H i C N **H**ouseholds **i**n **C**onflict **N**etwork

The Institute of Development Studies - at the University of Sussex - Falmer - Brighton - BN1 9RE

www.hicn.org

Reputation, Group Structure and Social Tensions

Dominic Rohner*

dr527@york.ac.uk

HiCN Working Paper 40

January 2008

Abstract: Social tensions impede social cohesion and public goods provision. They can also be a driving force for more serious conflicts such as civil wars. Surprisingly, however, the emergence of social tensions has only rarely been studied in the literature. In the present contribution a game-theoretic model highlights how reputation concerns and the structure of group cleavages matter for the emergence of social tensions. In particular, the respective effects of fractionalisation, polarisation and segregation are assessed. The predictions of the model can account for recent empirical evidence on ethnic conflicts. The framework can also be applied to the study of social capital and merger failures.

Keywords: Conflict, Information, Reputation, Ethnicity, Social Capital

JEL Codes: C73, D74, L14, Z13

Acknowledgements: I would like to thank Partha Dasgupta and Karl Ove Moene for especially helpful suggestions. I am also grateful to Samuel Bowles, Robert Evans, Michael Findley, Masayuki Kudamatsu, John Miller, David Myatt, Scott Page, Rajiv Sethi, Christopher Wallace, Jörgen Weibull, Diego Winkelried and Elizabeth Wood for their helpful comments. As well, useful discussions with conference and seminar participants in Amsterdam, Santa Fe NM, Columbus OH, Chicago, London, Berlin, Oslo, New York, Paris, Oxford, York and Cambridge are gratefully acknowledged. An earlier version of this paper has been circulated under the title "Information, Reputation and Ethnic Conflict".

* Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, United Kingdom.

Reputation, Group Structure and Social Tensions

Dominic Rohner
University of York and University of Cambridge

November 18, 2007

Abstract

Social tensions impede social cohesion and public goods provision. They can also be a driving force for more serious conflicts such as civil wars. Surprisingly, however, the emergence of social tensions has only rarely been studied in the literature. In the present contribution a game-theoretic model highlights how reputation concerns and the structure of group cleavages matter for the emergence of social tensions. In particular, the respective effects of fractionalisation, polarisation and segregation are assessed. The predictions of the model can account for recent empirical evidence on ethnic conflicts. The framework can also be applied to the study of social capital and merger failures.

JEL Classification: C73, D74, L14, Z13.

Keywords: Conflict, Information, Reputation, Ethnicity, Social Capital.

1 Introduction

Social tensions are important for several reasons. They have a *direct* impact on social welfare by threatening social cohesion and by impeding collective goods provision and economic performance. Their *indirect* impact, however, is even more worrisome. Social tensions are, together with the feasibility of collective action, a necessary condition for the break out of more serious forms of conflict, such as civil wars (cf., Collier and Hoeffler, 2004; 2007, for a related discussion).

[0] *Address:* Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, United Kingdom, dr527@york.ac.uk.

Acknowledgements: I would like to thank Partha Dasgupta and Karl Ove Moene for especially helpful suggestions. I am also grateful to Samuel Bowles, Robert Evans, Michael Findley, Masayuki Kudamatsu, John Miller, David Myatt, Scott Page, Rajiv Sethi, Christopher Wallace, Jörgen Weibull, Diego Winkelried and Elizabeth Wood for their helpful comments. As well, useful discussions with conference and seminar participants in Amsterdam, Santa Fe NM, Columbus OH, Chicago, London, Berlin, Oslo, New York, Paris, Oxford, York and Cambridge are gratefully acknowledged. An earlier version of this paper has been circulated under the title "Information, Reputation and Ethnic Conflict".

Surprisingly, the emergence of social tensions has only received very little attention in the literature. In the present contribution I would like to address this gap and show with the help of a game-theoretic model how information, reputation effects and the group structure of a society can favour or hinder the emergence of social tensions. The source of tensions in our framework will be disputed economic relations and the impact of fractionalisation, polarisation and segregation on the reputation cost of defection will be examined.

These concepts are defined in the following way. A highly *fractionalised society* is defined as one with a great number of divisions and of distinct groups. Following Montalvo and Reynal-Querol (2005, p.797), a highly *polarised society* is defined as a society "where a large minority group faces a majority group". Polarisation is greatest if a society consists of two groups of equal size, and is smallest if a society consists of one homogenous group. *Segregation* is defined as the extent to which different groups in the society are kept separate. More formal definitions of these concepts are included in the sections four and five.

Our model of social tensions is flexible enough to capture different forms of social tensions for which reputation and group structure play important roles. One natural and particularly important application of the model is ethnic conflict, as the structure of social tension is the driving force behind this form of conflict. The predictions of the model about the emergence of social tensions and grievances is complementary to the existing (formal) literature on ethnic conflict that focuses on collective action issues and the supply of hate-inducing actions by politicians (see for example, Glaeser, 2005; Esteban and Ray, 2006a; Caselli and Coleman, 2006). The model can also contribute to the literature on social capital and merger failures.

The remainder of the paper is organised as follows. Section 2 will be devoted to a discussion of the related literature and applications of the model, while in section 3 a basic model of peaceful versus conflicting economic interactions is built for a homogenous society. In section 4, group structure will be introduced in the model, and the impact of polarisation and segregation will be assessed. The model will be extended to n-groups in section 5 and the effects of fractionalisation will be studied. Section 6 concludes.

2 Related Literature and Applications

First, we shall discuss the literature that is related to the building blocks of the model, then the focus will lie on applications of the model, i.e. ethnic conflict, social capital and merger failures. The modelling framework of the present contribution builds up on the literature on commitment, reputation and contract enforcement in trade and business (see Greif, 1993; Greif, Milgrom and Weingast, 1994; Tirole, 1996; Dixit, 2003)[1].

[1] There is also a related literature in evolutionary biology focusing on how reputation and image scores and indirect reciprocity have favoured the emergence of cooperation (e.g., Nowak and Sigmund, 1998).

The literature on trade and conflict is also relevant for the present paper, which derives conditions under which peaceful trade (i.e. cooperation) prevails over appropriative activities (i.e. defection). In political science there is a large empirical literature on the "liberal peace" (see Polachek, 1980; Oneal and Russett, 1999; Gartzke, Li, and Boehmer, 2001)[2]. More recently, economists have shown interest in the interaction between trade and conflict, concluding that the overall impact of trade is ambiguous and depends on the price effects (Skaperdas and Syropoulos, 2001) and the bilateral versus multilateral nature of trade (Martin, Mayer and Thoenig, 2006).

We shall now discuss three applications of the model.

2.1 Ethnic Conflict

Often civil wars occur along ethnic lines. In countries as diverse as Rwanda, Sudan, Guatemala or Angola, ethnicity has played a major role in the breakout of hostilities.

There is a growing literature about ethnicity and civil wars. The empirical evidence is not conclusive so far. Fearon and Laitin (2003) and Collier and Hoeffler (2004) find that ethnic fractionalisation does not increase the risk of civil war outbreaks. By contrast, Vanhanen (1999), Sambanis (2001) and Collier, Hoeffler and Rohner (2006) conclude, using other data sources and measures, that ethnic fractionalisation increases the risk of civil wars.[3] Cederman and Girardin (2007) and Bates (1999) explain the occurrence of conflict with ethnonationalist exclusiveness, respectively one group's fear of permanent political exclusion.

Reynal-Querol (2002) and Montalvo and Reynal-Querol (2005) find that what drives ethnic conflict is not fractionalisation, but polarisation.

Ethnic segregation has also received considerable attention in the literature. Some scholars have claimed that segregation increases the risk of ethnic conflict (Diez Medrano, 1994; Olzak, Shanahan, and McEneaney, 1996), while others have argued that segregation, taking the form of "partition", could be a solution to ethnic conflict (Horowitz, 1985)[4].

Given the inconclusiveness of the empirical literature, it seems important to show theoretically through which channels ethnic cleavages could matter. Surprisingly, however, theoretical studies building formal models of ethnicity and conflict have been scarce. Esteban and Ray (1999) develop a behavioral model of contest and bargaining between groups according to the distribution of certain

[2] Most contributions in this literature find that trade relationships between countries can reduce the scope of inter-state war by increasing the long-run gains from economic cooperation.

[3] A nonlinear impact of ethnicity on conflict could explain the contradictory findings of the empirical literature. Horowitz (1985) has performed a comparative analysis and has found that for fully homogenous and for fully heterogenous societies the risk of ethnic conflict is small, whereas the risk is greater for fewer big groups confronting each other. Using cross-country evidence, Collier and Hoeffler (1998) have come to a similar conclusion, namely that intermediate levels of fractionalisation are the most risky.

[4] Sambanis (2000) concludes, using cross-country evidence, that partition does not significantly prevent conflict occurrence.

characteristics. In more recent papers (2006a; 2006b), the same authors focus on ethnic mobilisation and rent-seeking and on the question why in a society with class and ethnic cleavages, the latter tend to be more salient than the former[5]. Also Robinson (2001) enquires under what conditions ethnic conflict is more or less salient and destructive than class conflict. Another formal model addressing ethnic conflict has been built by Caselli and Coleman II (2006). They study the interaction between coalitions of different groups, whereas ethnicity increases the risk of conflict by enforcing coalition membership. Fearon and Laitin (1996) emphasise intra-group enforcement of group members' cooperation with players from other ethnic groups. It is shown that policing inside a group can ensure peaceful relations outside the group.[6]

All these papers focus on aggregate players ("interest groups") rather than on individual players, and do not treat the respective effects of ethnic polarisation, fractionalisation and segregation.

The present contribution would like to address the shortcomings of the existing literature on ethnic conflict by building a theoretical model in which the group structure of a society matters through the channels of information and reputation, and that is able to assess the impact of ethnic polarisation, segregation and fractionalisation on the likelihood of social tensions between ethnic groups. Rather than focusing on coalition building and contests between aggregate groups, as has been done in previous studies, I will emphasise the social tensions on the level of individual players.

2.2 The Emergence of Social Capital

In recent years scholars in economics have developed a strong interest in the impact of social aspects of society on economic performance. These underlying characteristics of a society have often been referred to as "social capital". There is some ambiguity about what is actually meant by "social capital" (cf. Dasgupta, 2005), but most studies link it to trust, civic participation and social networks. These aspects have been shown to have an important positive impact on economic and social outcomes (Putnam, 1995; Knack and Keefer, 1997; Temple and Johnson, 1998; Francois and Zabojnik, 2005).

Knowing that social capital matters, scholars have started to empirically study the factors that can favour or hinder its emergence. A key finding is that racial and ethnic cleavages and heterogeneity can result in several (generally) undesirable social and civic outcomes associated with more social tensions and disputes and lower levels of social capital. In particular, a more fractionalised and polarised group structure has been shown to result in less civic participation (Alesina and La Ferrara, 2000, Vigdor, 2004), less trust (Alesina and La Ferrara,

[5] In another interesting paper Esteban and Ray (2006c) treat the contrasting effects of fractionalisation and polarisation on conflict onsets and intensity. They do not include ethnicity and reputation concerns in their model. Rather, fractionalisation, polarisation, as well as political institutions matter by affecting the cost of conflict for different groups in a collective action framework.

[6] The creation and impact of ethnic identities is another important topic in the literature about ethnic conflict (Basu, 2005; Sen, 2006).

2002) and lower levels of public goods provided (Alesina, Baqir and Easterly, 1999; Luttmer, 2001; Miguel and Gugerty, 2005; Lind, 2007).

The present contribution will explain theoretically how group cleavages, i.e. fractionalisation, polarisation and segregation, favour or hinder the emergence of social tensions, which can to many respects be regarded as the opposite force to social cohesion and social capital, decreasing trust and civic collaboration. Thus, our theory of social tensions hopes to contribute to the study of forces impeding the building of social capital.

2.3 Merger failures

Social tensions among employees have been shown to result in a suboptimal performance of companies (Rob and Zemsky, 2002). This is especially relevant for mergers, as in many mergers of companies the post-merger profitability is lower than before. This at first sight surprising outcome can be explained by "cultural conflict" between employees of the companies merged (Weber and Camerer, 2003). Fulghieri and Hodrick (2006) and Banal-Estanol, Macho-Stadler and Seldeslachts (2006) have modeled internal conflict after mergers, but do not take reputation and group size effects into account.

The predictions of our model of social tensions are able to account for interesting findings of this literature on "merger failures". For example, the result (derived in proposition 4) that maximum polarisation increases the scope for social tensions is consistent with the finding that "mergers of equals" of two similar sized companies seem to be especially problematic[7].

3 The Basic Model

In what follows I build a model of how reputation and information matter for determining whether the economic interaction between players is characterised by "defection" or "cooperation"[8]. The concepts of "defection" and "social tension" are linked in the following way.

Definition 1 *Social tension* $\equiv$ *"number of matches with defection" divided by "total number of matches".*

The more players defect, the higher is the level of social tensions.

3.1 Strategies, Payoffs, Information

The following assumptions are made:

[7] Cf. Weber and Camerer (2003) for a discussion of the (largely unsuccessful) merger between Daimler-Benz and Chrysler. Agrawal, Jaffe and Mandelker (1992) provide empirical evidence that firms do on average especially poorly when acquiring "preys" that are large relative to themselves (although the relationship is non-monotonic).

[8] These concepts are more formally defined under assumption G.2 below.

G.1 - General setting: The game lasts for an infinite number of periods. Players discount the future and take into account that, with some probability, they will "die" in a given future period. The players who die are replaced by newly born players. There is a large number of players who match randomly.

G.2 - Actions: First, players choose between entering into contact with the opponent or staying out. If players enter, they select the level of the variable F (referring to "fighting effort"), where $F \in [0, 1]$. Although we allow for intermediate levels of F, it follows from the specification of the payoff function (as shown below) that the variable F always takes the extreme values 0 (which we call "cooperation") or 1 (which we call "defection").

G.3 - Payoff function: In all periods all players receive a payoff of 0 if they stay out. If they enter into contact with their match, they have the payoff function displayed in equation (1)[9].

$$V_i = (\frac{1}{2} + \theta(\rho F_i - \psi F_j))S - cF_i - gF_j \tag{1}$$

where i,j=players, θ=parameter capturing the decisiveness of fighting effort (with $0 \leq \theta \leq 0.5$), ρ=parameter indicating the fighting technology (ability) of player i ($0 \leq \rho \leq 1$), F=level of fighting effort ($0 \leq F \leq 1$), ψ=fighting technology of player j ($0 \leq \psi \leq 1$), S=economic gains (surplus) from interaction, c=parameter related to the cost of player i's fighting effort, and g=parameter measuring player i's cost inflicted by the fighting effort of player j.

The total economic gains of the interaction S are multiplied by the linear difference-form contest success function[10], $(\frac{1}{2}+\theta(\rho F_i-\psi F_j))$, which refers to the share that player i receives of the gains. The relative share of player i depends linearly on the differences in fighting effort between the players. The shares of both players sum up to 1.

The parameter θ measures the decisiveness of the differences in the fighting effort between the two players. If $\theta = 0$, both opponents receive half of the surplus S, independently of their fighting effort. By contrast, if $\theta = 1$, the level of fighting has a strong impact on the distribution of S. Further, the parameters ρ and ψ reflect the fighting technology of the two players. $\rho = 0$ indicates a total inefficient fighting technology of player i, where an increased fighting effort of i does not result in his obtaining a greater share. $\rho = 1$ corresponds to a very efficient fighting technology of i. It is analogous for player j's fighting technology ψ.

Player i's payoff function (1) also includes the parameters c and g which relate to the cost of his own fighting effort, respectively the destruction inflicted by the opponent's fighting effort.

[9] The payoff function of player j is analogous: $V_j = (\frac{1}{2} + \theta(\psi F_j - \rho F_i))S - cF_j - gF_i$.

[10] "Contest success functions" relate fighting efforts to the share of a "prize" received by a particular player (see Hirshleifer, 1989; Skaperdas, 1996). The present contest success function, $(\frac{1}{2} + \theta(\rho F_i - \psi F_j))$, is similar to the one used in Rohner (2006), although the present contribution introduces the fighting technology differently.

G.4 - Types: There are two types of players who differ in their fighting ability ρ. The players referred to as "strong" ("weak") have $\rho = \alpha$ ($\rho = \beta$), where $\alpha > \beta$. A proportion p of the population are assumed to be "strong" types.

G.5 - Information: i) The players are incompletely informed about the type of the other players. All other features of the game such as the form of the payoff function, the strategy space and the distribution of the two types are common knowledge.

ii) In general, players only observe the actions played in the interactions in which they are involved. However, it is assumed that, if a player defects and his opponent cooperates, a proportion q of the players becomes informed about the defection. One could think, for example, of a player telling his friends about the bad behaviour of his last opponent. If both players defect, they do not inform their friends about the interaction. The intuitive reason is that they do not want to appear in a bad light, as they have defected themselves as well.

iii) The players are assumed to have imperfect recall. The players who learn in a given period about the defection of another player will remember the fact that this player has defected in the past, without however remembering in which period(s) it happened[11]. Also, players do not remember any other aspects of past interactions.

G.6 - Solution concept: The equilibrium concept used is the "Perfect Bayesian Equilibrium".

3.2 The equilibria of the stage game

First, I derive results that are valid for the stage game in any period, then I focus on the reputation cost of defection, which is related to the "shadow of the future". Thus, for the moment we can think of the game as a one-shot game. It is analysed under what conditions players will enter into contact with their match, and whether they choose cooperation or defection.

When players decide to enter the game, they choose defection ($F_i = 1$) rather than cooperation if $\rho > \rho^* = \frac{c}{S\theta}$[12]. This cut-off level ρ^* is crucial for the equilibrium of the game.

Further, players only decide to enter the game if the expected payoff V_i (given the optimal levels of F_i and F_j chosen thereafter) is greater than their outside option of staying out, which equals 0.

If both types have high fighting abilities, i.e. $\alpha > \rho^*$ and $\beta > \rho^*$, both will fight if they choose to enter. Equation (2) displays the condition under which players of a given type choose to enter. They enter if the expected payoff of entering is greater than or equal to zero[13].

[11] This assumption eases the exposition by assuring that only stationary strategies will be played in equilibrium.

[12] This follows from the first derivative of V_i with respect to F_i in equation (1).

[13] As tie-breaking rule, it is assumed that in case of indifference, the players enter the interaction.

$$(\frac{1}{2} + \theta(\rho - \widetilde{p}\alpha - (1-\widetilde{p})\beta))S - c - g \geq 0, \; where \; \rho \in \{\alpha, \beta\} \tag{2}$$

where $\widetilde{p} = \left(\frac{t_S p}{t_S p + t_W(1-p)}\right)$, p=proportion of the population being "strong" types, t_S=proportion of "strong" types entering, t_W=proportion of "weak" types entering.

The parameter $\widetilde{p}$ refers to the proportion of the entering players that are of a "strong" type. According to the values of the different parameters there are three possible outcomes: both types stay out, "strong" types enter and defect and "weak" types stay out or both types choose to enter and defect. For some ranges of values multiple equilibria arise. As the focus of the present contribution is the reputation effect of group cleavages, which is only relevant to the case of $\alpha > \rho^* > \beta$ treated further below, we will not go into more detail for the case of $\alpha > \rho^*$, $\beta > \rho^*$.

If both types have not very effective fighting technologies, $\alpha < \rho^*$, $\beta < \rho^*$, they will both choose full economic cooperation, where $F_i = 0$. For both players cooperating, $F_i = F_j = 0$, the payoff of entering the game is always positive. Thus, for ineffective fighting technologies in equilibrium all players choose the actions (enter, cooperate) in all periods.

The case which is most interesting and relevant to our research question is when $\alpha > \rho^* > \beta$. From now on we will focus on this case. For $\alpha > \rho^* > \beta$, "strong" types would in a one-shot game, if they enter, always choose defection ($F_i = 1$), and "weak" types would, if they do not stay out, always choose to cooperate ($F_i = 0$). As before, different cases can be distinguished according to the decision of the two types to enter or stay out. "Strong" types enter the interaction if condition (3) holds:

$$(\frac{1}{2} + \theta(\alpha - \widetilde{p}\alpha))S - c - \widetilde{p}g \geq 0 \tag{3}$$

Please note that, for assuring correct and consistent beliefs, "strong" types must have the beliefs of all "strong" types entering if condition (3) holds. Thus, this implies that t_S=1 and condition (3) becomes: $(\frac{1}{2} + \theta(\alpha - \widetilde{p}'\alpha))S - c - \widetilde{p}'g \geq 0$, where $\widetilde{p}' = \left(\frac{p}{p + t_W(1-p)}\right)$.

"Weak" types enter the interaction if condition (4) holds:

$$(\frac{1}{2} - \theta\widetilde{p}\alpha)S - \widetilde{p}g \geq 0 \tag{4}$$

Given that $\alpha > \rho^* = \frac{c}{S\theta}$, "strong" types have always greater incentives to enter the interaction than "weak" types. Put differently, equation (3) always holds if equation (4) holds.

To make the analysis interesting, we can assume that condition (3) always holds, for all levels of $\widetilde{p}$. This is assured by assumption G.7 below.

G.7 - Condition assuring that strong types always enter: It is assumed that $\frac{S}{2} - c - g \geq 0$.

It follows that "strong" types always enter, and "weak" types only enter the interaction with a given opponent if the probability $\widetilde{p}$ of the opponent being "strong" is smaller than some threshold level $\widetilde{p}^*$. Formally, condition (4) can be rewritten as:

$$\widetilde{p} \leq \widetilde{p}^* \equiv \frac{S}{2(S\theta\alpha + g)} \tag{5}$$

Again, for making the analysis interesting, we assume that the proportion p of "strong" types is relatively small and that condition (5) holds if all "weak" players enter the game ($t_W = 1$). It is also assumed that $0 < \widetilde{p}^* < 1$. Accordingly, if condition (5) holds and if a "weak" type matches with some trader of whom she has not had any *a priori* information, she will enter the interaction and then choose to trade ($F_i = 0$). Please note that there is also another equilibrium where all "weak" players stay out, even though entering would be profitable if all "weak" types were to enter, and where accordingly $\widetilde{p} = 1$. For the rest of the analysis we will focus on the most interesting case where the conditions (3) and (5) hold, and where without *a priori* information "weak" players enter the game.

As shown in the next subsection, if a player has learnt that her present opponent has defected in the past, she can deduce (using Bayesian updating) that her opponent is with probability $\widetilde{p}$=1 a "strong" type and that therefore condition (5) does not hold. Thus, she will not enter the interaction.

3.3 The reputation cost of defection

So far, the stage game has been analysed. Now, inter-temporal considerations are included. As players both discount future benefits and take into account the possibility of dying in future periods, we multiply future benefits with multiples of the parameter $\delta = \delta_0 h$, where δ_0=discount factor ($0<\delta_0<1$) and h=probability of being still alive in a given period ($0<h<1$).

First of all, we have to state the infinite periods equivalent of equations (3) to (5), in order to know the conditions under which "strong" and "weak" types enter the game if they have not received any information about the past behaviour of the opponent (in the case of receiving information about a past defection their belief structure is different, as we will see at a later stage). Taking into account the probabilities of opponents being of a "strong" type, and of them defecting[14], conditional on having received no information, the equations (4) and (5) become, respectively, (4') and (5'). Assumption G.7, i.e. $\frac{S}{2} - c - g \geq 0$, still assures that defecting "strong" types always enter. Equation (4') now refers to the entering condition for a cooperating "strong" type or a "weak" type. It translates in equation (5'), as before.

$$(\frac{1}{2} - \theta\widehat{p}\widehat{z}\alpha)S - \widehat{p}\widehat{z}g \geq 0 \tag{4'}$$

[14] As shown in the proof of proposition 1, there exists no equilibrium where "weak" types defect.

$$\widehat{p} \leq \widehat{p}^* \equiv \frac{S}{2(S\theta\alpha + g)\widehat{z}} \tag{5'}$$

where $\widehat{p}$=probability of the opponent being a "strong" type conditional on receiving no information about past defections, $\widehat{z}$=probability of the opponent defecting conditional on having not being informed about past defections.

We will focus on the case where the condition (5') holds and where accordingly all "weak" players enter the game if they have no (negative) *a priori* information about their match. As before, we assume that $0 < \widehat{p}^* < 1$. The optimal strategy of the "weak" types, and the optimal strategy of "strong" types who get informed about the opponent's past defection are treated in proposition 1. Below, we will derive the optimal strategy of "strong types" when they receive no information about the past behaviour of their opponent.

There is a reputation cost for "strong" types choosing defection in the first period, as informed "weak" players will not enter in economic interaction with them. If this reputation cost is big enough, "strong" players will in the first period choose cooperation ($F_i = 0$) rather than defection in order to avoid the reputation cost.

The condition under which "strong" players choose cooperation rather than defection in the first period can be obtained by comparing the expected values of cooperation and defection. Usually such a problem would be very complex as one would have to consider an infinite number of strategies. Fortunately, the structure of the reputation cost of defection implies that "strong" types have only two potential strategies which are a best-reply for some parameter values: First, defection in the first period and always thereafter. Second, cooperation in all periods. In what follows it is shown that these two strategies are Perfect Bayesian Equilibria for some parameter values. The preliminary results needed are derived in the lemmas 1 and 2.

As outlined earlier in the assumptions G.1 and G.5, at the beginning of each period a proportion of players die (some of which are informed about past defections), and then further people become informed about the defection in the past period. People who are informed, stay informed until their death. Allowing for some probability of "forgetting" would not affect our results.

To show that "always defect" can be an equilibrium strategy for "strong" types for some parameter values, we have to show that "strong" types who find it in their interest to defect in some period (if the reputation cost of defection is not big enough) will only continue to choose defection if the reputation cost of doing so is non-increasing, which is the case in our framework. The intuition of the proof is as follows:

A player will only defect in a given period if the initial gain of defection is greater than the loss due to the additional number of players informed about the defection. At the beginning of the first period in which defection is chosen by a given player, nobody is informed about a past defection, as there was no past defection. As always, a part q of the "weak" non-informed players get

informed[15]. After a second defection a proportion q of the non-informed "weak" players would become informed. As there are now less non-informed players (as some of the informed players of the previous period survive), the additional reputation cost of defection in this second defection period would be smaller and so forth. Thus, once the player has defected, the reputation costs of future defections is ever decreasing.

Also, until the first defection the incentive structure of a given player is stationary. Thus, if he finds it in his interest to defect in a given period τ, he would already have incentives to start defecting at any period $t < \tau$. It follows that he will start defection in the first period.

The reasoning above is summarised in lemma 1.

Lemma 1 *A player who ever chooses to defect will start to do so in the first period, and will stick to defection in all future periods.*

Proof. Please refer to Appendix A. ■

The reasoning for the case of players choosing in all periods to cooperate, treated in lemma 2, is similar to the reasoning applied for lemma 1. As long as a player never chooses defection, his incentive structure is the same for all periods and, if he has no incentives to choose defection in a given period τ, he will not find it in his interest to do so in any period $t \neq \tau$.

Lemma 2 *A player who finds it in her interest to cooperate in a given period, will also choose cooperation in all previous and future periods.*

Proof. Please refer to Appendix A. ■

Following the results of lemmas 1 and 2, and assuming that the equation (4') holds, we can derive for "strong" types the conditions under which the equilibria "always cooperate" or "always defect" are selected when receiving a signal "N" (no information about the opponent's past behaviour is revealed).

From our assumption G.7, i.e. $\frac{S}{2} - c - g > 0$, follows that "strong" players will choose (enter, defection) whenever they observe a signal "I" (information that the opponent has defected in the past), as defecting on another defector will not result in a reputation cost (cf. assumption G.5).

Equation (6) represents the inter-temporal expected value for a given player i, who is of a "strong" type, to choose defection in the first period and always thereafter. This expected value computation takes into account that in equilibrium opponents who have defected in the past after observing "N" will defect again, and that "weak" players who get informed in the future about player i's defection will stay out, while informed "strong" opponents will defect.

[15] Also the same proportion q of the already informed players get informed about the defection in that particular period, but this has no impact as they were already previously informed.

$$\widehat{q}\left[\frac{S}{2}-c-g\right]$$
$$+(1-\widehat{q})\left\{\begin{array}{c}(\frac{1}{2}+\theta\alpha(1-\widehat{z}\widehat{p}))S-c-\widehat{z}\widehat{p}g\\ +\left(\frac{\delta}{1-\delta}-\widetilde{q}\right)\widehat{p}\left[(\frac{1}{2}+\theta\alpha(1-\widehat{z}))S-c-\widehat{z}g\right]\\ +\widetilde{q}\widehat{p}\left[\frac{S}{2}-c-g\right]+\left(\frac{\delta}{1-\delta}-\widetilde{q}\right)(1-\widehat{p})\left[(\frac{1}{2}+\theta\alpha)S-c\right]\end{array}\right\} \quad (6)$$

where $\widehat{q}$=expected probability of receiving a signal "I" (information that the present opponent has defected in the past), $\widehat{z}$=expected proportion of potential "strong" type opponents who defect conditional on a signal of "N" (no information about the opponent's past behaviour is revealed), $\widehat{p}$=expected proportion of opponents being of a "strong" type conditional on having received a signal "N", $\widetilde{q}$=present value of the proportion of players who are informed in the different future periods about player i having defected.

Equation (7) reports the expected value for "strong" types of choosing cooperation in all periods when receiving a signal "N", and choosing defection when receiving a signal "I".

$$\widehat{q}\left[\frac{S}{2}-c-g\right]+(1-\widehat{q})\left\{\begin{array}{c}(\frac{1}{2}-\theta\widehat{z}\widehat{p}\alpha)S-\widehat{z}\widehat{p}g\\ +\frac{\delta}{1-\delta}\widehat{p}\left[(\frac{1}{2}-\theta\widehat{z}\alpha)S-\widehat{z}g\right]+\frac{\delta}{1-\delta}(1-\widehat{p})\left[\frac{S}{2}\right]\end{array}\right\} \quad (7)$$

The expected value of cooperation is greater if the proportion of players who are informed about the previous periods' defection(s) is big enough. "Strong" types will choose cooperation rather than defection if condition (8) holds[16][17].

$$\widetilde{q}\geq\widetilde{q}^{*}=\frac{\frac{1}{1-\delta}(\theta\alpha S-c)}{\widehat{p}\left[(\theta\alpha S+g)(1-\widehat{z})\right]+(1-\widehat{p})\left[(\frac{1}{2}+\theta\alpha)S-c\right]} \quad (8)$$

Please note that the variable $\widetilde{q}$ is strictly increasing in q (this can be seen from the equations used in the proof of lemma 1). This permits us to focus in the following analysis on q (the cut-off level of q corresponding to $\widetilde{q}^{*}$ can be denoted as q*).

The equilibrium of the game is summarised in proposition 1. The beliefs about the opponent are denoted by μ, where μ=expected probability that the opponent is a "strong" type.

Proposition 1 *The following set of strategies and beliefs constitutes a Perfect Bayesian Equilibrium for the usual assumptions G.1 to G.7, if equation (4') holds and if* $\alpha>\rho^{*}>\beta$*:*

[16] The behaviour of the other "strong" types (i.e. $\widehat{z}$) is linked to whether condition (8) holds or not. For example, for condition (8) holding, it must be that $\widehat{z}=0$. For some ranges of parameter values there are multiple equilibria.

[17] For simplicity and without loss of generality, it is assumed that players choose cooperation when they are indifferent.

The first case is if equation (8) does not hold, i.e. if $\widetilde{q}$ is small (low reputation cost of defection). "Strong" types always choose (enter, defect; $\mu = 1$) for a signal "I" and (enter, defect; $\mu = \widehat{p}$) for a signal "N". "Weak" types play (out; $\mu = 1$) if they observe a signal "I", and play (enter, cooperate; $\mu = \widehat{p}$) if they observe "N".

The second case is if equation (8) holds, i.e. if $\widetilde{q}$ is big (high reputation cost of defection). "Strong" types play (enter, defect; $\mu = 1$) for a signal "I" and (enter, cooperate; $\mu = \widehat{p}$) for a signal "N", "weak" types play (out; $\mu = 1$) after observing "I" and (enter, cooperate; $\mu = \widehat{p}$) after "N".

This is the unique equilibrium for the "weak types" entering the game.

Proof. Please refer to Appendix A. ■

There are also two equilibria where the "weak" types do not enter the game. They are treated in the proof of proposition 1. For the analysis of group conflict in the next sections, only the case referred to in proposition 1 is relevant.

In the next section it will be assessed how group cleavages affect the (stage game) "reputation cost" of defection, q, and in this way influence the scope of cooperation and defection.

4 Introducing Group Cleavages in the Model

So far, the variable q has been regarded as exogenous. At present, q will become endogenous to the model and it will be discussed how the group structure affects the level of q[18].

In a homogenous society with only one group, the probability of the next match of a player being informed about his past defection corresponds simply to the number of uninformed players ("friends") who become informed by each player who has been betrayed in the previous period, divided by the total number of players in the population. Thus, $q = k$, where k=part of uninformed players who become informed about the defection.

Introducing group cleavages in the model leads to additional assumptions and features of the game. These are listed below.

G.8 - Two groups: Initially, we assume that the population is composed of two groups, which for example differ in ethnic characteristics (in section 5 the model will be extended to n-groups). The first group amounts to a share w of the whole population ($0 \leq w \leq 1$). Accordingly, the part v=(1-w) of the population belongs to the second group.

[18] The equilibrium strategies derived in the previous section remain valid for this extended version of the model. The strategies for meeting players from the own group and for meeting players from other groups remain stationary. However, non-informed strong players can for example choose to always cooperate with players of their own group and always defect on players of the other groups. For some settings it can occur that after several periods of such a strategy players will find it in their interest to also defect on players of their own group. This, however, does not affect the qualitative predictions of the model.

G.9 - Part time d: Players spend a certain part d (whereas 0<d<1) of their total time endowment (which is normalized to 1) for within-group activities. For expositional ease, it is assumed in the main text that this part d is fixed and does not depend on the relative group sizes. The results are all robust for a more general framework, with the part of time spent on intra-group interactions depending on the relative group sizes. This is derived in Appendix B.

The model presented allows for different levels of d for different groups, and this general case will also be emphasised for the analysis of intra-group conflict. For inter-group interaction, however, we will assume without loss of generality that all groups have the same level of d, which eases the exposition.

The part d could for example be interpreted as the time spent on tribal gatherings or religious ceremonies and on other intra-group interaction. Similarly, the fraction of time (1-d) is spent with people from the other group. Typically, people from both groups are assumed to spend more than the proportional share of their time on intra-group activities. This can be expressed as $d_i > w$, $d_j > v = (1 - w)$ for the analysis of intra-group interaction, and $d > w$, $d > v = (1 - w)$ for inter-group interaction.

G.10 - Matching: For some values of d and w, not all players will find a trade partner. In this case they are assumed to be outside the game for one period and to get some compensation (for example through an insurance scheme), leaving their expected inter-temporal payoff unchanged.

4.1 The Likelihood of Intra-Group and Inter-Group Tension

The probability of the next period's match being informed about the present defection in the case of interaction with a member of the same group i is given by q_S below (computation in Appendix A)[19].

$$q_S = k \left[\frac{d_i^2}{w} + \frac{(1 - d_i)^2}{(1 - w)} \right] \tag{9}$$

where, k=part of uninformed players who become informed about the defection ("friends"), w=relative size of the own group i relative to the whole population ($0 \leq w \leq 1$), d_i=the part of the time a given player spends with people from her own group i ($0 \leq d_i \leq 1$).

If a player defects on an opponent from another group, the probability q_D of the next period's match being informed becomes as displayed below (computation in Appendix A). As mentioned earlier, we assume for convenience that $d_i = d_j \equiv d$.[20]

[19] The expression is analogous for the other group j.

[20] Please note that the level of inter-group defection is the same for both groups, independently of the parameter values. Knowing that w+v=1, we can easily see from equation (10) that $q_D^i = q_D^j$ always holds.

$$q_D = k\left[\frac{d(1-d)}{w(1-w)}\right] \tag{10}$$

Group cleavages affect at the same time the unconditional probability of meeting people and the conditional probability of their being informed. The interaction of changes in these two values leads to non-linear effects of introducing group cleavages on the likelihood of the next match being informed. In some cases, group division can lead to increased social tensions due to a lower reputation cost of defection. In other cases, the likelihood of social tensions can be reduced. We shall now compare the scope for social tensions within groups, between groups and in homogenous societies.

Proposition 2 *The likelihood of intra-group tension initiated by a member of a group i in heterogenous societies is lower than the likelihood of inter-group tension for the usual assumption that the time spent for intra-group interaction is greater than group i's proportional share in the population ($d > w$), resp. ($d > 1 - w$).*

The likelihood of intra-group tension initiated by a member of a group i in heterogenous societies is always lower than the likelihood of tension in a homogenous society without cleavages, given the usual assumptions.

The likelihood of inter-group tension initiated by a member of group i in heterogenous societies is higher than the likelihood of tension in a homogenous society without cleavages, given the usual assumption ($d > w$), resp. ($d > 1-w$).

Proof. Please refer to Appendix A. ■

The intuition behind these results is that a high likelihood of future interaction with informed players makes intra-group defection especially costly. Given that this likelihood is lower for inter-group interaction, the reputation cost of defection is lower for that case; it is intermediate for homogenous societies.

4.2 The Impact of Polarisation

An important issue is how increases or decreases in polarisation affect the likelihood of intra-group and inter-group tension. In the present paper I focus on the case of polarisation between two groups, which is defined in the following way.

Definition 2 *Polarisation*$\equiv 1 - |w - v|$*, where w=population share of group i, v=population share of group j.*

The more similar the shares of the two population groups, the higher is the level of polarisation in a given society. This way of introducing polarisation in our theoretical framework is consistent with the commonly used definitions and measures of polarisation (see Montalvo and Reynal-Querol, 2005).

The impact of changes in the population share of a group on its likelihood of intra-group defection and tension is given by the first derivative of q_S with respect w, as displayed in equation (11):

$$\frac{\partial q_S}{\partial w} = k\left[\frac{-d_i^2}{w^2} + \frac{(1-d_i)^2}{(1-w)^2}\right] \tag{11}$$

This expression is negative ($\frac{\partial q_S}{\partial w} < 0$), for the usual assumption $d_i > w$, i.e. when people spend more than the proportional share of their time on intra-group interaction. This implies that increases in the size of his own group w (for a given level of group integration d) lead to more defection and thus a higher likelihood of intra-group tension for group i. An increase in w would however correspond to a decrease in v (as v=1-w), lowering the likelihood of intra-group tension for the second group j.

The impact of changes in w on the likelihood of inter-group tension is given by expression (12).

$$\frac{\partial q_D}{\partial w} = k\left[\frac{d(1-d)}{(w(1-w))^2}\right](2w-1) \tag{12}$$

The expression $\frac{\partial q_D}{\partial w}$ becomes positive for w>0.5.

We shall now analyse what happens if initially polarisation is at a minimum (w=1, v=0) and increases afterwards, i.e. w decreases and v increases. The effects on intra-group defection for the more numerous group (here group i) and the less numerous group are summarised in proposition 3:

Proposition 3 *A marginal increase in polarisation (decreasing the population share of the more numerous group and increasing the share of the less numerous group for a given level of d_i) results in a lower level of intra-group tension inside the more numerous group and in a higher level of intra-group tension inside the smaller group.*

Proof. For $d_i > w$, in equation (10) we have $\frac{\partial q_S}{\partial w} < 0$. Thus, increasing (decreasing) w results in a lower (higher) q_S, and therefore a higher (lower) likelihood of defection and social tension. ■

Intuitively, for the more numerous group, a decrease in its size increases the likelihood that fellow group members will become informed (i.e. $\frac{d_i^2}{w}$ increases in equation (9)) and thus increases the reputation cost of defection. The effect goes in the opposite direction for an increase in the group size of the less numerous group.

However, in most countries that suffer from political instability and from social tensions, the constraint that is binding is the condition for inter-group tension (given that $q_D < q_S$, as derived in proposition 2). It is summarised in proposition 4 how an increase in polarisation matters with that respect.

Proposition 4 *A marginal increase in polarisation results in a lower reputation cost of defection q_D (for both groups) and accordingly in a higher level of inter-group tension.*

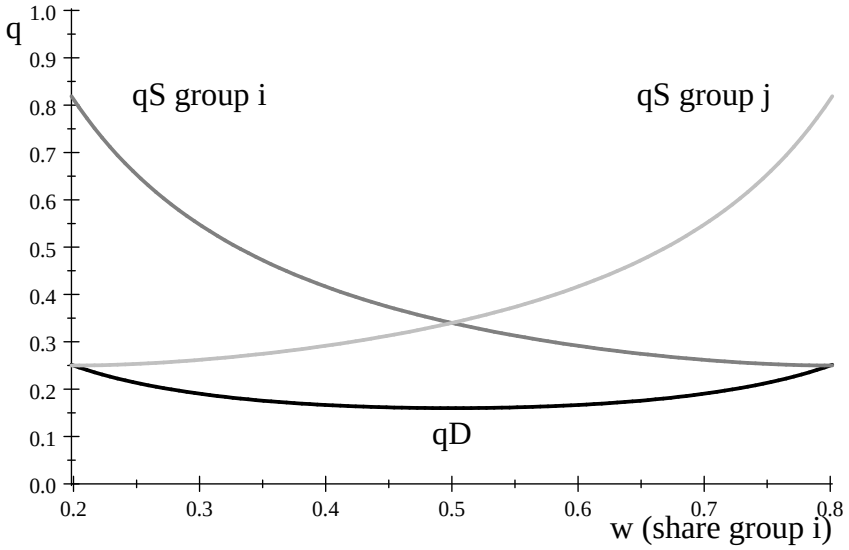

Figure 1: The impact of polarisation on qS and qD

Proof. For w>0.5 in equation (12) we have $\frac{\partial q_D}{\partial w} > 0$, and accordingly for w<0.5 we have $\frac{\partial q_D}{\partial w} < 0$. Thus, decreasing w of the more numerous group results in a lower q_D (as w>0.5 $\Rightarrow \frac{\partial q_D}{\partial w} > 0$), whereas increasing v of the smaller group results in a lower q_D as well (as v<0.5 $\Rightarrow \frac{\partial q_D}{\partial v} < 0$). ■

The intuition for this result is as follows. As shown in the Appendix A (see equation A.7), the reputation cost q_D is composed of two terms, the probability of players from the own group being informed and the probability of players from the other group being informed. If the groups become very unequal in size, one of these two terms will sharply increase, given that the relative group sizes are in the denominator of these terms. This results in a large q_D. For high levels of polarisation, both terms remain moderate in size, and q_D stays low.

The present framework provides a theoretical explanation as to why high levels of polarisation can result in social tensions between ethnic groups. If collective action is feasible, this increase in social tensions can lead to ethnic civil wars. It has been found in the empirical literature (for example, Montalvo and Reynal-Querol, 2005) that polarisation increases the risk of civil wars, but theoretical models focusing on this issue have so far been sparse.

Figure 1 plots as a numerical example[21] the levels of q_S and q_D for different levels of w. The values of q_D are lower than the values of q_S, indicating that the likelihood of inter-group tension is higher than of intra-group tension. The

[21] The following parameter values have been used: d=0.8, k=0.25.

parameter q_D takes its lowest value at w=0.5 (maximum polarisation). It follows that the more polarised a society, the greater is the likelihood of inter-group tension.

4.3 The Impact of Segregation

The concept of segregation refers to the separation and lack of interaction between different groups.[22] The extent of segregation is measured in our model by the parameter d. The following definition applies:

Definition 3 *Segregation≡d, where d=part of time spent for intra-group interaction.*

High values of d correspond to strong segregation, with only little inter-group interaction. Low values of d indicate a very integrated society with a lot of inter-group interaction.

Remember that we had defined social tension as the proportions of interactions that were conflicted, i.e. where defection occurred. This concept has so far been appropriate, as until now d was constant. For the analysis of segregation, however, d is not constant. Thus, we shall define a further concept that not only captures the likelihood of defection per interaction but also the number of interactions.

Definition 4 *Social disputes ≡ total number of defections.*

Please note that "social disputes" ≡ (total number of defections) = (social tensions) × (number of interactions). We shall first establish the impact of changes in the relevant parameter (which is d at present) on the likelihood of intra-group tension. Equation (13) displays the first derivative of q_S with respect to d_i.

$$\frac{\partial q_S}{\partial d_i} = 2k \left[\frac{d_i}{w} - \frac{(1-d_i)}{(1-w)} \right] \tag{13}$$

We have $\frac{\partial q_S}{\partial d_i} > 0$ for the usual assumption $d_i > w$. Increases in d_i result in increases in q_S, and thus lead to reduced scope for intra-group tension. However, the impact of segregation on the total number of disputes is ambiguous. Although tensions are reduced, intra-group interaction becomes more frequent. The proposition below summarises this trade-off:

Proposition 5 *More segregation (i.e., a higher d_i) increases q_S and thereby results in less intra-group tension. If the increase in q_S is substantial, it can lead to initially conflicted interactions ($q_S < q^*$) becoming peaceful ($q_S \geq q^*$), thereby reducing total intra-group disputes. In contrast, for smaller increases in q_S, segregation can result in a higher level of total social disputes by making intra-group interaction more frequent.*

[22] Whether segregation policies are politically successful and morally justifiable is controversial. The present theoretical framework provides a *positive* analysis, and does not treat the *normative* aspects of this issue.

Proof. From equation (13) follows $d_i > w \Leftrightarrow \frac{\partial q_S}{\partial d_i} > 0$. ■

This result is intuitive, as more intra-group interaction increases the monitoring of intra-group defection, reducing in this way social tensions. Now we shall analyse the effects of segregation on inter-group interactions.

$$\frac{\partial q_D}{\partial d} = k\frac{(1-2d)}{w(1-w)} \tag{14}$$

It follows from equation (14) that $d > 0.5 \Leftrightarrow \frac{\partial q_D}{\partial d} < 0$. As discussed earlier, the assumption G.9 that d>w, d>(1-w) implies that d>0,5, as $\max[w, 1-w] > 0.5$. Thus, segregation leads, as players from different groups meet less often, to a lower reputation cost of defection q_D, and to more social tension. At the same time, segregation reduces inter-group interaction, making the overall effect on inter-group disputes ambiguous.

Proposition 6 *Segregation increases inter-group tensions. The impact on inter-group disputes is ambiguous. Full segregation (d=1) eliminates inter-group disputes entirely. For intermediate levels of segregation (0<d<1), and initially conflicted inter-group interactions ($q_D < q^*$), segregation reduces the occurrence of inter-group disputes by decreasing inter-group interaction (as already $q_D < q^*$, a further decrease in q_D does not matter). For initially honest and peaceful inter-group interaction ($q_D \geq q^*$), segregation can increase the scope for inter-group disputes, if the decrease in q_D is large enough such that afterwards $q_D < q^*$ holds.*

Proof. Follows from the reasoning discussed above. ■

5 Social Tension in an n-group Framework

For analysing issues like polarisation it made sense to limit ourselves to a 2-group framework that allowed for an unequal size of the groups. However, for analysing fractionalisation, as well as for testing the robustness of previous results, it is helpful to use a n-group framework, with more than two groups, each of an equal size. Fractionalisation is defined as below:

Definition 5 *Fractionalisation*$\equiv 1 - \frac{1}{r}$*, where r=number of groups.*

The level of fractionalisation increases in the number of groups in the society.

For intra-group defection in a n-group setting, the likelihood q_S of the next period's opponent being informed is given by equation (15), which corresponds to q_S in the 2-group framework with $w = \frac{1}{r}$, where r is the number of groups.

$$q_S = k\left[\frac{{d_i}^2}{1/r} + \frac{(1-d_i)^2}{(1-1/r)}\right] \tag{15}$$

The main difference between the n-group and the 2-group framework is that in the n-group case strangers from other groups do not all belong to the *same* other group. Thus, if a player from a group i defects on an opponent of a given group j, this will result in a relatively high probability that other players of group j are informed of the defection. However, players from another "foreign" group m will be as badly informed about the defection as the players of the "home" group i. Thus, it is necessary to take into account the probability of matching people from all different groups as well as their conditional probability of being informed. This is done in Appendix A.

The likelihood of the next period's match being informed about inter-group defection is given by equation (16).

$$q_D = k\frac{(1-d)r}{(r-1)}\left[2d + \frac{r-2}{r-1}(1-d)\right] \tag{16}$$

Propositions 5 and 6, summarising the effects of segregation on social tensions and disputes, also hold in a n-group framework if the first derivative of q_S with respect to d_i is positive, and the first derivative of q_D with respect to d is negative. As q_S is the same in the n-group as in the 2-group framework (for $w = \frac{1}{r}$), the results on intra-group interaction of the 2-group setting remain valid for n-groups. As far as inter-group interaction is concerned, the first derivative of q_D with respect to d is displayed in equation (17).

$$\begin{aligned}\frac{\partial q_D}{\partial d} &= k\frac{r}{(r-1)}\left[-(2d + \frac{r-2}{r-1}(1-d)) + (1-d)(2 - \frac{r-2}{r-1})\right] \\ &= k\frac{2r}{(r-1)^2}\left[1 - dr\right] \end{aligned} \tag{17}$$

We have $\frac{\partial q_D}{\partial d} < 0 \Leftrightarrow d > \frac{1}{r} = w$. Thus, the conclusions of proposition 6 in the previous section hold as well for the n-player framework (given the usual assumption G.9).

For assessing the impact of fractionalisation on social tensions, one can focus on the derivatives $\frac{\partial q_S}{\partial r}$ and $\frac{\partial q_D}{\partial r}$. For obtaining $\frac{\partial q_S}{\partial r}$, one can simply refer to the discussion of $\frac{\partial q_S}{\partial w}$ in the previous section. As $w = \frac{1}{r}$, $\frac{\partial q_S}{\partial r}$ has just the opposite sign as $\frac{\partial q_S}{\partial w}$ before. This leads to proposition 7.

Proposition 7 *A marginal increase in fractionalisation (i.e., increasing the number of groups, r, in the population) results in a higher level of q_S, provided that $d > \frac{1}{r}$, and accordingly in a lower level of intra-group tensions.*

Proof. See proof of proposition 3. ■

Intuitively, as groups become smaller, intra-group monitoring increases. For assessing the impact of fractionalisation on inter-group tensions, one can take the derivative of q_D with respect to r, displayed below. We obtain $\frac{\partial q_D}{\partial r} < 0 \Leftrightarrow d > \frac{1}{r}$, which leads to proposition 8.

$$\begin{aligned}\frac{\partial q_D}{\partial r} &= k(1-d)\left[\left(\frac{-1}{(r-1)^2}\right)(2d+\frac{r-2}{r-1}(1-d))+\left(\frac{r}{r-1}\right)\left(\frac{1-d}{(r-1)^2}\right)\right] \\ &= k\frac{2(1-d)}{(r-1)^3}\left[1-dr\right] \qquad (18)\end{aligned}$$

Proposition 8 *Fractionalisation decreases q_D and thus leads to more inter-group tensions if $d > \frac{1}{r}$.*

Proof. Follows from the discussion above. ■

This finding is intuitive, as the reduced relative size of the opponent's group in a fractionalised society makes it less likely to meet someone of that group in the future, decreasing thereby the reputation cost of defection. Our theoretical result is consistent with recent empirical evidence on fractionalisation increasing social tensions and conflicts between ethnic groups (cf., for example, Collier, Hoeffler and Rohner, 2006).

6 Conclusion

The present contribution has examined how group cleavages matter for the emergence of social tensions. In the model players had the choice between staying out, cooperating or defecting. It has been shown how the reputation cost of future opponents being informed about defection can enforce cooperation. This reputation cost depended critically on group structure.

Increases in polarisation have been found to increase intra-group tensions in the less numerous group and decrease tensions inside the more numerous group. It also leads to more social tensions between groups. The impact of increased segregation are less clear-cut, as segregation not only affects the level of social tensions, but also the frequency of interaction. If the effect of segregation on q_S is large, it can reduce intra-group disputes, whereas otherwise it can result in a rise of intra-group disputes. Inter-group disputes may be increased by segregation if initially relations are peaceful, while segregation can reduce inter-group disputes if initially interactions are conflicted. Fractionalisation has been found to decrease social tensions within groups and increase social tensions between groups.

The findings of the model can account for recent empirical evidence on polarisation, fractionalisation and ethnic conflict by Montalvo and Reynal-Querol (2005) and Collier, Hoeffler and Rohner (2006). The predictions of the theoretical framework can also be applied to the literature on social capital and merger failures.

Further theoretical and empirical research on group structure, group identity and social tensions is encouraged.

References

[1] Agrawal, Anup, Jeffrey Jaffe and Gershon Mandelker. (1992). "The Post-Merger Performance of Acquiring Firms: A Re-Examination of an Anomaly", *Journal of Finance*, 47, 1605-21.

[2] Alesina, Alberto and Eliana La Ferrara. (2000). "Participation in Heterogenous Communities", *Quarterly Journal of Economics*, 115, 847-904.

[3] Alesina, Alberto and Eliana La Ferrara. (2002). "Who trusts others?", *Journal of Public Economics*, 85, 207-34.

[4] Alesina, Alberto, Reza Baqir and William Easterly. (1999). "Public Goods and Ethnic Divisions", *Quarterly Journal of Economics*, 114, 1243-84.

[5] Banal-Estanol, Albert, Ines Macho-Stadler and Jo Seldeslachts. (2006). "Endogenous Mergers and Endogenous Efficiency Gains: The Efficiency Defence Revisited", forthcoming in the *International Journal of Industrial Organization*.

[6] Basu, Kaushik. (2005). "Racial conflict and the malignancy of identity", *Journal of Economic Inequality*, 3, 221-41.

[7] Bates, Robert. (1999). "Ethnicity, Capital Formation, and Conflict", CID Working Paper no. 27, Harvard University.

[8] Caselli, Francesco, and Wilbur John Coleman II. (2006). "On the Theory of Ethnic Conflict", mimeo, London School of Economics and Duke University.

[9] Cederman, Lars Erik, and Luc Girardin. (2007). "Beyond Fractionalization: Mapping Ethnicity onto Nationalist Insurgencies", *American Political Science Review*, 101, 173-85.

[10] Collier, Paul and Anke Hoeffler. (1998). "On Economic Causes of Civil Wars", *Oxford Economic Papers*, 50, 563-73.

[11] Collier, Paul and Anke Hoeffler. (2004). "Greed and grievance in civil war", *Oxford Economic Papers*, 56, 563-95.

[12] Collier, Paul and Anke Hoeffler. (2007). "Civil War", in Todd Sandler and Keith Hartley, *Handbook of Defense Economics* (Vol. 2), Amsterdam, Elsevier. Chapter 23.

[13] Collier, Paul, Anke Hoeffler, and Dominic Rohner. (2006). "Beyond Greed and Grievance: Feasibility and Civil War", mimeo, University of Oxford and University of Cambridge.

[14] Dasgupta, Partha. (2005). "Economics of Social Capital", *Economic Record*, 81, S2-S21.

[15] Diez Medrano, Juan. (1994). "The Effects of Ethnic Segregation and Ethnic Competition on Political Mobilization in the Basque Country, 1988", *American Sociological Review*, 59, 873-89.

[16] Dixit, Avinash. (2003). "Trade Expansion and Contract Enforcement", *Journal of Political Economy*, 111, 1293-1317.

[17] Esteban, Joan, and Debraj Ray. (1999). "Conflict and Distribution", *Journal of Economic Theory*, 87, 379-415.

[18] Esteban, Joan, and Debraj Ray. (2006a). "A Model of Ethnic Conflict", mimeo, Institut d'Anàlisi Económica and New York University.

[19] Esteban, Joan, and Debraj Ray. (2006b). "On the Salience of Ethnic Conflict", mimeo, Institut d'Anàlisi Económica and New York University.

[20] Esteban, Joan, and Debraj Ray. (2006c). "Polarization, Fractionalization and Conflict", mimeo, Institut d'Anàlisi Económica and New York University.

[21] Fearon, James, and David Laitin. (1996). "Explaining Interethnic Cooperation", *American Political Science Review*, 90, 715-35.

[22] Fearon, James, and David Laitin. (2003). "Ethnicity, Insurgency, and Civil War", *American Political Science Review*, 97, 75-90.

[23] Francois, Patrick and Jan Zabojnik. (2005). "Trust, Social Capital, and Economic Development", *Journal of the European Economic Association*, 3, 51-94.

[24] Fulghieri, Paolo and Laurie Simon Hodrick. (2006). "Synergies and Internal Agency Conflicts: The Double-Edged Sword of Mergers", *Journal of Economics and Management Strategy*, 15, 549-76.

[25] Gartzke, Erik, Quan Li, and Charles Boehmer. (2001). "Investing in the Peace: Economic Interdependence and International Conflict", *International Organization*, 55, 391-438.

[26] Glaeser, Edward. (2005). "The Political Economy of Hatred", *Quarterly Journal of Economics*, 120, 45-86.

[27] Greif, Avner. (1993). "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition", *American Economic Review*, 83, 525-48.

[28] Greif, Avner, Paul Milgrom, and Barry Weingast. (1994). "Coordination, Commitment, and Enforcement: The Case of the Merchant Guild", *Journal of Political Economy*, 102, 745-76.

[29] Hirshleifer, Jack. (1989). "Conflict and rent-seeking success functions: Ratio vs. difference models of relative success", *Public Choice*, 63, 101-12.

[30] Horowitz, Donald. (1985). *Ethnic Groups in Conflict*, Berkeley, University of California Press.

[31] Knack, Stephen, and Philip Keefer. (1997). "Does Social Capital Have An Economic Payoff? A Cross-Country Investigation", *Quarterly Journal of Economics*, 112, 1251-88.

[32] Lind, Jo Thori. (2007). "Fractionalization and the size of government", *Journal of Public Economics*, 91, 51-76.

[33] Luttmer, Erzo. (2001). "Group Loyalty and the Taste for Redistribution", *Journal of Political Economy*, 109, 500-28.

[34] Martin, Philippe, Thierry Mayer, and Mathias Thoenig. (2006). "Make Trade not War?", mimeo, University of Paris 1 and University of Geneva.

[35] Miguel, Edward and Mary Kay Gugerty. (2005). "Ethnic diversity, social sanctions, and public goods in Kenya", *Journal of Public Economics*, 89, 2325-68.

[36] Montalvo, José, and Marta Reynal-Querol. (2005). "Ethnic Polarization, Potential Conflict, and Civil Wars", *American Economic Review*, 95, 796-815.

[37] Nowak, Martin, and Karl Sigmund. (1998). "Evolution of indirect reciprocity by image scoring", *Nature*, 393, 573-7.

[38] Olzak, Susan, Suzanne Shanahan and Elizabeth McEneaney. (1996). "Poverty, Segregation, and Race Riots: 1960 to 1993", *American Sociological Review*, 61, 590-613.

[39] Oneal, John, and Bruce Russett. (1999). "Assessing the Liberal Peace with Alternative Specifications: Trade Still Reduces Conflict", *Journal of Peace Research*, 36, 423-42.

[40] Polachek, Solomon. (1980). "Conflict and Trade", *Journal of Conflict Resolution*, 24, 55-78.

[41] Putnam, Robert. (1995). "Bowling Alone: America's Declining Social Capital", *Journal of Democracy*, 6, 65-78.

[42] Reynal-Querol, Marta. (2002). "Ethnicity, Political Systems, and Civil Wars", *Journal of Conflict Resolution*, 46, 29-54.

[43] Rob, Rafael and Peter Zemsky. (2002). "Social capital, corporate culture, and incentive intensity", *RAND Journal of Economics*, 33, 243-57.

[44] Robinson, James. (2001). "Social identity, inequality and conflict", *Economics of Governance*, 2, 85-99.

[45] Rohner, Dominic. (2006). "Beach holiday in Bali or East Timor? Why conflict can lead to under- and overexploitation of natural resources", *Economics Letters*, 92, 113-7.

[46] Sambanis, Nicholas. (2000). "Partition as a Solution to Ethnic War: An Empirical Critique of the Theoretical Literature", *World Politics*, 52, 437-83.

[47] Sambanis, Nicholas. (2001). "Do Ethnic and Nonethnic Civil Wars Have the Same Causes?", *Journal of Conflict Resolution*, 45, 259-82.

[48] Sen, Amartya. (2006). *Identity and Violence: The Illusion of Destiny*, New York, Norton.

[49] Skaperdas, Stergios. (1996). "Contest success functions", *Economic Theory*, 7, 283-90.

[50] Skaperdas, Stergios and Constantinos Syropoulos. (2001). "Guns, Butter, and Openness: On the Relationship between Security and Trade", *American Economic Review*, 91, 353-7.

[51] Temple, Jonathan and Paul Johnson. (1998). "Social Capability and Economic Growth", *Quarterly Journal of Economics*, 113, 965-90.

[52] Tirole, Jean. (1996). "A Theory of Collective Reputations (with Applications to the Persistence of Corruption and to Firm Quality)", *Review of Economic Studies*, 63, 1-22.

[53] Vanhanen, Tatu. (1999). "Domestic Ethnic Conflict and Ethnic Nepotism: A Comparative Analysis", *Journal of Peace Research*, 36, 55-73.

[54] Vigdor, Jacob. (2004). "Community composition and collective action: Analyzing initial mail response to the 2000 census", *Review of Economics and Statistics*, 86, 303-12.

[55] Weber, Roberto and Colin Camerer. (2003). "Cultural Conflict and Merger Failure: An Experimental Approach", *Management Science*, 49, 400-15.

Appendix A - Derivations of the mathematical results of the main text

Proof of Lemma 1:

In each period first a proportion (1-h) of all players die and are replaced by newly born players, then a proportion q of players become informed if there was a defection in the previous period. Players remain informed until they die.

As for "strong" players, the condition $\alpha > \rho^* = \frac{c}{S\theta}$ holds, the initial gain of defection (labelled G_τ) in a given period τ equals $G_\tau = \theta\alpha S - c > 0$, which is the same in all periods.

The present value of the reputation cost L_τ of defecting for the first time in a given period τ equals the sum of all foregone gains in the future due to previously non-informed players getting informed about this particular defection. The exact cost of defection for a player i in a given period t depends on the actions chosen in the future. However, for any given strategy the cost of defection decreases in the number of past defections. Without loss of generality we can focus on the comparison between the reputation cost of defecting in period τ and defecting in period $\tau+1$ (the reasoning is the same for defecting in a period $t \geq \tau + 2$). It is found that $L_{\tau+1} = L_\tau(1 - qh) < L_\tau$, where q>0, h>0.

Below, this finding is illustrated by the derivation of the results for the two stationary cases of "always cooperate" and "always defect".

For the case of always playing (enter, defect) in all future periods $t \geq \tau + 1$, the loss of defecting in period τ (the first period of defection) equals $L_\tau = (\widetilde{q}^F_\tau - \widetilde{q}^T_\tau)y$, where $y = \widehat{p}\,[(\theta\alpha S + g)(1 - \widehat{z})] + (1 - \widehat{p})\left[(\frac{1}{2} + \theta\alpha)S - c\right]$ and where $\widetilde{q}^F_\tau = \left[q_\tau + \delta q_{\tau+1} + \delta^2 q_{\tau+2} + ...\right]$, with $q_\tau = 0$, $q_{\tau+1} = q$, $q_{\tau+2} = q_{\tau+1}h + (1 - q_{\tau+1}h)q = qh + (1-qh)q$ and so forth, and where $\widetilde{q}^T_\tau = \left[q_\tau + \delta q_{\tau+1} + \delta^2 q_{\tau+2} + ...\right]$, with $q_\tau = 0$, $q_{\tau+1} = 0$, $q_{\tau+2} = q$ and so forth. $\widetilde{q}^F_\tau$ ($\widetilde{q}^T_\tau$) refers to the present value of the proportion of players being informed in the future if defection (cooperation) is chosen in period τ. The term capturing the reputation loss, $(\widetilde{q}^F_1 - \widetilde{q}^T_1)$ of defecting for the first time in period 1, becomes $(\widetilde{q}^F_\tau - \widetilde{q}^T_\tau) = \left[\delta q + \delta^2(1 - q)qh + ...\right]$.

If a player continues to defect in period $\tau + 1$ after having defected for the first time in period τ, the immediate gains G of defection remain the same, but the reputation cost is different. Without loss of generality we can treat the case of a defection in period $\tau+1$, after having already defected in period τ. The loss becomes $L_{\tau+1} = (\widetilde{q}^F_{\tau+1} - \widetilde{q}^T_{\tau+1})y$, where $\widetilde{q}^F_{\tau+1} = \left[q_{\tau+1} + \delta q_{\tau+2} + \delta^2 q_{\tau+3} + ...\right]$, with $q_{\tau+1} = q$, $q_{\tau+2} = qh + (1 - qh)q$ and $q_{\tau+3} = (qh + (1 - qh)q)h + (1 - (qh + (1 - qh)q)h)q$ etc, and $\widetilde{q}^T_{\tau+1} = \left[q_{\tau+1} + \delta q_{\tau+2} + \delta^2 q_{\tau+3} + ...\right]$, with $q_{\tau+1} = q$, $q_{\tau+2} = qh$ and $q'_{\tau+3} = qh^2 + (1 - qh^2)q$ etc. It follows that $(\widetilde{q}^F_{\tau+1} - \widetilde{q}^T_{\tau+1}) = \left[\delta q(1 - qh) + \delta^2(1 - q)qh(1 - qh) + ...\right] = (\widetilde{q}^F_\tau - \widetilde{q}^T_\tau)(1 - qh)$. Given that q>0, h>0, we know that $(\widetilde{q}^F_{\tau+1} - \widetilde{q}^T_{\tau+1})$ is smaller than $(\widetilde{q}^F_\tau - \widetilde{q}^T_\tau)$, and that defecting becomes less and less costly the more a player has defected in the past.

For the case of always playing (enter, cooperate) in all future periods $t \geq \tau + 1$, the gains of defection are as before, and the reputation cost of defecting in period τ (the first period of defection) equals $L_\tau = (\widetilde{q}^F_\tau - \widetilde{q}^T_\tau)y$, where $\widetilde{q}^F_\tau =$

$\left[q_\tau + \delta q_{\tau+1} + \delta^2 q_{\tau+2} + ...\right]$, with $q_\tau = 0$, $q_{\tau+1} = q$, $q_{\tau+2} = qh$ and so forth, and where $\widetilde{q}_\tau^T = 0$. It follows that $(\widetilde{q}_\tau^F - \widetilde{q}_\tau^T) = \widetilde{q}_\tau^F = \left[\delta q + \delta^2 qh + ...\right]$.

If a player defects in period $\tau + 1$ after having defected for the first time in period τ, and plays (enter, cooperate) in all future periods $t \geq \tau + 2$, the reputation cost of defection corresponds to $L_\tau = (\widetilde{q}_{\tau+1}^F - \widetilde{q}_{\tau+1}^T)y$, where $\widetilde{q}_{\tau+1}^F = \left[q_{\tau+1} + \delta q_{\tau+2} + \delta^2 q_{\tau+3} + ...\right]$, with $q_{\tau+1} = q$, $q_{\tau+2} = qh + (1 - qh)q$, $q_{\tau+3} = [qh + (1 - qh)q]\, h$ etc. Further, $\widetilde{q}_{\tau+1}^T = \left[q_{\tau+1} + \delta q_{\tau+2} + \delta^2 q_{\tau+3} + ...\right]$, with $q_{\tau+1} = q$, $q_{\tau+2} = qh$, $q_{\tau+3} = qh^2$ etc. Again, we obtain $(\widetilde{q}_{\tau+1}^F - \widetilde{q}_{\tau+1}^T) = \left[\delta q(1 - qh) + \delta^2 qh(1 - qh) + ...\right] = (1 - qh)(\widetilde{q}_\tau^F - \widetilde{q}_\tau^T)$.

The results obtained above for the stationary cases of always playing (enter, cooperate) or always playing (enter, defect) in the future also hold for all non-stationary cases (if they were to exist), where players cooperate in some periods and defect in others.

To summarise, once a player chooses to defect in some period τ, the reputation cost of defection in any future period $t > \tau$ will be smaller than it was in period τ. It follows that players who defect once will continue to defect in all future periods. Further, up to the period when the first defection occurs, the game is stationary and the incentives faced in each period are the same; therefore, if a player has incentives to first defect in a period τ, he would also have had incentives to defect in an earlier period $\tau' < \tau$. Thus, we have $\tau = 1$, i.e. the player defects in the first period.

Proof of Lemma 2:

A player will only choose cooperation in a given period τ if the present value of choosing "cooperate" is greater than of playing "defect". It is intuitive that if the reputation cost of defection is big enough, "strong" types would choose cooperation. Since in this case the stationary incentive structure would be the same for each period, if they are better off cooperating, it is in their interest to start with cooperation immediately in the first period.

Proof of Proposition 1:

First, we can treat all strategies where both players do not condition their actions on the signal observed about the opponent.

1) Both "weak" and "strong" types always selecting (out; $\mu \in [0, 1]$) is an equilibrium, as no player would be better off deviating.

2) "Weak" types always choosing (out; $\mu \in [0, 1]$) and "strong" types always choosing (enter, cooperate; $\mu \in [0, 1]$) is not an equilibrium, as "strong" types would deviate and play (enter, defect; $\mu \in [0, 1]$).

3) "Weak" types always choosing (out; $\mu \in [0, 1]$) and "strong" types always choosing (enter, defect; $\mu \in [0, 1]$) is an equilibrium, as nobody would deviate.

4) "Weak" types always choosing (enter, cooperate; $\mu \in [0, 1]$) and "strong" types always choosing (out; $\mu \in [0, 1]$) or (enter, cooperate; $\mu \in [0, 1]$) are not equilibria, as "strong" types would deviate and play (enter, defect; $\mu \in [0, 1]$).

5) "Weak" types always choosing (enter, cooperate; $\mu \in [0, 1]$) and "strong" types always choosing (enter, defect; $\mu \in [0, 1]$) is not an equilibrium, as "weak" types would deviate and play (out; $\mu = 1$) if they observe the signal "I".

6) Any cases where "weak" types always choose (enter, defect; $\mu \in [0,1]$) and "strong" types do not condition their actions on their signals cannot be equilibria, as "weak" types would deviate.

Now, at least one type conditions his actions on the signal observed.

7) There can be no cases where "weak" types do not condition their actions on the signal received and "strong" types do, as "strong" types would be better off not conditioning and always playing (enter, defect; $\mu \in [0,1]$).

At present, the cases of the "weak" type conditioning, and the "strong" type not conditioning are assessed.

8) Any cases where "strong" types always choose (out; $\mu \in [0,1]$) or (enter, cooperate; $\mu \in [0,1]$) and "weak" types condition their actions on their signals cannot be equilibria, as the best reply of the "weak" types would be to always play (enter, cooperate; $\mu \in [0,1]$).

9) When "strong" types always choose (enter, defect; $\mu \in [0,1]$), and when "weak" types play (out; $\mu = 1$) if they observe signal "I", and play (enter, cooperate; $\mu = \widehat{p}$) if they observe "N", it is an equilibrium for equation (8) not holding, i.e. if $\widetilde{q}$ is small. In this case nobody has incentives to deviate and the beliefs are consistent. If equation (8) holds, this would not be an equilibrium, as "strong" types would be better off playing (enter, defect; $\mu = 1$) for a signal "I" and (enter, cooperate; $\mu = \widehat{p}$) for a signal "N".

10) "Strong" types always selecting (enter, defect; $\mu \in [0,1]$) and "weak" types playing (out; $\mu \in [0,1]$) for signal "I", and (enter, defect; $\mu \in [0,1]$) for "N" is not an equilibrium, as "weak" types would deviate.

11) When "strong" types always choose (enter, defect; $\mu \in [0,1]$) and "weak" types play (enter, cooperate; $\mu \in [0,1]$) if they observe signal "I", and play (out; $\mu \in [0,1]$) if they observe "N", this is not an equilibrium, as "weak" types would deviate.

12) Consider "strong" types always choosing (enter, defect; $\mu \in [0,1]$) and "weak" types playing (enter, cooperate; $\mu \in [0,1]$) if they observe signal "I", and (enter, defect; $\mu \in [0,1]$) for signal "N". This could not be an equilibrium, as an "innocent" "weak" player i knows that his opponent will always defect, as she will receive a signal "N". The only reason for player i to defect is to be rewarded by a "weak" player choosing (enter, cooperate; $\mu \in [0,1]$) in a future interaction. However, as his current match defects with certainty, a defection of player i would not result in anyone being informed. Thus, he is better off choosing (enter, cooperate; $\mu \in [0,1]$) or (out; $\mu \in [0,1]$) according to the parameter values for observing "N".

13) Further cases are when "strong" types always choose (enter, defect; $\mu \in [0,1]$) and "weak" types play (enter, defect; $\mu \in [0,1]$) if they observe signal "I", and play (out; $\mu \in [0,1]$) or (cooperate; $\mu \in [0,1]$) after "N". A signal "I" can only come from a "strong" type who has defected on "weak" types. Thus, "weak" types would deviate, and it is not an equilibrium.

Now, both players condition their actions on their signal.

14) Consider "weak" types playing (out; $\mu = 1$) after observing "I" and (enter, cooperate; $\mu = \widehat{p}$) after "N". There is an equilibrium when "strong" types play (enter, defect; $\mu = 1$) for a signal "I" and (enter, cooperate; $\mu = \widehat{p}$)

for a signal "N" if equation (8) holds. The beliefs are consistent, and nobody has incentives to deviate. The case of equation (8) not holding is treated under 9).

15) "Weak" types playing (out; $\mu \in [0,1]$) after observing "I" and (enter, defect; $\mu \in [0,1]$) after "N" is not an equilibrium with conditioning, as the "strong" types' best reply would be to always play (enter, defect; $\mu \in [0,1]$).

16) "Weak" types playing (enter, cooperate; $\mu \in [0,1]$) after "I" and (out; $\mu \in [0,1]$) after "N" is not an equilibrium with conditioning, as the "strong" types' best reply would be to always play (enter, defect; $\mu \in [0,1]$).

17) When "weak" types play (enter, cooperate; $\mu \in [0,1]$) after observing "I" and (enter, defect; $\mu \in [0,1]$) after "N", this is not an equilibrium with conditioning, as the "strong" types' best response would be to always play (enter, defect; $\mu \in [0,1]$).

18) "Weak" types playing (enter, defect; $\mu \in [0,1]$) after "I" and (out; $\mu \in [0,1]$) after "N" is not an equilibrium with conditioning, as the "strong" types' best reply would be to always play (enter, defect; $\mu \in [0,1]$).

19) For "weak" types playing (enter, defect; $\mu \in [0,1]$) after observing "I" and (enter, cooperate; $\mu \in [0,1]$) after "N" the best reply of "strong" types would be to either always play (enter, defect; $\mu \in [0,1]$) or to play (enter, defect; $\mu \in [0,1]$) after observing "I" and (enter, cooperate; $\mu \in [0,1]$) after "N", according to the parameter values. In either case "weak" types would be better off deviating after observing "I".

Computing the probability of the next period's match being informed about the defection:

For intra-group defection, the overall probability q_S of the next match being informed is given by equation (A.1).

$$q_S = P(S)P(k \mid S) + P(D)P(k \mid D) \tag{A.1}$$

where, $P(S)$=Probability of meeting a player belonging to the same group, $P(k \mid S)$=Probability of the match being informed, conditional on being from the same group, $P(D)$=Probability of meeting a player belonging to another group, $P(k \mid D)$=Probability of the match being informed, conditional on being from another group.

By definition, the probability of a match in the next period with a player from one's own group is: $P(S) = d_i$, whereas d_i=the part of the time a given player spends with people from her own group ($0 \leq d_i \leq 1$). Accordingly, the probability of matching with someone outside the group becomes $P(D) = (1 - d_i)$.

Further, we have

$$P(k \mid S) = \frac{d_i}{w} k \tag{A.2}$$

where, k=part of uninformed players who become informed about the defection ("friends"), w=size of the own group relative to the whole population ($0 \leq w \leq 1$).

$$P(k \mid D) = \frac{(1-d_i)}{(1-w)}k \tag{A.3}$$

Introducing (A.2) and (A.3), together with $P(S) = d_i$, and $P(D) = (1-d_i)$, in (A.1), we obtain:

$$q_S = d_i \left[\frac{d_i}{w}k\right] + (1-d_i)\left[\frac{(1-d_i)}{(1-w)}k\right] = k\left[\frac{{d_i}^2}{w} + \frac{(1-d_i)^2}{(1-w)}\right] \tag{A.4}$$

For inter-group defection, the overall probability, q_D, of the next match being informed is again given by equation (A.1). As before, the likelihood of matching with a player of one's own group equals $P(S) = d$ (for convenience, it is assumed that for the case of inter-group interaction $d_i = d_j \equiv d$). The probability of matching in a given period with someone of the group of last period's betrayed opponent is $P(D) = (1-d)$. The conditional probabilities $P(k \mid S)$ and $P(k \mid D)$ become as displayed in equations (A.5) and (A.6).

$$P(k \mid S) = \frac{(1-d)}{w}k \tag{A.5}$$

$$P(k \mid D) = \frac{d}{(1-w)}k \tag{A.6}$$

Introducing $P(S)$, $P(D)$, (A.5), and (A.6) in (A.1), we obtain (A.7).

$$q_D = d\left[\frac{(1-d)}{w}k\right] + (1-d)\left[\frac{d}{(1-w)}k\right] = k\left[\frac{d(1-d)}{w(1-w)}\right] \tag{A.7}$$

Proof of Proposition 2:

Equation (8) implies that a higher probability of the next match being informed, q, reduces the incentives for defection. It follows that intra-group tensions are lower than inter-group tensions if $q_S > q_D$. We have $q_S = k\left[\frac{d^2}{w} + \frac{(1-d)^2}{(1-w)}\right]$ and $q_D = k\left[\frac{d(1-d)}{w(1-w)}\right]$. Proposition 2 is valid if condition (A.8) holds.

$$q_S > q_D \Leftrightarrow \frac{d^2}{w} + \frac{(1-d)^2}{(1-w)} > \frac{d(1-d)}{w(1-w)} \tag{A.8}$$

Condition (A.8) holds if $(2d-1)(d-w) > 0$. This is the case when the usual assumption $d > \max\{w, 1-w\}$ holds (it implies that $d > w$ and $d > 0.5$).

Intra-group tensions are lower than social tensions in homogenous societies, if $q_S > q$. We have $q = k$ and $q_S = k\left[\frac{d^2}{w} + \frac{(1-d)^2}{(1-w)}\right]$. Setting $k\left[\frac{d^2}{w} + \frac{(1-d)^2}{(1-w)}\right] > k$, we obtain after reformulation condition (A.9), which always holds.

$$q_S > q \Leftrightarrow (d-w)^2 > 0 \tag{A.9}$$

Inter-group tensions are greater than tensions in a homogenous society, if condition (A.10) holds.

$$q_D < q \Leftrightarrow \left[\frac{d(1-d)}{w(1-w)}\right] < 1 \tag{A.10}$$

This condition holds if $((d+w)-1)(d-w) > 0$, which is the case for the usual assumption $d > \max\{w, 1-w\}$.

Computing q_D for n-groups:

For inter-group defection, the overall probability, q_D, of the next match being informed is given by equation (A.11).

$$q_D = P(S)P(k \mid S) + P(C)P(k \mid C) + P(T)P(k \mid T) \tag{A.11}$$

where, $P(S)$=Probability of meeting a player belonging to the same group, $P(k \mid S)$=Probability of the match being informed, conditional on being from the same group, $P(C)$=Probability of meeting a player belonging to the group of the present opponent, $P(k \mid C)$=Probability of the match being informed, conditional on being from the group of the present opponent, $P(T)$=Probability of meeting a player belonging to some third group, $P(k \mid T)$=Probability of the match being informed, conditional on being from some third group.

As before, the likelihood of matching with a player of one's own group equals $P(S) = d$. The probability of matching in a given period with someone of the group of last period's betrayed opponent is $P(C) = (1-d)\frac{1}{r-1}$, where r=number of groups. Further, $P(T) = (1-d)\frac{r-2}{r-1}$.

The conditional probabilities are as follows:

$$P(k \mid S) = P(k \mid T) = \frac{(1-d)\frac{1}{r-1}}{\frac{1}{r}}k \tag{A.12}$$

$$P(k \mid C) = \frac{d}{\frac{1}{r}}k \tag{A.13}$$

Introducing $P(S)$, $P(C)$, $P(T)$, (A.12), and (A.13) in (A.11), we obtain (A.14).

$$\begin{aligned} q_D &= d\left[\frac{(1-d)\frac{1}{r-1}}{\frac{1}{r}}k\right] + (1-d)\frac{1}{r-1}\left[\frac{d}{\frac{1}{r}}k\right] + (1-d)\frac{r-2}{r-1}\left[\frac{(1-d)\frac{1}{r-1}}{\frac{1}{r}}k\right] \\ &= k\frac{(1-d)r}{(r-1)}\left[2d + \frac{r-2}{r-1}(1-d)\right] \end{aligned} \tag{A.14}$$

Appendix B - Derivations for the extent of intra-group interaction depending on group sizes

In the present appendix the computations of sections 4 and 5 are performed for the case where the part of time spent for intra-group interaction is not constant, but depends on the group size. The results of sections 4 and 5 are robust to this extension.

It is at present assumed that the players spend some fixed amount of time on intra-group (d_i) and inter-group (e_i) interaction, and that another part of their time is attributed to intra- or inter-group interaction depending on the relative group sizes. As before, we will first focus on the 2-group case. The new probabilities P(S) and P(D) are displayed below.

$$P(S) = d_i + (1 - d_i - e_i)w \tag{B.1}$$

$$P(D) = e_i + (1 - d_i - e_i)(1 - w) \tag{B.2}$$

where w=relative size of the player's own group relative to the whole population ($0 \leq w \leq 1$).

As previously, it holds that $P(S) = 1 - P(D)$. Since players tend to spend more time with other players belonging to their own group, it is assumed that $d_i \gg e_i$ and that $P(S) > w$, which is the case for small or intermediate values of w and for the condition $d_i \gg e_i$ being fulfilled.

Computing the probability of the next period's match being informed of the defection:

As before, *for intra-group defection*, the overall probability q_S of the next match being informed is given by equation (A.1).

Given (B.1) and (B.2), the conditional probabilities become:

$$P(k \mid S) = \frac{d_i + (1 - d_i - e_i)w}{w} k \tag{B.3}$$

where k=percentage of uninformed players who become informed about the defection ("friends").

$$P(k \mid D) = \frac{e_i + (1 - d_i - e_i)(1 - w)}{(1 - w)} k \tag{B.4}$$

Introducing (B.1), (B.2), (B.3) and (B.4) in (A.1), we obtain:

$$q_S = k \left[\frac{d_i^2}{w} + \frac{e_i^2}{1 - w} + 1 - (d_i + e_i)^2 \right] \tag{B.5}$$

For inter-group defection, the overall probability, q_D, of the next match being informed is again computed according to the same formula as in (A.1),

and the values of $P(S)$ and $P(D)$ are the same as before (see (B.1), respectively (B.2)). The new conditional probabilities are displayed in the equations (B.6) and (B.7).

$$P(k \mid S) = \frac{e_j + (1 - d_j - e_j)w}{w} k \tag{B.6}$$

$$P(k \mid D) = \frac{d_j + (1 - d_j - e_j)(1 - w)}{(1 - w)} k \tag{B.7}$$

Introducing (B.1), (B.2), (B.6) and (B.7) into (A.1), we obtain (B.8).

$$q_D = k \left[\frac{d_i e_j}{w} + \frac{d_j e_i}{1 - w} + 1 - (d_i + e_i)(d_j + e_j) \right] \tag{B.8}$$

Proposition 2:

The probability $P(S)$ can be expressed as $P(S) = d_i + (1 - d_i - e_i)w \equiv D_i$. It follows that $P(D) = 1 - D_i$, $q_S = k \left[\frac{D_i^2}{w} + \frac{(1-D_i)^2}{(1-w)} \right]$ and $q_D = k \left[\frac{D(1-D)}{w(1-w)} \right]$.[23] The formula of q_S and q_D, expressed in terms of D are exactly equivalent to the formula of (9) and (10) expressed in terms of d. Thus, the results of proposition 2 also hold for the new specification.

Polarisation:

For assessing the impact of changes in polarisation on intra-group defection, we have to take the first derivative of q_S with respect to w.

$$\frac{\partial q_S}{\partial w} = k \left[\frac{-d_i^2}{w^2} + \frac{e_i^2}{(1 - w)^2} \right] = k \left[\frac{e_i^2 w^2 - d_i^2 (1 - w)^2}{w^2 (1 - w)^2} \right] < 0 \tag{B.9}$$

This derivative is negative, as before in the main text, if e is relatively small, and w not too large. That is the case for small and intermediate values of w and if our initial assumption of people spending a more than proportional part of their time on intra-group interaction holds (i.e. $d_i \gg e_i$). Thus, as before our model predicts that more polarisation leads to more (less) intra-group tensions inside the group that sees its population share increase (decrease).

As in the main text, we set for simplicity for the analysis of inter-group interaction always $d_i = d_j \equiv d$, $e_i = e_j \equiv e$.

$$\frac{\partial q_D}{\partial w} = k \left[\frac{-de}{w^2} + \frac{de}{(1 - w)^2} \right] = k \left[\frac{de(2w - 1)}{w^2 (1 - w)^2} \right] \gtrless 0 \Leftrightarrow w \gtrless 0.5 \tag{B.10}$$

As before, the derivative is positive for w>0.5 and negative for w<0.5, indicating that increases in polarisation (making the population shares of the two groups more equal) result in more tensions.

[23] As before, we set here for expositional ease $D_i = D_j \equiv D$.

Segregation:

In the model of the main text we had $P(S) = d_i$, and more segregation simply corresponded to an increase in d_i. At present, $P(S) = d_i + (1 - d_i - e_i)w$, and again increased segregation is represented by a greater (fixed) part of time spent on intra-group interaction, d_i. The impact of segregation on intra-group tensions is displayed below in equation (B.11).

$$\frac{\partial q_S}{\partial d} = k\left[\frac{2d_i}{w} - 2(d_i + e_i)\right] = 2k\left[d_i\left(\frac{1}{w} - 1\right) - e_i\right] > 0 \tag{B.11}$$

The derivative $\frac{\partial q_S}{\partial d}$ is positive if e_i is relatively small compared to d_i, $d_i \gg e_i$, and if w is not too large. Again, this confirms the main text's previous findings. The results for inter-group tensions are as follows:

$$\frac{\partial q_D}{\partial d} = k\left[e\left(\frac{1}{w} + \frac{1}{1-w}\right) - 2(d + e)\right] < 0 \tag{B.12}$$

We obtain a negative derivative, as in the main text, if $d \gg e$.

Comparative statics of q_S for n-groups:

The analysis of segregation in a n-group framework is identical to a two-group framework, as the relevant equations are congruent for $w = \frac{1}{r}$.

As in the main text, the effect of fractionalisation on the likelihood of intra-group tension, $\frac{\partial q_S}{\partial r}$, is simply the inverse of the effect of $\frac{\partial q_S}{\partial w}$ computed in equation (B.9).

Comparative statics of q_D for n-groups:

As before, for inter-group defection, the overall probability, q_D, of the next match being informed is given by equation (A.11). The relevant expressions become (for simplicity we consider $d_i = d_j = d_k \equiv d$, $e_i = e_j = e_k \equiv e$):

$$P(S) = d + (1 - d - e)\frac{1}{r} \tag{B.13}$$

$$P(C) = \frac{e}{r-1} + (1 - d - e)\frac{1}{r} \tag{B.14}$$

$$P(T) = e\left[\frac{r-2}{r-1}\right] + (1 - d - e)\frac{r-2}{r} \tag{B.15}$$

$$P(k \mid S) = P(k \mid T) = \frac{\frac{e}{r-1} + (1 - d - e)\frac{1}{r}}{\frac{1}{r}} \tag{B.16}$$

$$P(k \mid C) = \frac{d + (1 - d - e)\frac{1}{r}}{\frac{1}{r}} \tag{B.17}$$

Introducing equations (B.13) to (B.17) in (A.11), we obtain:

$$q_D = k\left[\frac{2der}{r-1} + \frac{e^2(r-2)r}{(r-1)^2} + 1 - (d+e)^2\right] \tag{B.18}$$

The impact of changes in segregation is analysed below:

$$\frac{\partial q_D}{\partial d} = 2k\left[\frac{e}{r-1} - d\right] < 0 \tag{B.19}$$

As in the main text, we have $\frac{\partial q_D}{\partial d} < 0$ for $d > e$ and at least two groups, i.e. $r \geq 2$.

For assessing the impact of fractionalisation on inter-group tensions we take the first derivative of q_D with respect to r.

$$\frac{\partial q_D}{\partial r} = 2ek\left[\frac{e - d(r-1)}{(r-1)^3}\right] < 0 \tag{B.20}$$

The derivative $\frac{\partial q_D}{\partial r}$ becomes negative for $d > e$ and at least two groups, i.e. $r \geq 2$, which corresponds to the result in the main text.