CIREQ

Centre interuniversitaire de recherche en économie quantitative

Cahier 03-2021

On the Emergence of Consensus or Conflicting Social Norms

Juan I. Block, Rohan Dutta and David K. Levine

Cahier 03-2021

On the Emergence of Consensus or Conflicting Social Norms

Juan I. Block, Rohan Dutta and David K. Levine

CIREQ, Université de Montréal
C.P. 6128, succursale Centre-ville
Montréal (Québec) H3C 3J7
Canada
Téléphone : (514) 343-6557
Télécopieur : (514) 343-7221
cireq@umontreal.ca
http://www.cireqmontreal.com

Dépôt légal - Bibliothèque nationale du Canada, 2021, ISSN 0821-4441
Dépôt légal - Bibliothèque et Archives nationales du Québec, 2021
ISBN-13 : 978-2-89382-778-0

On the Emergence of Consensus or Conflicting Social Norms[1]

Juan I. Block[2], Rohan Dutta[3], David K. Levine[4]

Abstract

We propose a simple model to study why in societies consisting of two distinct groups with their own norms some achieve consensus while others contend with conflicting norms. In addition to the usual individual incentive compatibility assumption (as in a Nash equilibrium), each group coordinates on their "preferred" incentive compatible action profile. This delivers a unique equilibrium prediction that permits a sharp charaterization of the relationship between the emergence (and intensity) of conflict and segregation, fractionalization and other economic variables.

JEL Classification Numbers: C72, D71, D74, J15, R23, Z13

Keywords: consensus, conflict, fractionalization, collective decision-making

[1]We thank Andrea Mattozzi, and Salvatore Modica. Financial support from the EUI Research Council and the Cambridge-INET Institute is gratefully acknowledged.

[2]Faculty of Economics, University of Cambridge; email: jb2002@cam.ac.uk

[3]Department of Economics, McGill University and CIREQ; email: rohan.dutta@mcgill.ca

[4]Department of Economics and RSCAS at the EUI and WUSTL; email: david@dklevine.com

Preprint submitted to Mimeo: dklevine.com *December 29, 2020*

1. Introduction

Social tensions are a common feature of heterogeneous societies in which the different groups follow conflicting norms. At the same time, not all heteregeneous societies feature such conflict. As in numerous instances of assimilation, minority groups often adopt the majority norm and consensus emerges. Social tensions arising from conflicting norms have important economic consequences, from underprovision of public goods to ethnic conflicts, eventually leading to slower economic development (see Bazzi et al. (2019)). We propose a simple model to study why consensus emerges in some societies while others must contend with conflicting norms.

Individuals in our model are either leaders or followers and belong to one of two groups. Each group has a corresponding norm, one that is easier (less costly) to follow for the group members compared to those in the other group. The game has three stages. First, group leaders suggest social norms. Second, followers choose a norm. Third, followers engage in a round of social interaction consisting of pairwise matches. The chance of a match occurring within a group versus between groups depends on the degree of segregation. Upon being matched, each agent observes the norm chosen by their partner and imposes a fixed punishment on them if it is different from their own.

Our model of group behavior is one in which leaders are constrained by the optimal play of their followers. That is, they are free to coordinate their followers on a social norm provided it is incentive compatible for the followers. Formally, the model is one of a collusion constrained equilibrium (CCE), introduced in Dutta et al. (2018). A key feature of the model is that leaders care about the proportion of the population that adhere to their preferred social norm.

Our main result characterizes the generically unique equilibrium, which results in either a consensus or conflict. In a consensus the leaders of only one group impose their preferred norm, which the leaders of the other group, abandoning their own norm, also propose. Everyone adheres to the norm. In a conflict, group members adhere to their leaders' preferred norm and punishment occurs when members of different groups interact.

Some recent papers (see Advani and Reich (2015), Bazzi et al. (2019), Sato and Zenou (2019) and Goyal et al. (2020)) study models with similar features to ours in settings involving networks or multiple generations. They share our motivation in trying to explain why some heterogeneous societies engage in conflict while others do not. They do not, however, model the coordination process within the group. Our uniqueness result is in sharp contrast to these earlier papers and allows for a simple characterization and unambiguous comparative statics. Consensus emerges at low levels of segregation and is typically the majority group's norm for large enough majority size. Atypically, when the minority norm is preferred by all and the majority share is not too large, the minority norm is adopted in a consensus. Conflict obtains whenever the minority share exceeds a threshold, which depends on the level of segregation. Similarly, conflict arises whenever the degree of segregation exceeds a threshold, which depends on the relative group sizes and also on the net costs to a member from following the opposing groups norm instead of her own. Typically, greater fractionalization lowers this segregation threshold. The intensity of conflict decreases with greater segregation and increases with higher fractionalization, consistent with the empirical findings in Bazzi et al. (2019) , Easterly and Levine (1997) and Esteban et al. (2012).[5]

A marginal increase in segregation reduces welfare in all scenarios but one. In the exception, a society originally in conflict and with a costly enough punishment benefits mechanically from greater segregation through fewer instances of inter-group matches. We conclude by discussing how introducing simple dynamics into our model generates the phenomenon of tipping as discussed in Schelling (1971).

At the heart of this paper is the idea that groups are able to coordinate their actions through tools (peer pressure, ostracism), whose efficacy depends on the choices made by other groups. In this we follow the classic works of Olson Jr. (1965) and Ostrom (1990) and more recently Levine and Modica (2016) and Levine and Mattozzi (2020).

2. Model

2.1. Environment

Consider a society consisting of two groups $J \in \{A, B\}$. There is a continuum of individuals of unit mass, with a fraction $0 < \phi_A < 1$ who are members of group A, and with the remaining fraction $\phi_B = 1 - \phi_A$ being members of group B. In addition to group members, each group has leaders of infinitesimal mass.

There are two social norms $j \in \{a, b\}$. Social norms are group specific in that norm j corresponds to group J. For any member of group J adhering to the social norm k has an individual cost of c_{Jk}. These costs could take negative values, thereby representing benefits. We assume each member of a group likes their own social norm better than members of the other group do.

Assumption 1. *$c_{Jj} \leq c_{Kj}$ for $K \neq J$.*

Notice that members of one group may prefer the other group's social norm, that is we allow $c_{Jj} > c_{Jk}$ for $k \neq j$. We allow this because one social norm may be intrinsically more

[5] Since there are only two groups the measures of fractionalization and polarization coincide (see Esteban and Ray (2011)).

costly than the other, for example, by imposing more requirements on adherents such as working hard, mastering several languages or following particular dietary restrictions.

The leaders of each group specify simultaneously and independently the social norm that should be followed by each member of their group. Each individual takes as given the norm chosen by everyone else in society (including fellow members) and adheres to a norm to maximize expected utility. The leaders of group J prefer their own social norm j to that of the other group $k \neq j$. Leaders have self-serving interests: their objective function is the fraction of the population that adheres to their preferred social norm. In particular, leaders do not care about costs to members.

After the social norms are determined by the leaders, group members engage in a round of social interaction. Specifically, individuals are matched randomly in pairs: with probability $1 - \sigma$ the entire population is matched randomly, and with probability σ each group is matched randomly with own group members only. We refer to σ as the *degree of segregation*.

Upon being matched, each member observes whether the matched partner adhered to the same social norm or not. Social norms are assumed to rely on *peer enforcement* by which individuals must penalize deviations from accepted behavior. We assume that there is a fixed punishment $P > 0$ that is imposed by a group member on a partner who fails to comply. This may be in the form of informal social sanctions such as peer pressure and social ostracism, or other kinds of physical or material sanctions.[6]

2.2. Equilibrium

We assume leaders have a limited ability to specify the norm in the sense that group members will only adopt the social norm proposed by their leaders if it is incentive compatible. Hence, leaders may choose only such social norms. A norm is incentive compatible for a group if no group-member can better off by following a different norm while everyone else in the group follows the norm. Of course whether a norm is incentive compatible for a group depends on the actions of the other group. The notion of equilibrium that captures this idea is collusion constrained equilibrium (CCE).[7] CCE applies broadly to any non-cooperative game in which the players are partitioned into collusive groups, and is defined in an appropriately subtle way to avoid non-existence problems. In the current context, however, the set of CCE would be identical to the prediction from the following simpler equilibrium notion, which for the sake of brevity we continue to refer to as CCE.

[6] We note that the punishment of non-adherents is a characteristic of social norms: it appears, for example, even in the written constitution of a prison gang (see, e.g., Skarbek (2014)). For theoretical consideration, see Levine and Modica (2016).

[7] See Dutta et al. (2018) for a formal justification for using this solution concept when studying interaction between groups.

Definition 1. A *collusion constrained equilibrium in the social norm game (CCE)* is a choice of a social norm by the leaders of each group such that, given the choice of the leaders of the other group, it is incentive compatible for members to adhere to the norm and no other incentive compatible norm is preferred by either leader.

If in equilibrium leaders of both groups choose the same social norm we refer to *consensus*, and if in equilibrium leaders of each group choose their preferred social norm we refer to *conflict*.

We require that the punishments be large enough to induce compliance with social norms. To this end, we assume that the cost of being punished is greater than the cost of switching social norms.

Assumption 2. *For any group J and $k \neq j$, $P > |c_{Jk} - c_{Jj}|$.*

To avoid special cases due to group members being ex ante indifferent, we make the following genericity assumption:

Assumption 3. *For each group J and $k \neq j$, $\phi_J \neq \dfrac{1-2\sigma}{2(1-\sigma)} - \dfrac{(c_{Jk} - c_{Jj})/P}{2(1-\sigma)}$.*

3. Consensus and Conflict

In this section, we characterize the conditions that lead to a consensus or conflict, and which social norm is adopted when there is consensus.

To describe our results in a parsimonious manner, define $d_J \equiv (c_{Jk} - c_{Jj})/P$ as the net cost relative to being punished for group J members who adhere to the opposing social norm $k \neq j$. Our assumptions about c_{Jk} are reflected in the following key properties of d_J:

Lemma 1. *$d_A + d_B \geq 0$ and $-1 < d_A, d_B < 1$.*

Our main result shows that generically there is a unique collusion constrained equilibrium.

Proposition 1. *If $\phi_J > \dfrac{1 + d_{-J}}{2(1-\sigma)}$, there is a unique collusion constrained equilibrium with consensus j. Otherwise, there is a unique collusion constrained equilibrium with conflict.*

Our formal proof is in the AppendixA; here we discuss the idea. Note that if both norms are incentive compatible for a group, the group leader would strictly prefer to propose her preferred social norm. Then we need to characterize the conditions under which it is incentive compatible for the group members to adhere to their leader's preferred norm given the other group's behavior. To this end, it is optimal for group J members to adhere to (their own) social norm j while the other group members follow social norm $k \neq j$ if the population share ϕ_J is above the following threshold

$$\underline{\phi}_J(\sigma, d_J) \equiv \frac{1 - 2\sigma - d_J}{2(1-\sigma)}.$$

It is incentive compatible for members in group K to adhere to the norm j when the population share ϕ_J is above the threshold

$$\overline{\phi}_J(\sigma, d_{-J}) \equiv \frac{1 + d_{-J}}{2(1-\sigma)}.$$

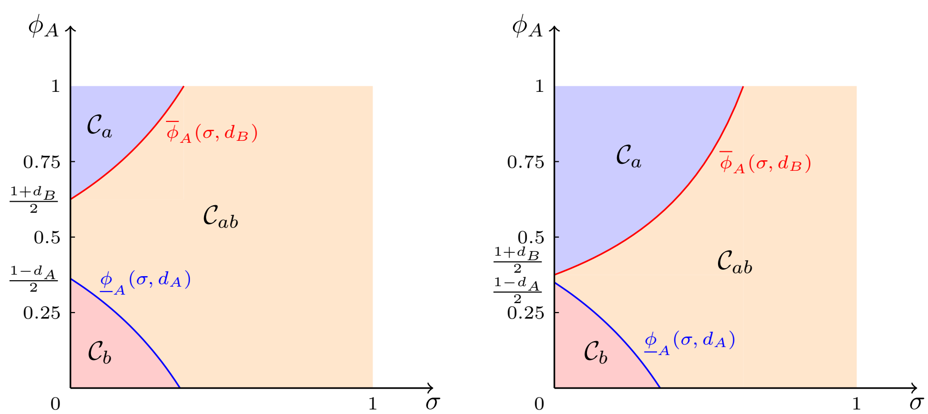
Observe that $\overline{\phi}_J \geq \underline{\phi}_J$ by Lemma 1 and that $\overline{\phi}_J = 1 - \underline{\phi}_{-J}$. These thresholds are sufficient to describe the collusion constrained equilibrium as described in the Proposition.

Figure 1 illustrates the conflict and consensus regions in the unique equilibrium described in Proposition 1. The two Panels in Figure 1 differ only on switching costs. Given d_A, d_B, the conflict region $\mathcal{C}_{ab}$ in equilibrium based on (σ, ϕ_A) is defined by

$$\mathcal{C}_{ab}(d_A, d_B) \equiv \left\{(\sigma, \phi_A) \mid \underline{\phi}_A(\sigma, d_A) < \phi_A < \overline{\phi}_A(\sigma, d_B)\right\},$$

and the consensus j region $\mathcal{C}_j$ in equilibrium is given by

$$\mathcal{C}_j(d_{-J}) \equiv \left\{(\sigma, \phi_A) \mid \phi_J > \overline{\phi}_J(\sigma, d_{-J})\right\}.$$



(a) $d_J > 0$ for all J.

(b) $d_B < 0 < d_A$.

Figure 1: Consensus and conflict as functions of σ and ϕ_A.

One interesting implication of Proposition 1 is that the majority group norm may not be adopted in a consensus equilibrium. This happens when the minority norm is preferred by all, segregation is low and the majority is not too large. By contrast, if the members of each group prefer their own norm, then a consensus equilibrium must feature the majority norm. When the minority norm is preferred by all, segregation is low and the majority is sufficiently large an intriguing equilibrium emerges. There is consensus over a norm that is privately rejected by all individuals.

3.1. Segregation, Fractionalization and Consensus

We next analyze how the emergence of consensus and conflict depend on the economic environment.

Segregation. Greater segregation implies less interaction between groups, making peer enforcement within groups stronger. As a result, the required incentive to adhere to the leaders' preferred norm may be achieved with a smaller group size, as the following observation states.

Corollary 1. *The thresholds $\overline{\phi}_J$ and $\underline{\phi}_J$ are increasing and decreasing in σ, respectively.*

An increase in segregation lowers the punishment cost of adhering to the leaders' preferred norm while increasing the punishment cost of violating it, irrespective of what the other group is doing. So, for sufficiently high segregation two conflicting norms would have to be an equilibrium, as the following corollary shows.

Corollary 2. *If $\sigma > (1 - \min\{d_J\})/2$, there is conflict regardless of group sizes.*

This result suggests that regardless of population shares, more segregated societies are more likely to have groups adhering to conflicting norms.[8] As figure 1makes clear, the relationship between segregation and conflict is monotone, conflict at a given level of segregation implies conflict at all higher levels. Indeed we can characterize the level of segregation at which the equilibrium (if ever) at consensus switches to conflict.

Proposition 2. *The lowest level of segregation consistent with equilibrium conflict as a function of population shares and switching costs, $\underline{\sigma}(\phi_A, d_A, d_B)$, satisfies*

$$\underline{\sigma}(\phi_A, d_A, d_B) = \max\left\{0, \frac{1 - 2\phi_A - d_A}{2\,(1 - \phi_A)}, \frac{1 - 2\phi_B - d_B}{2\,(1 - \phi_B)}\right\}.$$

The graph of $\underline{\sigma}(\cdot, d_A, d_B)$ as a function of ϕ_A can be seen in figure 1 as the curve separating the consensus and conflict regions. As the figure shows this minimum level of segregation required for conflict is higher when one group dominates in size.

[8] This is consistent with the findings in Corvalan and Vargas (2015) on the effect of ethnic and language segregation on the incidence of civil conflicts at any intensity level.

Fractionalization. We next turn to the role played by the degree of diversity in society in the emergence of consensus and conflict. Fractionalization is a commonly used measure of such diversity and captures the probability with which two randomly selected individuals from society belong to two different groups (see, for instance Easterly and Levine (1997) and Collier and Hoeffler (1998)). In our model fractionalization is captured exactly by the product $\phi_A \phi_B$. The relationship between fractionalization and conflict is not straightforward in that a given level of fractionalization may or may not be consistent with conflict depending on switching costs, size of the majority group and degree of segregation. To describe the relationship indirectly we characterize the effect of fractionalization on the lowest level of segregation consistent with equilibrium conflict. A decreasing (increasing) effect means that greater fractionalization encourages conflict (consensus).

Proposition 3. *(i) If $d_J > 0$ for both $J \in \{A, B\}$, then $\underline{\sigma}(\phi_A, d_A, d_B)$ is strictly decreasing in fractionalization for $\phi_J > \overline{\phi}_J$.*

(ii) If $d_K < 0$ and $J \neq K$, then $\underline{\sigma}(\phi_A, d_A, d_B)$ is decreasing in fractionalization for $\phi_J \geq 1/2$ or $\phi_J \leq \underline{\phi}_J$ and increasing in fractionalization for $\overline{\phi}_J \leq \phi_J < 1/2$.

(iii) If $\underline{\phi}_J < \phi_J < \overline{\phi}_J$ then $\underline{\sigma}(\phi_A, d_A, d_B) = 0$ (and therefore independent of fractionalization).

In words, so long as the population distribution is consistent with consensus at some level of segregation then, with one exception, an increase in fractionalization lowers the minimum level of segregation that supports conflict. Increased fractionalization increases the relative population share of the minority group, making its peer enforcement stronger and therefore less reliant on segregation. The exception arises when members of both groups prefer the minority group leaders' norm, which is also the only candidate for consensus due to the small size of the majority group (see Figure 1b with $\overline{\phi}_A \leq \phi_A < 1/2$). Here increased fractionalization by increasing the relative population share of the minority group makes the minority group leaders' norm even more attractive for the majority group, thereby requiring even greater segregation to make the majority group stick to their own norm.

Costs. Greater preference for one's own group norm is reflected in higher values of d_J. As can be seen from contrasting Figures 1a and 1b, higher d_J increases the range of parameters σ and ϕ_A where conflict occurs. The reason is obvious. If it was an equilibrium to follow your own group's norm, then increasing your preference for that norm would only reinforce the equilibrium. Indeed, now such an equilibrium could emerge with lesser segregation than was earlier required. The following result states this observation formally.

Proposition 4. *The conflict region $\mathcal{C}_{ab}(d_A, d_B)$ is monotonic in d_J, i.e. if $d_J \leq d'_J$ then $\mathcal{C}_{ab}(d_J, d_{-J}) \subseteq \mathcal{C}_{ab}(d'_J, d_{-J})$.*

4. Intensity of Conflict

In our model different conflict equilibria typically yield different amounts of conflict because punishment occurs only when agents adhering to different social norms interact. With almost complete segregation, for example, there is conflict only in the hypothetical sense that if anyone actually met they would punish the partner. Given a conflict equilibrium, the relevant measure of the level of conflict is therefore the expected cost of punishment per capita

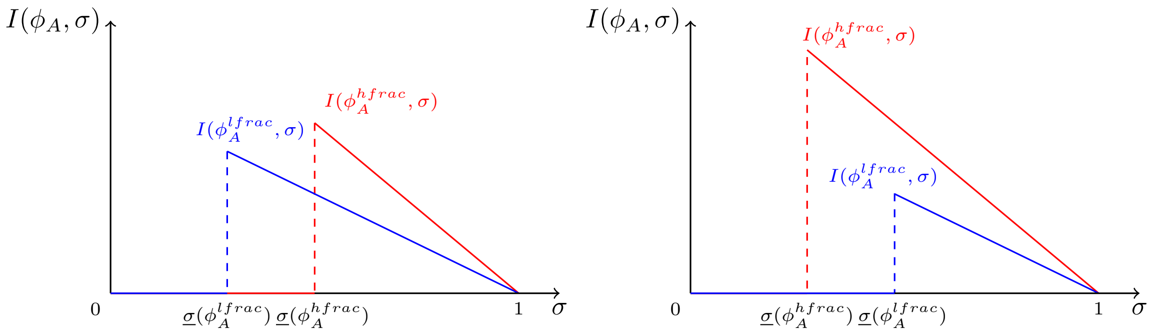
$$I(\phi_A, \sigma) = (1 - \sigma)\phi_A \phi_B P,$$

which we label the *intensity of conflict.* The next proposition simply lists which parameters influence this intensity, and how and follows directly from the equation above.

Proposition 5. *Conditional on conflict, the intensity of conflict is decreasing in the degree of segregation σ, increasing in fractionalization $\phi_A \phi_B$ and independent of switching costs.*

Starting from a consensus equilibrium, increasing segregation eventually triggers the switch to conflict. At this point the intensity of conflict is at its maximum. Further segregation only dampens the intensity since despite clear hostile intent (opposing norms) the two groups meet less and less often . This matches the empirical finding in Bazzi et al. (2019), where segregation dampens the increase in conflict brought about by higher fractionalization. Note that since we focus on just two groups while Bazzi et al. (2019) consider multiple groups, greater polarization in their study corresponds to greater fractionalization in ours. The result is also consistent with the evidence of Field et al. (2008) that incidents of violence were more likely to occur in integrated neighborhoods in the 2002 riots in India.

The change in intensity of conflict with respect to segregation is shown in Figure 2 for different levels of fractionalization. ϕ_A^{hfrac} corresponds to the more fractionalized society and ϕ_A^{lfrac} to the less. Panel (a) corresponds to the exceptional case in proposition 3 part (ii) where an increase in fractionalization increases the lowest level of segregation consistent with conflict. Panel (b) captures the more standard case in which higher fractionalization requires less segregation to generate conflict. Notice though that in both cases, if the level of segregation is consistent with conflict at both levels of fractionalization, the intensity of conflict is higher with higher fractionalization.



(a) $d_A < 0 < d_B$.

(b) $0 < d_A < d_B$.

Figure 2: Intensity of conflict with respect to σ.

Figure 3 offers a different perspective by mapping the intensity of conflict as a function of ϕ_A for different degrees of segregation, $\sigma^{low} < \sigma^{medium} < \sigma^{high}$. It highlights the key observation that while low levels of segregation allow for conflict for a more limited range of population distributions, when it does it generates more intense conflicts.

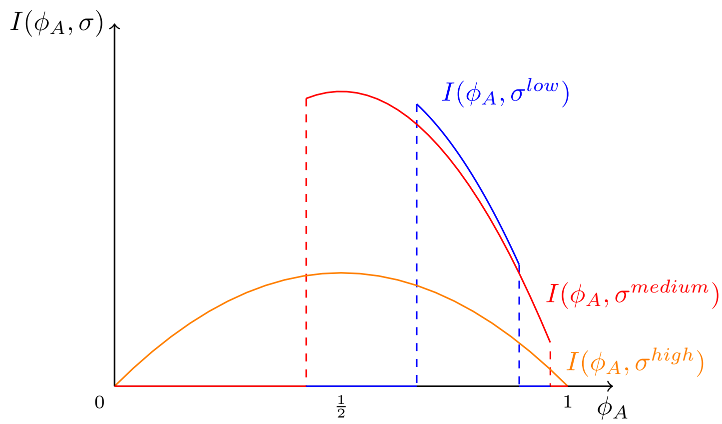


Figure 3: Intensity of conflict with respect to ϕ_A.

5. Segregation and Welfare

This section explores how the degree of segregation affects welfare in equilibrium. In doing so we must allow interactions with people outside one's group to deliver additional benefits, for example, by offering a different perspective or skill (see Hong et al. (1998) and Alesina et al. (2000)). We allow for this by assuming a benefit $U \geq 0$ for a member who meets a member of the other group. Thus, intergroup interactions may be more valuable than intragroup interactions.

A key observation here is that allowing for this additional benefit from inter-group interactions keeps our strategic analysis above unchanged. Indeed, Proposition 1 carries through, as do all subsequent results. The reason is simple. The probability of an inter-group match is unaffected by individual choices. Further, the benefit accrues from meeting someone in the other group, independent of the norms being followed by either party. Therefore the difference in payoffs to an individual from the two norm choices, holding everyone else's strategy fixed, is the same irrespective of the value of U. Finally, the leaders' continue to prefer their own group's norm. As a result, the equilibrium prediction remains the same. With this in mind we turn to welfare analysis. The average expected payoff under conflict and consensus j, respectively, are

$$W_{ab}(\phi_A, \sigma) = 2(1-\sigma)\phi_A\phi_B(U-P) - \phi_A c_{Aa} - \phi_B c_{Bb}, \tag{1}$$

$$W_j(\phi_A, \sigma) = 2(1-\sigma)\phi_A\phi_B U - \phi_A c_{Aj} - \phi_B c_{Bj}. \tag{2}$$

As discussed in Subsection 3.1, a marginal increase in segregation can lead to three possible scenarios. A consensus equilibrium remains a consensus equilibrium, a conflict equilibrium remains as such and finally a consensus equilibrium switches to conflict. In the next proposition we summarize the impact on welfare in these three cases.

Proposition 6. *Suppose there is a marginal increase in segregation.*

(i) At a consensus equilibrium if the type of equilibrium is unchanged then welfare strictly decreases if $U > 0$ and is constant otherwise.

(ii) At a conflict equilibrium if the type of equilibrium is unchanged then welfare decreases if and only if $U \geq P$.

(iii) If the equilibrium switches from consensus to conflict then welfare decreases.

Parts (i) and (ii) follow immediately from equations 1 and 2. If inter group meetings generate a net surplus, then clearly greater segregation reduces welfare. For part (iii), notice that for the group whose norm was the consensus, say J, a move to conflict brings the penalty P from being matched with the other group, $-J$. The latter faces the same penalty but now may face a lower cost from following their own norm. Nevertheless, at the point where the equilibrium switches, it must be that following their own norm is a weak within-group equilibrium for $-J$. In other words, assuming each group follows their own leaders preferred norm then $-J$ members are indifferent between following norm j and getting punished by group $-J$ members and following $-j$ and getting punished by group J members. The expected payoff from the latter is exactly the per capita contribution of group $-J$ members in the conflict equilibrium welfare computation. It must therefore be lower than their contribution to the consensus equilibrium welfare which is simply their payoff from following norm j with no punishment.

An important normative question that remains unanswered is whether a conflict equilibrium can ever generate greater welfare compared to consensus. Proposition 6 shows that to answer this it is sufficient to compare consensus without segregation $W_{ab}(\phi_A, 1)$ to conflict with total segregation $W_j(\phi_A, 0)$.

Proposition 7. *Suppose ϕ_A is consistent with consensus j for low enough segregation. Then, $W_{ab}(\phi_A, 1) \geq W_j(\phi_A, 0)$ if and only if $d_{-j} \geq 2\phi_A U/P$.*

Intuitively, for conflict with total segregation to generate higher welfare the consensus norm must be costly enough for the group with the other norm to outweigh the benefit U from a complete lack of segregation.

6. Discussion

We have examined the relationship between segregation and the choice of norms in a static model. Nevertheless it is easy to see the implications of certain dynamics. Suppose in particular that conflicting norms lead to greater segregation. In this case Figure 1 confirms that once in conflict, such a society would enter a cycle of increasing segregation and persistent conflict, each reinforcing the other. It is not necessary, though, that conflict would then lead to a totally segregated society in a hurry. Recall that the intensity of conflict decreases with segregation. If segregation is increasing in the intensity of conflict, then our model would predict a slowing down of segregation over time. We would expect to see societies caught in a conflict-segregation cycle but sufficiently far from complete segregation.

Schelling (1971) discusses the phenomenon of tipping wherein a minority group enters a neighbourhood in sufficient numbers causing the majority residents to begin evacuating. The key feature is a critical threshold for the minority share, a tipping point, below which not much changes and above which the original majority residents eventually all leave. Card et al. (2008) find evidence of tipping behaviour in a number of US cities, with tipping points ranging from 5% to 20% minority share. Our model coupled with the simple dynamic in the paragraph above generates tipping behaviour. Assuming A to be the majority group, a society with initial segregation σ would have a tipping point of $\underline{\phi}_B(\sigma, d_B) = 1 - \overline{\phi}_A(\sigma, d_B)$. In our theory it is the minority group's choice of norm rather than its mere presence that determines the dynamics of segregation. Interestingly, the tipping point depends on the preferences of the minority and (perhaps more surprisingly) not on that of the majority. The rationale is that the distaste for conflict is what persuades the majority to move. The minority share threshold above which the minority stop adopting the majority norm and instead hold their own, resulting in conflict, is wholly determined by the preferences of the minority.

References

Advani, A. and Reich, B. (2015). Melting pot or salad bowl: the formation of heterogeneous communities. Technical report, IFS Working Papers.

Alesina, A., Spolaore, E., and Wacziarg, R. (2000). Economic integration and political disintegration. *American Economic Review*, 90(5):1276–1296.

Bazzi, S., Gaduh, A., Rothenberg, A. D., and Wong, M. (2019). Unity in diversity? how intergroup contact can foster nation building. *American Economic Review*, 109(11):3978–4025.

Card, D., Mas, A., and Rothstein, J. (2008). Tipping and the dynamics of segregation. *The Quarterly Journal of Economics*, 123(1):177–218.

Collier, P. and Hoeffler, A. (1998). On economic causes of civil war. *Oxford Economic Papers*, 50(4):563–573.

Corvalan, A. and Vargas, M. (2015). Segregation and conflict: An empirical analysis. *Journal of Development Economics*, 116:212–222.

Dutta, R., Levine, D. K., and Modica, S. (2018). Collusion constrained equilibrium. *Theoretical Economics*, 13(1):307–340.

Easterly, W. and Levine, R. (1997). Africa's growth tragedy: policies and ethnic divisions. *The quarterly journal of economics*, 112(4):1203–1250.

Esteban, J., Mayoral, L., and Ray, D. (2012). Ethnicity and conflict: An empirical study. *American Economic Review*, 102(4):1310–42.

Esteban, J. and Ray, D. (2011). Linking conflict to inequality and polarization. *American Economic Review*, 101(4):1345–74.

Field, E., Levinson, M., Pande, R., and Visaria, S. (2008). Segregation, rent control, and riots: The economics of religious conflict in an indian city. *American Economic Review*, 98(2):505–10.

Goyal, S., Hernández, P., Martínez-Cánovas, G., Moisan, F., Munoz-Herrera, M., and Sánchez, A. (2020). Integration and diversity. *Experimental Economics*, pages 1–27.

Hong, L., Page, S. E., et al. (1998). Diversity and optimality. Santa Fe Institute.

Levine, D. K. and Mattozzi, A. (2020). Voter turnout with peer punishment. *American Economic Review*, 110(10):3298–3314.

Levine, D. K. and Modica, S. (2016). Peer discipline and incentives within groups. *Journal of economic behavior & organization*, 123:19–30.

Olson Jr., M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups.* Harvard Economic Studies.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action.* Cambridge University Press.

Sato, Y. and Zenou, Y. (2019). Assimilation patterns in cities. Technical report, IFN Working Paper.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186.

Skarbek, D. (2014). *The social order of the underworld: How prison gangs govern the American penal system.* Oxford University Press.

AppendixA. Proofs

Proof of Lemma 1. Write $P(d_A + d_B) = c_{Ab} - c_{Aa} + c_{Ba} - c_{Bb} = (c_{Ab} - c_{Bb}) + (c_{Ba} - c_{Aa})$. By Assumption 1 and $P > 0$ it follows $d_A + d_B \geq 0$. By Assumption 2, $-P < c_{Jk} - c_{Jj} < P$; $-1 < d_J < 1$ follows by definition. □

Proof of Proposition 1. Suppose all norms are incentive compatible for both group members. The payoffs to the group leaders from the two choices of social norm are given by

		B leaders	
		a	b
A leaders	a	$1, 0$	ϕ_A, ϕ_B
	b	ϕ_B, ϕ_A	$0, 1$

Observe that for the leaders of J their own social norm j strictly dominates $k \neq j$.

We next study incentive compatibility for group members. The expected payoff of a group J member adhering to norm j is given by $-c_{Jj} - \mu_{Jj}P$, where μ_{Jj} is the probability of meeting a partner adhering to a different norm. By Assumption 2, if both groups follow a social norm, it is optimal for everyone to do so. If $(1-\sigma)(1-\phi_J)P < d_J P + (\sigma + (1-\sigma)\phi_J)P$ it is incentive compatible for members of J to adhere to j even if $-J$ members do not and strictly not incentive compatible when the inequality is reversed. This is without loss of generality by Assumption 3. Rewrite this as

$$\phi_J > \frac{1 - 2\sigma - d_J}{2(1-\sigma)} \equiv \underline{\phi}_J(\sigma, d_J). \tag{A.1}$$

If this is the case then the leaders of J will choose j as this is their most preferred norm.

If inequality (A.1) holds for J leaders and the opposite for $-J$ leaders, namely, the following condition is satisfied

$$\phi_{-J} < \frac{1 - 2\sigma - d_{-J}}{2(1-\sigma)}, \tag{A.2}$$

then J leaders will choose j and $-J$ leaders will have no choice but to conform, resulting in the consensus j. The latter, inequality (A.2), may be rewritten using $\phi_{-J} = 1 - \phi_J$ as

$$\phi_J > \frac{1 + d_{-J}}{2(1-\sigma)} \equiv \overline{\phi}_J(\sigma, d_{-J}). \tag{A.3}$$

By Lemma 1, since $d_{-J} \geq -d_J$, we have

$$\frac{1 + d_{-J}}{2(1-\sigma)} \geq \frac{1 - 2\sigma - d_J}{2(1-\sigma)},$$

so that if inequality (A.3) holds so does inequality (A.1). Hence consensus is the unique equilibrium when inequality (A.3) holds for one of the two groups.

By Assumption 3, there are two other possibilities. If both group leaders' dominant strategies are incentive compatible then there is a unique equilibrium where they follow these strategies resulting in conflict. Alternatively, none of the group leaders choosing their own social norm in the face of their opponents choosing theirs is incentive compatible for their members. The theorem follows from ruling out this latter possibility. We show that at least leaders of one group are able to implement their preferred norm in the face of the other leaders doing the same.

Suppose that it is not feasible for J leaders to implement their own social norm in the face of $-J$ leaders implementing $-j$. From reversing inequality (A.1) and by Assumption 3 this requires that

$$\phi_J < \frac{1 - 2\sigma - d_J}{2(1-\sigma)}.$$

Using $\phi_J = 1 - \phi_{-J}$ and $d_J \geq -d_{-J}$ this can be written as

$$\phi_{-J} > \frac{1 - 2\sigma - d_{-J}}{2(1-\sigma)},$$

which implies that it is feasible for $-J$ to implement $-j$ even when J implements j. $\square$

Proof of Proposition 6(iii). Without loss of generality, assume the consensus equilibrium was a. Consider the welfare difference

$$W_a(\phi_A, \sigma) - W_{ab}(\phi_A, \sigma) = -\phi_B(c_{Ba} - c_{Bb}) + 2(1-\sigma)\phi_A\phi_B P.$$

For this to be positive requires

$$P > \frac{c_{Ba} - c_{Bb}}{2(1-\sigma)\phi_A}.$$

Since we are evaluating this inequality at the point where the equilibrium switches from consensus to conflict, we must set $\phi_A = \overline{\phi}_A(\sigma, d_B) = (1 + d_B)/(2(1-\sigma))$. Substituting this above gives

$$P > \frac{c_{Ba} - c_{Bb}}{1 + d_B}.$$

Recall that $d_B = (c_{Ba} - c_{Bb)}/P$. So we have

$$1 > \frac{c_{Ba} - c_{Bb}}{P + c_{Ba} - c_{Bb}}$$

which is always satisfied since $P > 0$. □

Proof of Proposition 7. We prove the statement for $j = a$. A symmetric argument applies to the other case.

$$\begin{aligned}
W_{ab}(\phi_A, 1) &\geq W_a(\phi_A, 0) \\
\iff -\phi_A c_{Aa} - \phi_B c_{Bb} &\geq 2\phi_A \phi_B U - \phi_A c_{Aa} - \phi_B c_{Ba} \\
\iff c_{Ba} - c_{Bb} &\geq 2\phi_A U \\
\iff d_B &\geq (2\phi_A U)/P.
\end{aligned}$$

□

Récents cahiers de recherche du CIREQ
Recent Working Papers of CIREQ

La liste complète des cahiers de recherche du CIREQ est disponible à l'adresse suivante : http://www.cireqmontreal.com/cahiers-de-recherche .

The complete list of CIREQ's working papers is available at the following address : http://www.cireqmontreal.com/cahiers-de-recherche .

08-2020	Benatia, D., R. Godefroy, J. Lewis, Estimating COVID-19 Prevalence in the United States : A Sample Selection Model Approach, avril 2020, 24 pages
09-2020	Kahn, L., F. Lange, D. Wiczer, Labor Demand in the Time of COVID-19 : Evidence from Vacancy Postings and UI Claims, avril 2020, 18 pages
10-2020	Lange, F., T. Papageorgiou, Beyond Cobb-Douglas : Flexibly Estimating Matching Functions with Unobserved Matching Efficiency, avril 2020, 37 pages
11-2020	Baccini, L., A. Brodeur, Explaining Governors' Response to the COVID-19 Pandemic in the United States, avril 2020, 12 pages
12-2020	Gomme, P., Labor Market and Fiscal Policy During and After the Coronavirus, avril 2020, 28 pages
13-2020	Horan, S., Y. Sprumont, Two-Stage Majoritarian Choice, mai 2020, 12 pages
14-2020	Bounie, D., Y. Camara, J. Galbraith, Consumers' Mobility, Expenditure and Online-Offline Substitution Response to COVID-19 : Evidence from French Transaction Data, mai 2020, 54 pages
15-2020	Beaulieu, M.-C., J.-M. Dufour, L. Khalaf, Arbitrage Pricing, Weak Beta, Strong Beta : Identification-Robust and Simultaneous Inference, avril 2020, 57 pages
16-2020	Bahel, E., Y. Sprumont, Strategy-proof Choice under Monotonic Additive Preferences, mai 2020, 7 pages
17-2020	Jones, S.R.G., F. Lange, W.C. Riddell, C. Warman, Waiting for Recovery: The Canadian Labour Market in June 2020, juin 2020, 29 pages
18-2020	Gottlieb, C., J. Grobovsek, M. Poschke, F. Saltiel, Lockdown Accounting, octobre 2020, 13 pages
19-2020	Matveev, D., F. Ruge-Murcia, Tariffs and the Exchange Rate : Evidence from Twitter, septembre 2020, 18 pages
20-2020	Riboni, A., F. Ruge-Murcia, The Power of the Federal Reserve Chair, juin 2020, 56 pages
21-2020	Bouakez, H., O. Rachedi, E. Santoro, The Sectoral Origins of the Spending Multiplier, décembre 2020, 66 pages
01-2021	Svensson, L.-G., T. Andersson, L. Ehlers, Non-Manipulable House Exchange under (Minimum) Equilibrium Prices, décembre 2020, 17 pages
02-2021	Basteck, C., L. Ehlers, Strategy-Proof and Envyfree Random Assignment, décembre 2020, 22 pages