gamma) + (1-\omega)(\frac{\alpha-\beta+\varepsilon}{2})$

Payoff from playing $nf = \omega(\alpha+\delta) + (1-\omega)(-\beta)$

Clearly, if $\omega \geq \frac{\alpha+\beta+\varepsilon}{\alpha+\beta+\varepsilon+2(\alpha+\delta+\gamma)}$, then playing nf is best response for G player. So this strategy profile constitutes a Bayesian Nash equilibrium if $\omega \geq \omega^* = \frac{\alpha+\beta+\varepsilon}{\alpha+\beta+\varepsilon+2(\alpha+\delta+\gamma)}$.

Next, we want to show that all players playing fight is the only equilibrium if $\omega < \omega^*$. It is trivial to check that all players playing f is a Nash equilibrium for all levels of beliefs. Therefore, we skip this and focus on uniqueness. We will prove this by contradiction. Suppose $\omega < \omega^*$ and there is an equilibrium such that players of at least one ethnicity play nf with strictly positive probability. Suppose the players play according to the following strategy profile:

$$E_1 \ plays \ - \ \ p_1(nf) + (1-p_1)f$$
$$E_2 \ plays \ - \ \ p_2(nf) + (1-p_2)f$$

Case 1 - $p_1 \neq p_2$.

WLOG, let $p_2 > p_1$. This implies that $p_2 > 0$ and $p_1 < 1$. p_1 cannot be equal to zero, else the best response for the E_2 ethnicity will be to play f with probability one but that would imply $p_2 = 0$. This is a contradiction. Thus, we have that $p_1 \in (0,1)$ i.e. players of ethnicity 1 are indifferent between the action fight and not fight.

Subcase 1 - $p_2 = 1$. In this case, for the E_1 ethnicity players to be indifferent between fight and not fight, we need the condition that the payoff from fighting is equal to the payoff from not fighting. Thus we have,

$$\omega(-\gamma) + (1-\omega)\left(\frac{(1-r+(1-p_1)r)\alpha}{1-r+(1-p_1)r+(1-r)} + \frac{(1-r)(-\beta+\varepsilon)}{1-r+(1-p_1)r+(1-r)}\right) = \omega(\alpha+\delta) + (1-\omega)(-\beta)$$

It can be easily checked that the ω which solves the this expression is above ω^*. However, we started with the case that $\omega < \omega^*$. So, this is a contradiction.

Subcase 2 - $p_2 < 1$. In this case, we must have that both ethnicities are indifferent between the two actions. However, it is easy to check that we cannot have common priors and have two symmetric ethnicities be simultaneously indifferent when mixing with different probabilities (since $p_2 \neq p_1$).

Case 2 - $p_1 = p_2$.

Subcase 1 - $p_1 = p_2 = 0$. This is not possible since we want an equilibrium in which players of at least one ethnicity play not fight with positive probability.

Subcase 2 - $p_1 = p_2 = 1$. By definition of ω^*, we know that in this case, there is a profitable deviation in switching to fight for any arbitrary player.

Subcase 3 - $p_1 = p_2 \in (0,1)$. In this case, players of both ethnicity are indifferent between fight and not fight and equal fractions of both ethnicity are playing fight. We can show quite easily that for mixing to be optimal, we need $\omega = \omega^*$. This is a contradiction because we started with $\omega < \omega^*$.

Thus, there is no other equilibrium when $\omega < \omega^*$ and therefore in this case, conflict is inevitable. Next we show - If $\omega > \omega^*$, then all players playing nf (not-fight) is the payoff dominant equilibrium for the G players.

Expected payoff from this equilibrium = $\omega(\alpha+\delta) + (1-\omega)(-\beta)$.[26] There is only one other equilibrium possible in pure strategies - an equilibrium in which both G types and B types play f. Payoff from this all fight equilibrium = $\frac{\alpha-\beta+\varepsilon}{2}$.

[26]We only consider the expected payoffs of the G type when thinking of Payoff dominance. Since the B types are always choosing to fight, clearly they are at least indifferent to the result of their actions.

It is easy to see that if $\omega > \omega^*$ and $\alpha + \varepsilon + 2\gamma < \beta$, then the ex-ante expected payoff from all fight equilibrium is lower than payoff from equilibrium in which G players don't fight. Let us check to see if there are any mixed strategy equilibria. First, we need this claim:

Claim 1. *In any mixed strategy equilibrium where the G types of both ethnicities play the same strategies, the weight on playing f has to be less than or equal to $c-(1-q)$.*

Proof. We will prove by contradiction. Suppose the players of any ethnicity play f with a strictly higher weight than $c-(1-q)$. Then the fraction of players playing f for that ethnicity is higher than c in any state of the world. This implies that conflict is inevitable. However, when conflict is inevitable then playing f is strictly dominant strategy. Thus, the ethnicities could not be mixing between f and nf. Contradiction. □

Consider the strategy where all the G players are playing fight with probability p where $p \leq c-(1-q)$ (this must hold else conflict is inevitable and p will have to be equal to 1). For mixing to be optimal, the payoff from f must be equal to the payoff from nf.

Payoff from playing $f = \omega(-\gamma) + (1-\omega)(\frac{\alpha-\beta+\varepsilon}{2})$

Payoff from playing $nf = \omega(\alpha+\delta) + (1-\omega)(-\beta)$

If the above payoffs are the same then we have: $\omega = \omega^*$

Payoff from this mixed strategy equilibrium = $\omega^*(\alpha+\delta) + (1-\omega^*)(-\beta)$

Since the ethnicities are symmetric, in any mixed strategy equilibrium, the G players of both ethnicities will play the same strategies. Suppose the G players of E_1 ethnicity were playing f with probability p_1 and the G players of E_2 were playing f with probability p_2 where $p_1 \neq p_2$. We can see quite easily from the above proof that a necessary condition for the players of E_1 ethnicity to mix is that $\omega = \omega_1$ and the E_2 ethnicity requires $\omega = \omega_2$ for them to mix in equilibrium where $\omega_1 \neq \omega_2$. Thus, an asymmetric mixed equilibrium is not possible. There is however a hybrid equilibrium where one ethnicity play pure strategy nf (not fight) and the other ethnicity mixes between fight and not fight (easily follows from Case 1, subcase 1 in the proof of unique equilibrium when $\omega < \omega^*$). In this case, since all players see nf as an optimal action, the payoff for all G type players is $\omega(\alpha+\delta) + (1-\omega)(-\beta)$ which is the same as the

payoff from the equilibrium in which all players play nf.

Comparing ex ante expected payoffs in the four possible equilibria, it is obvious now that if $\omega > \omega^*$, then the equilibrium in which all G players play nf is the highest payoff equilibrium (along with the hybrid equilibrium). □

A.3 Barely Informative Equilibrium with Strategic Informed Agent

In this equilibrium, players of the opposite ethnicity play pure strategy nf along equilibrium path i.e. they do not fight. This equilibrium points out that it is not necessarily the case that b sends completely uninformative signals to members of the opposite ethnicity. If the opposite ethnicity players are choosing the pure strategy nf then the informed agent is indifferent between giving no information and a very small amount of information which will still make make nf optimal for the opposite ethnicity. We describe this equilibrium next.

Proposition 3. *There exists $\underline{\omega}$ such that if $\omega \in (\underline{\omega}, \omega^*)$, then the following profile of strategies constitute an equilibrium :*

b's strategy : (2)

$$f_b(E_1,(r,r)) = LR^s$$

$$f_b(E_1,(q,q)) = LQ^s$$

$$f_b(E_2,(r,r)) = q_b^R LR^d + (1-q_b^R)LQ^d$$

$$f_b(E_2,(q,q)) = q_b^Q LR^d + (1-q_b^Q)LQ^d$$

Player's strategies

E_1 *ethnicity/Same ethnicity*

$$g^{E_1}(LQ^s) = nf$$

$$g^{E_1}(LR^s) = f$$

E_2 *ethnicity/Opposite ethnicity*

$$g^{E_2}(LQ^d) = nf$$

$$g^{E_2}(LR^d) = nf$$

$$q_b^R, q_b^Q \in [0,1]$$

Proof. Pick the same specification for $\underline{\omega}$ as in proposition 2 i.e. let $\underline{\omega} = \frac{(1-r)(\alpha+\beta)+\varepsilon}{(1-r)(\alpha+\beta)+\varepsilon+(\alpha+\delta+\gamma)(1+1-r)}$.

The proposition assumes that $\omega > \underline{\omega}$. This assumption guarantees that the prior beliefs of the players are not so pessimistic that the players choose to fight irrespective of how the other players are playing. In particular, it tells us that if all other players follow their equilibrium strategies, then it is optimal for the good type players of the opposite ethnicity to not fight if no other good type player from the opposite ethnicity is fighting (see proof of proposition 2).

Given an $\omega(> \underline{\omega})$, choose any $\varepsilon > 0$ such that $\omega' > \underline{\omega}$ holds for any $\omega' \in (\omega - \varepsilon, \omega + \varepsilon)$. Now define $Pr((q,q)|LQ^d; q_b^Q, q_b^R)$ and $Pr((q,q)|LR^d; q_b^Q, q_b^R)$ be the posteriors of the agents of the opposite ethnicity about the state (q,q) conditional on information given by the letters LQ^d and LR^d under the signal structure (q_b^Q, q_b^R). Now one can find informative signals (q_b^Q, q_b^R) such that both $Pr((q,q)|LQ^d; q_b^Q, q_b^R)$,$Pr((q,q)|LR^d; q_b^Q, q_b^R) \in (\omega - \varepsilon, \omega + \varepsilon)$. We now confirm that this is an equilibrium.

We show that the opposite ethnicities response to LR^d and LQ^d are as stated. Under both signals, the posteriors of the agent are in the interval $(\omega - \varepsilon, \omega + \varepsilon)$ and hence are greater than $\underline{\omega}$. Hence playing nf is strictly better at both LQ^d and LR^d. Notice that this signal structure is also optimal for player b. In any state of the world she would want as few of the opposite ethnicity to participate[27] and the suggested strategy achieves that objective. It can be checked that the incentives of the players at other information sets are also optimal. Hence the above specification is an equilibrium. □

Corollary 1. *b's messages to the opposite ethnicity in the strategies described in proposition 3 are barely informative about the state of the world.*

Proof. By barely, we mean that b is indifferent between sending completely uninformative signals and signals which contain so little information that opposite ethnicity players still want to play pure strategy nf. The proof follows from $q_b^Q \neq q_b^R$. □

[27] If the state is good she is indifferent between a small fraction of the opposite ethnicity players fighting (small enough to not induce conflict) and none of them fighting.

A.4 Other payoff types for the Informed Agent

Hitherto, we have used a utility specification for the informed agent which describes her as peace-loving player with a bias towards her own ethnicity. In this section we show that this specification is not necessary for our results to go through. In particular, even if player b prefers conflict if her own ethnicity wins to peace, and prefers peace to a conflict if her own ethnicity loses, we can still get the same equilibria as before. The only additional condition we need is that payoff from peace be above a cut off for player b.

Formally, let $u_b(CW)$ be the payoff to player b if conflict happens and her own ethnicity wins. Similarly, $u_b(CL)$ is the payoff to player b if conflict happens and her own ethnicity loses, and $u_b(NC)$ is the payoff to her when conflict does not happen. Our results up to this point have a used a utility specification where $u_b(NC) > u_b(CW) > u_b(CL)$. Consider the following payoff type for agent b. The informed agent's preference satisfies: $u_b(CW) > u_b(NC) > u_b(CL)$. This payoff type is interpreted as follows: agent b resembles the mindset and payoff specification of an extremist who would prefer conflict to peace but only as long her own ethnicity wins. The following result provides equilibrium possibilities with this payoff specification.

Proposition 4. *Let* $p_1 = \frac{1}{(1-q)+1}, p_2 = \frac{1}{(1-q+qp_d)+1}$. *Assume the conditions required for proposition 2 and proposition 3. Then the following hold:*

1. *If* $p_1 u_b(CW) + (1-p_1)u_b(CL) < u_b(NC)$, *then the strategies outlined in both proposition 2 and proposition 3 constitute equilibria.*

2. *If* $p_2 u_b(CW) + (1-p_2)u_b(CL) < u_b(NC) < p_1 u_b(CW) + (1-p_1)u_b(CL)$, *then the strategies outlined in proposition 2 constitute an equilibrium but the strategies described in proposition 3 do not.*

3. *If* $u_b(NC) < p_2 u_b(CW) + (1-p_2)u_b(CL)$ *then all fight is the unique equilibrium.*

Proof. We will prove this one by one for each of the points above.

1. Consider the strategies outlined in proposition 2 and 3. Notice that in both, optimality of strategy for players of either ethnicity is satisfied given the behaviour of the agent

b. Consider strategies outlined in proposition 3 and consider the incentives of b. If b knows that it is the bad state of the world, she would like to maximise the probability of winning for her own ethnicity, and this is achieved by her strategy. When b knows that it is the good state of the world, she can either avert conflict (which her strategy proposes) or deviate and induce conflict. This would gives her a probability of winning equal to $\frac{1}{1+1-q} = p_1$. Then, inducing conflict gives utility $p_1 u_b(CW) + (1-p_1)u_b(CL)$ and following her prescribed strategy gives utility $u_b(NC)$. Now:

$$Deviation\ payoff\ = p_1 u_b(CW) + (1-p_1)u_b(CL) < u_b(NC) = Strategy\ Payoff \quad (3)$$

Hence, averting conflict is better. Since $p_2 < p_1$, the same argument works for the strategies described in proposition 2 as well.

2. A similar argument as above gives the result.

3. The fact that the strategies described in proposition 2 and 3 are not equilibrium strategies any more follows from the arguments made above. We show that the unique equilibrium is *all fight*. Suppose not. Then, there is an equilibrium where conflict is averted when the state is good. In this case, b gets $u_b(NC)$. However, she can deviate and ensure everyone from her own ethnicity plays fight. It is easy to show that p_d is the highest fraction of opposite ethnicity players who fight in any equilibrium which results in peace in the good state. Thus, the informed agent can always secure the payoff $p_2 u_b(CW) + (1-p_2)u_b(CL)$ by inducing conflict. Since this is greater than $u_b(NC)$, the deviation is strictly profitable.

□