Racial conflict and the malignancy of identity

KAUSHIK BASU
Department of Economics, Cornell University, Ithaca, NY 14853, USA,
E-mail: kb40@cornell.edu

Abstract. This paper demonstrates how our sense of identity can emerge out of mere markers of social distinction that may have no innate significance, but, nevertheless, spread to various aspects of our lives and be the root of conflict. The basis of such conflicts could arise from the use of race to form conditional judgments about people's behavior. Moreover, there are contexts where racial conflict is inevitable even though, *if* individuals had common knowledge of one another's preferences, there would be no conflict. It is argued that this kind of conflict, where many individuals have no innate aggressive preference, is widespread and understanding the process that gives rise to such conflict is the key to crafting effective policy that contains it.

JEL Classifications: C72, D74, D80, Z10.

1. Identity and aggression

Our sense of self, or identity, can influence our social, economic and political behavior. There are people willing to lay down their lives for their flag, boycott products that they believe are produced by an enemy people, detonate bombs to hurt people of another ethnicity. While there may be ways to contort the traditional models of economics and society to explain such behavior, the more reasonable method is to make explicit room for identity in our models. This has recently been argued by Akerlof and Kranton [1, 2], picking up on a theme which, outside of economics, has had currency for quite some time [15, 18, 19, 47, 51].[1] The focus of this paper, however, is not on identity *per se*, but on how our sense of identity can emerge out of almost nothing and then spread to various aspects of our lives and be the root of conflict.

This is not to suggest that all identities take this form. Identity can often be an innate part of our preference, something that enters directly into our utility function (as in Akerlof and Kranton, [1]) and it can also be the glue that cements society. Hence, the paper makes no claims to being a general theory. Its concern is, instead, with a narrow, though important, feature of identity, to wit, how identity could often have little innate basis, but nevertheless acquire importance and even malignancy *ex post*, fomenting strife and conflict.

Bardhan ([6], p. 169) has observed, "[It] is not uncommon to see communities sharing some historical animosities coexisting peacefully [...] for generations (Serbs, Croats and Muslims in the former Yugoslavia, for example) and then something snaps and inter-community violence erupts. Understanding these eruptions requires us to go beyond pointing to primordial antagonisms, examine the nature of that crucial 'something,' and see if it can be placed within some meaningful framework." What the model in this paper does is provide the outlines of a formal game-theoretic model, which can be thought of as a meaningful framework for understanding this kind of racial conflict, which is not pegged to 'primordial antagonisms.'

There are three reasons for studying this. First, there seems to be at least casual evidence that human differences that for a long time are innocuous can over a short time become markers of passion and conflict. Second, the possibility of identity emerging out of virtually nowhere seems paradoxical and is therefore theoretically interesting. And, finally, given the rise in conflict in many parts of the world (the Middle East, Kashmir, Sri Lanka, Northern Ireland), the basis of which is patently that of identity (though, oftentimes, mere symbols of identity that may not have been the basis of conflict for long periods of history), this is a topic that deserves attention.

Each human being has thousands of identifying marks ranging from skin color, ethnicity, and civilizational history, to height, weight, and the contours of her thumb impression. Some of these identifying marks become symbols of identity and reasons of conflict while others are treated as personal idiosyncrasies. We hear of religious wars, ethnic tensions and the coming clash of civilizations but we do not hear of friction between short and tall people, between the bald and the hirsute, or between those who can do mathematics and those who cannot (though there is a friend of mine who believes that, that is the next big clash and one that will dwarf previous ones).

Using the sociology of human life, one can narrow down the list of identifying marks that can be the source of major conflict and those that cannot. For instance, it is unlikely that there will ever be pitched battle between men and women, since they have to share living space (there will, for one, be no reason to wait to get up in the morning to fight).[2] There is little scope for any lasting conflict between the old as a group and the collectivity of young because each of us will feel empathy for the other side, knowing that we will be (or have been) in the same boat at some time. But among the list that survives this paring down, there may be something vacuous about the ones that get chosen and acquire salience. And to the extent that the basis of conflict can undergo major shifts – what seems an unimportant distinction today may become a matter of pride and war tomorrow and *vice versa*. Nationalism and its concomitant, patriotism, seem a natural emotion today and to stand up for it is so often held up as noble. But I remain persuaded that, if wars do not bring the world to a premature end, there will come a time when we will find it baffling and shameful to contemplate that

we once went to war for reasons of patriotism or *national* identity.[3] Today it seems normal to have newspapers or television channels in the U.S. count the tragedy of war by reporting the number of Americans who die each day, but we would find it abhorrent if a newspaper did the same by reporting the number of Christian Americans dying each day or the number of Whites dying each day. But morally there is not much difference between these and hopefully there will come a time when there will be easy agreement on this and tragedies will be measured in terms of numbers of *human beings* who die or get hurt.

One often hears people of race X say that they have nothing against people of race Y, but the latter behave so aggressively towards them that they have no option but to fight back with aggression. As Weingast [55] has argued in the context of Yugoslavia, the damage caused by the 'other' group's aggression can be so large that 'this' group is often compelled to preemptively act aggressively. It will be argued here that one can construct game-theoretically consistent models to illustrate such behavior. Racial conflict may emerge from no innate racial preference. The basis of such conflicts could be founded in the use of race to form conditional judgments about people's behavior. Moreover, there are contexts where racial conflict is inevitable even though, *if* individuals had common knowledge of their preferences, there would be none. Understanding the process that gives rise to such conflict is the key to crafting policy to restrain such conflict. Though I do not comment directly on policy in this paper, the theory that is developed here is motivated by the need to inform policymakers about what causes identity strife and flare-ups and how they can be doused.

The kind of formal game-theoretic argument that I use here has been developed and used by others in related areas [5, 13, 33, 40]. The novelty of the paper lies in the context. The use of Bayes–Nash equilibrium to analyze racial conflict seems to be the natural choice but is rarely used in the literature. The aim of this paper is to propose and defend such an analysis and set the stage for analyzing game-theoretically the myriad roles that identity plays in our social and economic lives. It is shown that there are social contexts where racial conflict can get triggered even though most people concerned are bent on avoiding it. Game theory turns out to be a valuable instrument for understanding such social processes.

The use of game theory and strategic equilibrium concepts for understanding social tensions and political conflict occurs in a relatively brief literature (see [12, 43]). One paper that explicitly uses this method to study political uprising is by Bates et al. [11]. In their paper they try to understand the long period of murmuring dissent and the eventual uprising in 1991, which dislodged Kenneth Kaunda and his United National Independence Party from power in Zambia. As in our model, critical use is made of incomplete information among the agents. But the similarity ends there, since I am interested in economies where *every* individual is, in a sense, powerless, and individuals critically use race or other identity markers to form conditional expectations of other people's behavior.

One of the most critical results of this model is how the injection of a few people with special preferences can transform a society in harmony to one in which racial groups are Balkanized and pitted against one another. In particular, the arrival of one new person, aggressively inclined, can transform a peaceful nation into one in which people of two races or religions are pitted against one another (and that is in fact the only equilibrium). Hence, in curing the problem it is important to recognize that a randomly selected person who is aggressive against another group may have no *innate* aggressive preference. Hence, this paper resists the view of conflict where the 'other' side is thought of as innately wicked. This is a tendency that has been with us from primitive times to the contemporary – to view those on the other side of the bush as evil.[4]

This tendency has been a major hindrance to understanding conflict and stopping it. And so I hope that this theoretical inquiry will contribute to the design of more effective policy. But, as is discussed later, there is another way in which models of this kind can influence social outcomes. This is by prompting *individuals* (in contrast to governments and policymakers) to alter their behavior. What textbook economics does not recognize but is true is that human beings have an innate tendency to contribute to public goods and causes even when that is not in their direct self-interest. So, once we understand that a particular behavior of ours can contribute to a public good or a public bad, we are inclined to, respectively, undertake or desist from that behavior. This paper shows that if (1) all of us are a bit more cooperative towards 'others' (that is, people who carry different identity markers) or (2) we refuse to use race or religion to form conditional judgments (even when there is information in race or religion), then the social outcome could dramatically change from bad to good. What I am arguing is that the mere understanding of this logic can prompt individuals to behave as in (1) and (2). Hence, the model has 'policy' implications not just for policymakers and government but for ordinary individuals.

This approach should also open the way for richer analyses and the examination of other aspects of identity conflicts. It is, for instance, known that people have very different powers of recognition, when it comes to people of their own race and those of another. This is a topic on which much has been written (see, for example, [30, 44]). To a White colonial master all Rwandans might have appeared indistinguishable but among the Blacks of that region there may be no problem in telling who is a Hutu and who a Tutsi.[5] This general idea finds reinforcement in the experiments of Li et al. [28] that found that Whites, who are basketball fans and therefore familiar with Black features (a disproportionately large number of basketball players being Blacks), are better at recognizing Black faces than Whites who do not watch basketball. Subjects were shown a number of unknown Black faces on a video monitor and later tested for their ability to recognize those faces. Blacks and White basketball fans did equally well and significantly better than Whites who were not basketball fans.

Recognition at the level of a collectivity can give dignity to a group, as Taylor [49] has argued effectively. But recognition (in the sense of having the ability to distinguish between individuals or sub-groups) can do more. It can act as a brake in the escalation of identity wars. When a person from a race or religious group X acts aggressively towards another, if the latter can identify the aggressor as belonging to a smaller group, for instance a sect of X or a member of X with a specific history, then this act of aggression will not taint his view of everybody of race X and this can be a crucial step in the prevention of escalation of aggression. That one should get to know one's enemy is therefore not just highfalutin moral advice, but a recommendation for peace that has foundation in formal theory. While the paper does not go formally into these larger questions it tries to set the ground for such investigations in the future.

2. Basic model

Consider a nation with set S of individuals. Each individual has certain visible characteristics, that is, visible to all, and some invisible characteristics that she knows about but others do not. For reasons of simplicity, we will take the former to be race, where one's race can be Black, B, or White, W. And we will take the invisible characteristic to be (for now) a variable that captures a person's affinity for his own type (i.e., who is visibly like him). For person $i \in S$, this variable is captured by $c_i \in [0,3]$. The significance of c_i will become clear later on. For now it is enough to think of $c_i = 3$ as the kind that has the strongest innate affinity for his type. To sum up, each person i in this society is characterized by an ordered pair (x_i, c_i), where $x_i \in \{B,W\}$ and $c_i \in [0,3]$.

Interaction among human beings is described here in terms of a 'basic game,' G, described below.

Basic Game, G:

		Player j	
		A	C
Player i	A	2, 2	8, 1
	C	1, 8	10, 10

Individuals are drawn randomly from S and matched against one another to play G. The 'payoff' in G should be thought of as direct visible payoffs that players get. For simplicity, we may think of the numbers as dollar payments and these 'payoffs' are common knowledge. The strategies A and C denote 'aggressive' and 'cooperative' behavior. The game captures the idea that co-operative behavior leads to desirable outcomes, but people do not like to be cooperative when one's opponent is being aggressive. There are many different

interpretations possible for this game. Action A could be an act of violence against the other player and the game could represent people's penchant for hitting back at someone who hits you. Another interpretation, the more widely applicable one, is where C stands for cooperative behavior, for instance, in business, and A stands for cut-throat behavior in business. To be cooperative against a cut-throat is to do badly. If both can be cooperative, they both do well. It would have been possible to use a more elaborate game (as, for instance, in Basu, [8], Section 4.6) where being cut-throat in business is always in one's atomistic interest, but for the purposes of this paper the additional complication would not add much. In other words, the idea being captured here is in the spirit of Fukuyama's [17] discussion of how trust among a people can help progress (see, also, [37], Section 6.3). In this game, if a group of people can trust one another and play cooperatively, they will earn more and prosper.

In this basic game there are two Nash equilibria (A,A) and (C,C). It can be checked that, not only does (C,C) Pareto dominate (A,A), but (C,C) is risk dominant.[6] There are game theorists who believe that when of two equilibria one happens to be risk dominant, then that is the outcome that will be selected ([53], p. 31). Though nothing of great significance hinges on this let us here go along with this assumption. In addition, it has been argued by some that players have a tendency to choose the Pareto-optimal Nash equilibrium, either because of a focal point argument or the view that pre-play communication will lead to that ([37], Section 6.4). In this example risk dominance and Pareto dominance coincide. So from now on we will assume that whenever two players are called upon to play this game and the game is common knowledge, the outcome will be (C,C).

The game that people actually play is however not G because there may be psychic costs and benefits involved in being aggressive and cooperative with different people, over and above the commonly visible payoffs described in G.[7] The actual game that two players i and j, picked at random from S, play is given by G^*. Note that if $d_i = d_j = 0$, then G^* reduces to G.

Game G^*:

		Player j	
		A	C
Player i	A	2, 2	8 $1 - d_j$
	C	$1 - d_i$, 8	$10 - d_i$, $10 - d_j$

The d_i and d_j refer to the psychic payoffs. We will consider different interpretations of these. In this and the next sections it is assumed that these are determined as follows: If both players are of the same visible type (in our current example, race), that is, $x_i = x_j$, then $d_i = d_j = 0$. If, however, $x_i \neq x_j$, $d_i = c_i$ and $d_j = c_j$. Recall c_i and c_j are to be interpreted as follows: c_i is the innate unfriendliness

that person i feels towards an alien, that is, someone different from him. If $c_i = 0$, then person i makes no distinction between his own type and an alien. In the present context, he is race blind. As long as $c_i < 2$, a player prefers to respond to cooperative behavior with cooperation. We shall refer to c_i as person i's *latent characteristic*. It could be thought of as an index of a person's chauvinism. If $c_i > 2$, i prefers to be aggressive *vis-a-vis* anybody who is visibly different from him irrespective of what action the person chooses.[8]

What I first want to demonstrate is that, even if people have $c_i < 2$, playing aggressively against the other race may be the only rational strategy for them. Expecting aggression from the other side, we respond with aggression; and moreover, given some weak conditions, this is the *only* expectation (i.e., the expectation of aggression) that can occur in equilibrium. Huge amounts of racial animus can be founded on the most flimsy basis.

If i and j both belong to the same race, clearly there are two Nash equilibria and, in keeping with our assumption, we will expect them to reach the Pareto superior and risk-dominant outcome, namely, (C, C).

Now suppose $x_i \neq x_j$. That is, one player is W and the other B. In deciding what to choose, A or C, a player now has the handicap of not knowing the other player's latent characteristic and not knowing the other player's belief about his latent characteristic.[9] As usual, under such circumstances, the natural equilibrium notion to use is that of Bayes–Nash.

To start with, let me assume that the *a priori* distribution of latent characteristic in society is independent of the person's race. Let $F(c)$ be the fraction of population with $c_i \leq c$, within each race. By denying a subscript to F, I am assuming that both races have an identical distribution of latent characteristics.

I shall now specify a sufficient condition on F such that, given that condition, the only equilibrium will be one where *everybody* is aggressive *vis-a-vis* people of the other race.

ASSUMPTION 1. F is such that, for all $c \in [0,2]$, $F(c) < (1 + c)/3$.

In Figure 1 let the broken line, ABD, denote the distribution function $F^*(c) = min \{(1 + c)/3, 1\}$, $c \in [0,3]$. Hence a distribution function satisfying assumption 1 is any non-negative monotonic function F the graph of which is strictly below F^*, for all $c \in [0,2]$ and below F^* for all $c \in [2,3]$. GHD depicts an F satisfying Assumption 1.

It is obvious that an F, where $F(2)$ is virtually equal to 1, can satisfy Assumption 1. Hence Assumption 1 is compatible with a case where virtually no one prefers to be unilaterally aggressive *vis-a-vis* the other race.

THEOREM 1. Given Assumption 1, the only Bayes–Nash equilibrium is one in which, whenever a person is matched to play against a person of a different race, he chooses to be aggressive.
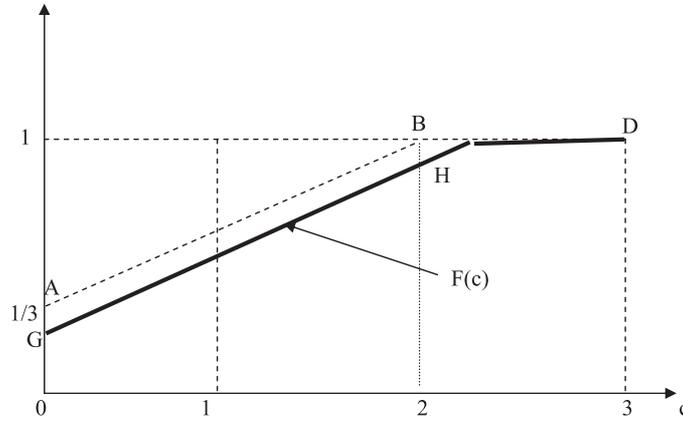
Figure 1. Distribution of latent characteristics.

Proof. The proof of this is straightforward. But instead of giving the shortest proof, I shall give one that has greater transparency and will aid us with some of the discussion that follows. For this it is useful to take the route of what may be christened 'quasi-reaction correspondences.' Let φ_B and φ_w be the probabilities of a randomly picked Black and White person playing A (aggressive) when they are pitched against each other. Clearly, in equilibrium, the values of φ_B and φ_w will depend on each other. A Black with latent characteristic c will be willing to play A if and only if:

$$2\varphi_w + 8(1 - \varphi_w) \geq (1 - c)\varphi_w + (10 - c)(1 - \varphi_w).$$

or

$$c \geq 2 - 3\varphi_w.$$

Hence, the probability of a randomly picked Black playing A is given by $1 - F$ $(2 - 3\varphi_w)$. Therefore, if Blacks responded optimally, given φ_w, we would have:

$$\varphi_B = 1 - F(2 - 3\varphi_w). \tag{1}$$

This is a slightly incorrect way of writing B's quasi-reaction correspondence since it presumes the probability of each c to be of measure zero and so hides the fact that when $\varphi_w = 2/3$, a positive measure of people are indifferent between choosing A and C and so φ_B is a correspondence. Figure 2 illustrates this correspondence. It is depicted by the graph EFJK. The segment EF reflects the segment GH in Figure 1. Since by Assumption 1, GH lies below AB, a little thought shows that the entire segment EF must lie below the 45° line through the origin in Figure 2 (shown by the broken line).

Since the two races are assumed to be identical, W's quasi-reaction correspondence, $E'F'J'K$ is symmetric to B's quasi-reaction correspondence.
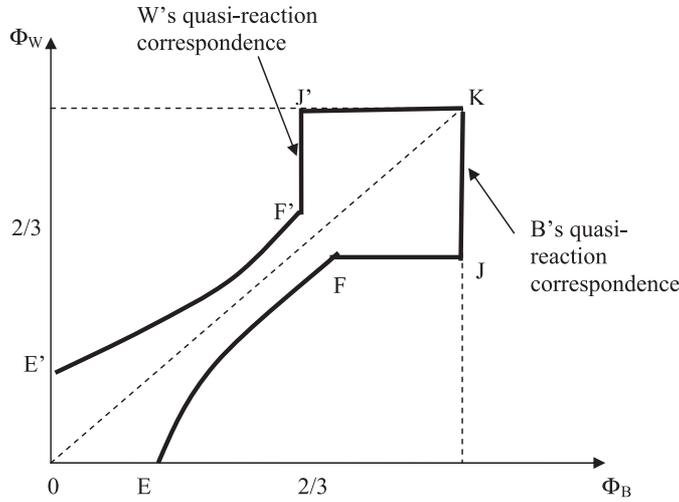
Figure 2.   Reaction correspondences.

If φ_B is the proportion of Blacks who are aggressive, we can use *W*'s quasi-reaction correspondence to read the proportion of Whites who will be aggressive. For φ_w, φ_B to be a Bayes–Nash equilibrium, each must give rise to the other. That is, they must satisfy (1) and

$$\varphi_w = 1 - F(2 - 3\varphi_B). \tag{2}$$

Given Assumption 1, the only point where this can happen is *K* in Figure 2, that is when all Blacks and all Whites choose to be aggressive towards the other race. This completes the proof of Theorem 1. □

The intuition behind the theorem is this. Consider a society in which there are two groups and no one wants, innately, to be aggressive towards members of the other group, though, like in all coordination or assurance games, if the likelihood of the other person being aggressive towards you is high then you will respond with aggression. How likely the other person's aggression has to be to provoke aggression by you, however, differs from person to person. In a large society the distribution of this parameter can be quite smooth, going from those who need very little likelihood to provoke them to those who need a very high likelihood. Now suppose this society has a few additional persons who are innately aggressive (that is, aggression is their dominant strategy). Now, this will mean that those individuals, who are prone to being aggressive if there is a small chance of aggression by their opponent, will now be aggressive (since people use group-conditional expectations when calculating the probability of aggression on the part of the other player). But once these people choose to be aggressive, others, who needed less provocation to be aggressive, might want to be aggressive. And once these people choose to be aggressive, those who needed even less pro-

vocation may choose to be aggressive. And so on. Cooperative behavior vanishes
like a trail of collapsing dominos.

If we are free to consider other kinds of F (that is, we are not restricted by
Assumption 1), it is easy to see that there may be other Bayes–Nash equilibria.
One such case is illustrated in Figure 3.

In this case there are three possible equilibria, of which two are stable – at 0
and K. I use the word 'stability' in the sense in which it is used in phase diagram
analyses. Suppose the society is in equilibrium U (in Figure 3). Hence, x (y) is
the proportion of Blacks (Whites) who are aggressive. These proportions are
mutually compatible, U being an equilibrium. But now suppose a few additional
Blacks act aggressive (by mistake or because of some exogenous shock). That is
x moves slightly to the right. We can see from W's quasi-reaction correspon-
dence that some more Ws will now choose to be aggressive. And this, in turn,
will make many more B's choose aggression (one can read this off B's reaction
correspondence). The process leads inexorably to point K. It is in this sense that
U is thought of as unstable and K and 0 stable.

If the society is at 0, then there is no inter-racial animosity. However, as must
be evident from the figure, the injection of a certain critical mass of people, who
have $c > 2$, (this is different from one or two persons being aggressive by
mistake) can cause a sharp change in inter-racial relations, with racial conflict
being the only equilibrium.[10]

To understand the paradoxical nature of this result let us see what the average
value of c for an $F(c)$ satisfying Assumption 1 looks like. So suppose

$$F(c) = \min\left\{\frac{1+c}{3} - \delta, 1\right\}$$
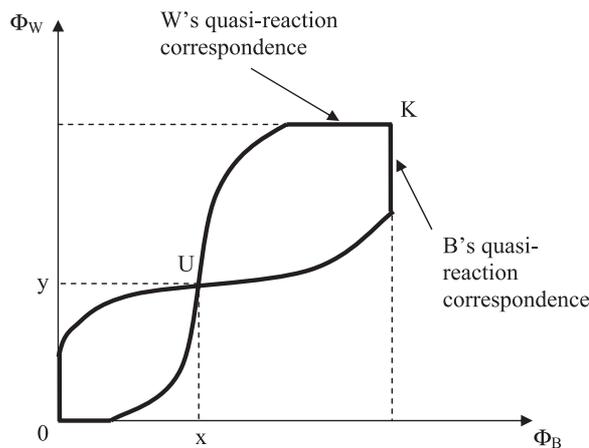
where $\delta \in \left(0, \frac{1}{3}\right)$.

Figure 3. Multiple equilibria.

Hence the average value of c, denoted by Ec, is given by:

$$Ec = \frac{(2 + 3\delta)^2}{6}.$$

Note that as $\delta \to 0$, $Ec \to \frac{2}{3}$. Hence, we can think of societies where Assumption 1 is satisfied and the average value of the latent characteristic is virtually equal to 2/3. Hence, in such a society when you confront a person of another race, you expect that the other person's latent characteristic to be way below 2. The other person expects the same of you. These expectations are in fact common knowledge. So, on the basis of expected preference, it seems plausible for both of you to play C and for both players to have an expected payoff almost equal to $9\frac{1}{3} = 10 - \frac{2}{3}$ each. But given the theorem just established we know that this cannot happen. The only rational behavior on the part of every agent when he confronts a person of the other race is to play A and so each player earns a payoff of 2. Conflict is the unique equilibrium.

It is worth noting that this model could be used to explain strife even in a homogenous society. Suppose people have no observable characteristic or, equivalently, everybody shares the same observable characteristic and the latent characteristic expresses a generalized aggressiveness towards any other player. In that case, as in our model, one person's dominant preference for aggression can lead to aggressive or non-cooperative behavior on the part of everybody.

3. Innate and ex post racism

One important way in which the above model can be generalized is to take away race-consciousness from everybody excepting the person(s) whose latent characteristic exceeds 2. What I mean is this. Suppose, for every person whose psychic cost, c_i, exceeds 2, (that is, those who prefer to be aggressive irrespective of how the other player plays), the cost occurs, as before, only across racial lines (that is, only when the other player is of a different race), but for everybody whose latent characteristic or psychic cost is less than 2 this cost occurs *no matter whom* he is playing against. That is, c_i (the unseen cost of being co-operative) occurs in these people no matter who they are playing against. In other words, everybody except the few people who had a dominant preference for aggressive play against other races now are totally race blind.

It is easy to see that the Bayes–Nash equilibrium remains exactly as described above (assuming Assumption 1 holds). People are always nasty across racial lines. Racial identity in this model emerges from virtually no race-identity variable in people's utility functions. The word 'virtually' is important because some innate identity preference is needed (in this example, we assume this is true for a few extreme people) for the balkanization of society to occur.

There is another possible modification which will illustrate how a *quasi-*Balkanization of society can occur with *no* identity variable in anybody's pref-

erence. Suppose in the model in the previous section there were n Whites and n Blacks and Assumption 1 were satisfied. Now consider a society in which there are n Blacks, as before, but only $nF(2)$ Whites and this is achieved by eliminating the Whites who had the highest values of c_i. Moreover, assume that all the people have preferences as described in the previous section, with the additional feature that the psychic cost c_i occurs no matter who one is playing against. In their preferences people have no sense of race (not even the few extremists who were allowed this in the example above). Note now that only $n[1 - F(2)]$ persons in the group B have an unconditional preference for A (against Whites *and* Blacks). Since race is a visible characteristic it is easy to see that every Bayes–Nash equilibrium will have the following feature: Blacks choose action A against Whites and Blacks, and Whites choose action A against Blacks. We can have a Bayes–Nash equilibrium which has this property and, in addition, has Whites playing cooperatively against Whites, which is not just Nash but risk dominant. This mimics the kind of colonial control the world witnessed in the eighteenth, nineteenth and early twentieth centuries when the colonial rulers acted cooperatively to control large native populations that were divisive and full of atomistic opportunism among themselves. In this case, the race-based behavior occurs though people have completely race-neutral preferences.

4. Idiosyncratic information, cultural traps and secret societies

The model can potentially be extended in many directions.

The assumption that human beings have some characteristics that are invisible and some that are visible (with the implicit assumption that these are visible to all) overlooks the other realistic possibility, where people have characteristics that are idiosyncratically visible, that is, visible to some and not to others. Consider a case where people know one another within their own racial groups, but cannot distinguish between individuals in the other racial group.

In other words, if person i, who is White, is aggressive to a White person, the information that registers in the latter's head is not simply that "one White person was aggressive to me" but "person i from the set of Whites was aggressive to me." On the other hand if person i, who is Black, were aggressive to the same White person, the information that registers in his head is that "a Black person was aggressive to me."

Since in the formal model here one does not learn about persons or races through one's actual experience (but carries probabilistic information in one's head, which is valid, in equilibrium), the observation in the previous paragraph cannot be formalized directly. One route for doing this, is to ask how one would, *within the framework of this model*, formalize a remark like "when Jack sees i, who is Black, Jack recognizes exactly who he is, instead of treating him as just one Black." Since, in this model the only latent variable is c_i, a reasonable way to

interpret the remark, "Jack recognizes person i," is to treat it as equivalent to "Jack knows the value of c_i."

What then would "recognizing *sub-groups*" mean? Suppose among Whites, there are two sub-groups, Croats and Serbs. Let us suppose that the distributions of the latent characteristic among Serbs, Croats and all of them pooled together are given by F_s, F_c, and F. Then, one way to interpret "recognition of the groups Serbs and Croats" is to say that when Jack sees a Serb, Jack knows that the person is drawn from F_s.

Whether one uses this or some other method to formalize this, the idea of 'recognizing faces or sub-groups' is evidently an important one that deserves formalization in the future. Such an analysis could provide a formal basis to the oft-heard recommendation that to control ethnic or racial conflict [52] it is important for people to 'get to know' the other side. The reason why this works may not be quite the same as what prompts this advice, which often has to do with how knowing someone fosters warmth and affection. The conjecture here is that if we know the other side, we are able to think of one act of aggression as an act of aggression by one particular person or a sub-group, instead of an aggression by a race or an entire ethnic group. This can prevent conflicts from cascading out of hand.

This can be combined with another interesting idea on partial information. Suppose a person's latent characteristic is not altogether latent, but instead a person emits an imperfect signal to the viewer, as in Morris and Shin [33]. Thus a person with a latent characteristic c emits a signal of his characteristic, which is chosen from a uniform distribution on $[c - \eta, c + \eta]$, where $\eta \geq 0$. Hence, when another player receives a signal c from his opponent in a game, he knows that she must have a latent characteristic in the interval $[c - \eta, c + \eta]$. If $\eta = 0$, then players have common knowledge of their types. However, it seems likely that if assumption 1 holds and η is above a certain size, then the only equilibrium will be the aggressive one. What will be interesting is to introduce the modification that the signal received depends on who the recipient of the signal is. If the recipient happens to be someone who knows the player well, then the η will be small. This may be an interesting route to explore in modeling the possibility of people knowing one another differentially.

Much has been written about cultural traps, why people fall into certain stereotypical behavior patterns even though that may not seem to yield any benefit [46, 56].[11] As Swidler ([46], p. 275) quizzes, "Why doesn't a member of the 'culture of poverty' [...] take advantage of opportunities to assimilate to the dominant culture in conduct and dress, acquire the appropriate educational credentials, and settle down to a steady job?" One way to answer this is to argue how human preferences can vary and, interestingly, there can be a self-fulfilling element to this. Once one recognizes oneself as belonging to a certain group one develops a preference for the symbols of that group.

But by using the structure of analysis developed in this paper one can also argue that even when people's preferences remain unchanged and they would

prefer to be in the shoes of the other race, they may not wish to exercise the option of changing their culture. The argument I am suggesting here has much to do with 'recognition' [49]. Suppose now that an individual has three traits: an innate, publicly observable trait (like race or nationality); a chosen cultural characteristic, such as dress-style, coiffure, or college degree; and a latent characteristic, as before. What is new here is the chosen cultural characteristic. Suppose this can be of two types: Green and Blue. That is, each person can choose to be Green or Blue. If the chosen characteristic is one that is idiosyncratically visible, in particular, it is visible only to your racial type, then it is easy to see why no one may want to switch. Consider a model just like the above one, but with fewer Blacks than Whites and with all Blacks choosing Green and Whites choosing Blue. Let us suppose the latent characteristic in this paper depicts one's animus towards anybody who looks different from oneself.

By now it is easy to see how we can have an equilibrium in which i and j are aggressive to one another if and only if they belong to different races. Since one earns a larger payoff the more one cooperates and there are fewer Blacks, Blacks are worse off in this equilibrium. Now consider one Black who wants to wave a conciliatory signal to the Whites to say that he is willing to cooperate with them (for instance, in business ventures). If he tries to do this by altering the chosen cultural signal, that is, by choosing Blue instead of Green, he will be worse off. Since, to Whites all Blacks are the same (that is, they do not take notice of the chosen variable), their behavior will not change towards this person. And since other Blacks will notice this person's altered cultural trait, some of them (those with sufficiently high c) will now be aggressive towards him.[12] Overall he will be worse off. Hence, Blacks choosing Green and Whites choosing Blue will persist.

The anthropologist, M.N. Srinivas, coined the term '*sanskritization*' [45] to denote the conscious change in culture that some lower caste groups in India affected in order to get some of the advantages of the upper castes, and also noted why this phenomenon was not more widespread. This had much to do with the opprobrium one faced for attempting to be different (see [7]). Not surprisingly, the successful cases of *sanskritization* were often associated with groups that migrated to and started life afresh in a new locale.

There are other interesting phenomena that this model can be used to address. The world has seen, mercifully not too often, the emergence and flourishing of 'secret societies.' By a 'secret society' I mean a group of people who know that they belong to a special group – let me call this group S – but the larger society to which they belong is unaware that they share a separate (and among-them known) group identity. Free Masons in some stages of history may have functioned as such a group. Underground revolutionary movements often have the character of a secret society. CIA agents, trying to infiltrate a society, may recognize one another as CIA agents though no one else may know this.[13] Such secret societies are a *social* counterpart of the military strategy of the Trojan

horse. They form a special, coherent group within a larger society that is unaware of their secret association.

To understand how this could play out in the model, we could consider a society of *2n* people, with two sub-societies, B and W, with a population of *n* each. A person of type B cannot tell the difference between a B and a W, whereas a W (maybe through a secret handshake, meaningful only to those who know the handshake) can tell if a person is W or B. In other words we are assuming that all people (B and W) have the same skin color and looks but when people shake hands a W person wiggles his middle finger and only a W person realizes that this is a signal that the wiggler is W. Clearly in this case a W person will be able to condition his behavior on whether his gaming partner is a W or a B, but a B person will not be able to use such a conditional strategy. This opens up a host of analytical possibilities that will be worth studying in the future.

5. Comment on policy

Despite the abstract character of the paper, its main motivation is to help us craft policy to diffuse racial and other identity-based tensions and turmoil. The first step towards this is to understand the politics of identity and how identity, through the simple mechanism of individuals using statistical information and group characteristics to form expectations about the behavior of individuals, can come to acquire certain malignancies. Innocuous differences turn into significant markers for which people are willing to die. To prevent this from happening we need to desist from certain individually rational behaviors, just as one way to control pollution is for us to behave socially responsibly and desist from smoking in a crowded waiting room and discarding plastic bags carelessly. Such behavior may not be individually rational, but civil society relies on the fact that once human beings realize that certain actions on their part entail minor individual sacrifices but lead to large social gains, they somehow develop those behavior traits as reflexive action.

Though this is contrary to standard wisdom in economics, understanding and appreciating the positive or negative externalities of our individual actions often encourages us to undertake the actions with greater or diminished zest even if this were contrary to one's self-interest. People are often persuaded to switch off the light when leaving a room by being made aware that if everybody was careless about this, the whole country could suffer an electricity shortage. Good citizens do not throw garbage out of their car windows, even if in their perception that makes no difference, simply because of the awareness that if everybody threw garbage out onto the streets that would make the city unclean.[14]

Hence, a better understanding of a phenomenon and our miniscule role in that can have a policy impact (even though this may imply a violation of the textbook rationality principle). If we realize that the other side is not innately evil and that our own tendency to form race-conditional expectations of other people con-

tributes to conflict, we are likely to resist this tendency, *even though each of us may individually benefit by using race-conditional expectations.*

Moreover, the model developed in this paper shows how a tiny change in our preference (everybody's or even one or two persons') can make a big difference.[15] Hence, this gives government and civil society some hints on how to 'educate' people.

The recognition of this externality problem also gives rise to the scope for within-group policy action. Suppose people can police the behavior of individuals in their own group, then there may be a case for each of us to try to change the behavior of individuals within his or her group. There are examples of this in real-life. In India, the Marwaris are a major trading group. Much of their success depends on how much people trust them. If people use group-conditional information to form judgments about how much to trust an individual Marwari trader, then it is in the interest of the group to monitor their own behavior so as to make them act more trustworthy, even though it may not be in the interest of each individual, atomistically, to do so. Kudaisya ([25], p.38) discusses how the "Marwari panchayat" (a local council of the Marwaris) took the initiative in punishing members of their own community who cheated customers. For instance, in 1917 (before the colonial government of India had enacted an anti-adulteration act) complaints were received of some Marwari traders adulterating food. In response, the Marwari community leaders investigated the matter, called a panchayat meeting and fined the traders found guilty and ex-communicated them. In situations of conflict, *of the kind* modeled in this paper, there is clearly advantage to instituting policies for in-group monitoring of behavior. In my model, if this is done successfully, it will alter the behavior of the other group as well.

What I have been talking about thus far are policy suggestions and recommendations for individual and within-group behavior that are prompted by the incomplete-information game analyzed in this paper. This must not detract from more commonly discussed policy recommendations, such as the need to have appropriate institutions, to have constitutional guarantees of basic rights to all individuals irrespective of their color or creed, to have a leadership that reflects a nation's diversity, and to have appropriately designed territorial and federal structures ([6, 23], Chapter 9). But it may be inadequate to stop at these conventional interventions. Also these conventional policies may be more costly than some of the ones that get prompted by the model in this paper. For one, the stress in this paper has been on the resolution of conflict by trying to alter the equilibrium, rather than to sustain a disequilibrium outcome through continued structural intervention.

In essence, this paper suggests a way of thinking about racial conflict or the clash of civilizations, which is at variance with popular thinking. On conservative radio and television channels one often hears remarks like: "The trouble with the 'other' community or religious group is that no one in that community expresses outrage against extremism and fundamentalism." Underlying statements of this

type is the presumption that the 'other community' is somehow different. The present paper, on the other hand, suggests that a more useful approach is to begin from the presumption that large groups are, at a sufficiently fundamental level, similar. So faced with a special problem or behavior trait of a community or religious group, the right approach is to inquire into what it is in the circumstances or experience of the group that makes them behave in that way. In other words, we are likely to be more successful if we seek the explanation, ultimately, in the *equilibrium* behavior (which may require us to go several steps back into previous experience), rather than in *innate* differences.

Acknowledgements

Notes

[1] In economics, identity appears in a more marginal way and that too mostly in recent times. See, for instance, Miguel and Gugerty [31], Platteau and Strzalecki [38], Rao and Walton [39], Austen-Smith and Fryer [4], Fryer and Jackson [16], Hoff and Pandey [22], Tangeras and Lagerlof [48], Loury [29], and Sen [41].

[2] This not to deny that there is conflict between the sexes in society and even within the household, but this takes very different forms from racial conflict. The conflict between the sexes often takes place within the broad parameters of cooperative behavior, thereby inviting the label 'cooperative conflict' [9, 42].

[3] A similar moral line is taken in several essays in Nussbaum and Cohen [35] and especially the opening essay by Nussbaum [34] which draws on some remarkable lectures given by the poet Rabindranath Tagore, arguing against nationalism. For a moving commentary on how nationalism can be put to venal use and give rise to 'collective psychosis' see Hedges [21].

[4] As Hume [24] wrote, "When our nation is at war with any other, we detest them under the character of cruel, perfidious, unjust and violent: But always esteem ourselves and allies equitable, moderate and merciful." To the extent that both sides in a war hold this view, at least one of them must be wrong. The paper argues that typically both are wrong.

[5] This may not be a particularly good example because the Hutu–Tutsi distinction is one of those constructed differences that arguably have few biological or distant-historical markers.

[6] The idea of risk-dominance, originally developed by Harsanyi and Selten [20], may be explained as follows. Consider a 2-by-2 coordination game in which each of two players can

choose between actions a and b and let $\pi_i(x, y)$ be the payoff earned by player i when players 1 and 2 play x and y, where $x, y \in \{a, b\}$. Since this is a coordination game, we assume that (a, a) and (b, b) are Nash equilibria. Of these, (x, x) is defined to be *risk dominant* if $(\pi_1(x,x) - \pi_1(y, x))(\pi_2(x, x) - \pi_2(x, y)) > (\pi_1(y, y) - \pi_1(x, y))(\pi_2(y, y) - \pi_2(y, x))$, where $x \in \{a, b\}$ and $y \neq x$. (Some authors call this *strictly risk dominant*.) In a symmetric game, such as G, risk-dominance translates into a simple and intuitive idea. Suppose each player believes that there is equal probability that the other player will choose each of the two strategies open to her. The strategy that each player chooses under this presumption is part of his risk-dominant strategy (see [57], chapter 4).

[7] This general point that the payoffs that are explicitly displayed may not be the actual payoffs perceived by the players, even in a laboratory setting (let alone life) has been discussed in the literature (see, e.g., [3, 54]).

[8] Rush Limbaugh is a name that easily springs to mind. As the newspaper columnist, Deborah Mathis, perceptively observed (*Ithaca Journal*, October 17, 2003, page 10A), Limbaugh owes his fame "to a generic, free-floating policy of zero-tolerance, which [he] has regularly applied to almost *any person unlike him*" (my italics).

[9] In this analysis it is important to distinguish between beliefs and higher-order beliefs (that is, beliefs about the other player's beliefs, and so on). This is of course central to game-theoretic reasoning but, even in sociology, there is a tradition that emphasizes these distinctions [32, 50].

[10] For a related exercise where a small initial disturbance can cause large-scale strife, see Kuran [26].

[11] In Basu and Weibull [10] the view is taken that while culture is important it is not innate. What is taken to be cultural behavior can, at an elemental level, be based on an individual's optimal response to his environment. And to the extent that an individual's environment could be determined by the 'cultural' behavior of others, we can have cultural traps. But these can erode under external pressure, as for instance, a migrant community settles in a new country. This is nicely documented for migrants in the U.S. in terms of demographic behavior patterns by Leiberson and Waters [27] and financial market behavior by Osili and Paulson [36].

[12] This is closely related to the recent literature on 'acting White,' which tries to explain why for certain minorities adoption of behavior that can improve one's economic status often give rise to peer rejection [4].

[13] At times the infiltration is less than perfect. At one stage the KGB presence in the U.S. was so deep that it was advised that if you suspected anybody to be a KGB agent, you should immediately call the CIA headquarters. But, as a late-night talk-show host warned, "Do not be surprised if the same person answers your call."

[14] And when there is not that awareness it gives rise to droll instructions like the one Davis ([14], p. 19) reports, while traveling by train in Colombia: "I noticed a sign stuck to the back of the seat in front of me. It politely asked all passengers to be civilized enough to throw garbage out the window of the train."

[15] Of course, for the big difference in outcome to occur *everybody's behavior* changes. What is interesting and is being stressed here is that this *behavior* change of everybody can be brought about by the *preference* change of one or two persons.

References

1. Akerlof, G. and Kranton, R.: Economics and identity, *Quarterly Journal of Economics* **115** (2000), 715–753.
2. Akerlof, G. and Kranton, R.: A model of poverty and oppositional culture, In: K. Basu, P.B.

Nayak and R. Ray (eds.), *Markets and Governments*, Oxford University Press, New Delhi, 2003.

3. Andreoni, J. and Samuelson, L.: Building rational cooperation, (mimeo), Department of Economics, University of Wisconsin, 2003.

4. Austen-Smith, D. and Fryer, R.: The economics of "Acting White," NBER Working Paper 9904, 2003.

5. Baliga, S. and Sjostrom, T.: Arms races and negotiations, (mimeo), MEDS, Northwestern University, 2002.

6. Bardhan, P.K.: *Scarcity, Conflict and Cooperation*, The MIT, Cambridge, Massachusetts, 2005.

7. Basu, K.: A theory of association: Social status, prices and markets, *Oxford Economic Papers* **41** (1989), 653–671.

8. Basu, K.: *Prelude to Political Economy: A Study of the Social and Political Foundations of Economics*, Oxford University Press, Oxford, 2000.

9. Basu, K.: Gender and say: A model of household behavior with endogenously determined balance of power, *Economic Journal* (2005), forthcoming.

10. Basu, K. and Weibull, J.: Punctuality: A cultural trait as equilibrium, In: R. Arnott et al. (eds.), *Economics for an Imperfect World*, MIT, Cambridge, Massachusetts, 2003.

11. Bates, R.H., de Figueiredo, R.J.P. and Weingast, B.R.: The politics of interpretation: Rationality, culture, and transition, *Politics and Society* **26** (1998), 603–642.

12. Bowles, S.: *Microeconomics: Behavior, Institutions and Evolution,* Princeton University Press, Princeton, 2003.

13. Carlsson, H. and van Damme, E.: Global games and equilibrium selection, *Econometrica* **61** (1993), 989–1018.

14. Davis, W.: *One River: Explorations and Discoveries in the Amazon Forest*, Touchstone, New York, 1997.

15. Ellemers, N., Spears, R. and Doosje, B. (eds.): *Social Identity*, Blackwell, Oxford, 1999.

16. Fryer, R.G. and Jackson, M.O.: Categorical cognition: A psychological model of categories and identification in decision making, (mimeo), Harvard University and Caltech, 2003.

17. Fukuyama, F.: *Trust: The Social Virtues and the Creation of Prosperity*, Free, New York, 1996.

18. Goffman, E.: *The Presentation of Self in Everyday Life*, Double Day, New York, 1959.

19. Goffman, E.: *Stigma: Notes on the Management of Spoiled Identity*, Simon and Schuster, New York, 1963.

20. Harsanyi, J. and Selten, R.: *A General Theory of Equilibrium Selection in Games*, The MIT Press, Cambridge, Massachusetts, 1988.

21. Hedges, C.: *War is a Force that Gives Us Meaning*, Public Affairs, New York, 2002.

22. Hoff, K. and Pandey, P.: Why are social inequalities so durable? An experimental test of the effects of indian caste on performance, (mimeo), World Bank, Washington, 2003.

23. Horowitz, D.: *Ethnic Groups in Conflict*, University of California Press, Berkeley, 1985.

24. Hume, D.: *A Treatise on Human Nature,* [Reprinted by Penguin, London, 1969], 1739.

25. Kudaisya, M.M.: *The Life and Times of G. D. Birla*, Oxford University Press, New Delhi, 2003.

26. Kuran, T.: Sparks and prairie fires: A theory of unanticipated political revolution, *Public Choice* **61** (1989), 41–74.

27. Leiberson, S. and Waters, M.C.: *From Many Strands: Ethnic and Racial Groups in Contemporary America*, Russell Sage Foundation, New York, 1988.

28. Li, J., Dunning, C. and Malpass, R.: Cross-racial identification among European Americans: basketball fandom and the contact hypothesis, (mimeo), University of Texas, El Paso, 1998.

29. Loury, G.: *The Anatomy of Racial Inequality,* Harvard University Press, Cambridge, Massachusetts, 2002.

30. Meissner, C. and Brigham, J.: Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review, *Psychology, Public Policy, Law* **7** (2001), 3–35.

31. Miguel, E. and Gugerty, M.K.: Ethnic diversity, social sanctions and public goods in Kenya, (mimeo), University of California, Berkeley, 2004.

32. Moore, J.C. Jr.: Role enactment and self-identity, In: J. Berger and M. Zelditch Jr. (eds.), *Status Rewards and Influence: How Expectations Organize Behavior*, Jossey-Bass, San Francisco, 1985.

33. Morris, S. and Shin, H.: Unique equilibrium in a model of self-fulfilling currency attacks, *American Economic Review* **88** (1998), 587–597.

34. Nussbaum, M.: Patriotism and cosmopolitanism, In: M. Nussbaum and J. Cohen (eds.), *For Love of Country?*, Beacon, Boston, 1996.

35. Nussbaum, M. and Cohen, J. (eds.): *For Love of Country?*, Beacon, Boston, 1969.

36. Osili, U.O. and Paulson, A.: Institutional quality and financial market development: Evidence from international migrants in the U.S., (mimeo), Federal Reserve Bank of Chicago, 2004.

37. Platteau, J.-P.: *Institutions, Social Norms and Economic Development*, Harwood Academic, Amsterdam, 2000.

38. Platteau, J.-P. and Strzalecki, T.: Heterogeneous loyalties and collective action: Micro-evidence from Senegal, (mimeo), University of Namur, 2003.

39. Rao, V. and Walton, M.: Culture and public action: Relationality, equality of agency and development, In: V. Rao and M. Walton (eds.), *Culture and Public Action*, Stanford University Press, Palo Alto, 2004.

40. Rubinstein, A.: The electronic mail game: Strategic behavior under almost common knowledge, *American Economic Review* **79** (1989), 385–391.

41. Sen, A.: Goals, commitment and identity, *Journal of Law, Economics, & Organization* **1** (1975), 341–355.

42. Sen, A.: Gender and cooperative conflict, In: I. Tinker (ed.), *Persistent Inequalities*, Oxford University Press, New York, 1990.

43. Shepsle, K.A. and Bonchek, M.S.: *Analyzing Politics: Rationality, Behavior and Institutions*, Norton & Co, New York, 1997.

44. Sporer, S.: Recognizing faces of other ethnic groups: An integration of theories, *Psychology, Public Policy, Law* **7** (2001), 36–97.

45. Srinivas, M.N.: The social system of a mysore village, In: M. Marriott (ed.), *Village India*, University of Chicago Press, Chicago, 1955.

46. Swidler, A.: Culture in action: Symbols and strategies, *American Sociological Review* **51** (1986), 273–286.

47. Tajfel, H.: Social identity and intergroup behavior, *Social Science Information* **13** (1974), 65–93.

48. Tangeras, T. and Lagerlof, N.-P.: Ethnic diversity and civil war, Working Paper No. 589 IUI, Stockholm, 2003.

49. Taylor, C.: The politics of recognition, In: A. Gutmann (ed.), *Multiculturalism*, Princeton University Press, Princeton, 1994.

50. Troyer, L. and Younts, C.W.: Whose expectations matter? The relative power of first- and second-order expectations in determining social influence, *American Journal of Sociology* **103** (1997), 692–732.

51. Turner, J.C.: Social comparison and social identity: Some prospects for intergroup behavior, *European Journal of Social Psychology* **5** (1975), 5–34.

52. Varshney, A.: *Ethnic Conflict and Civic Life: Hindus and Muslims in India*, Yale University Press, New Haven, 2002.

53. Weibull, J.: *Evolutionary Game Theory*, The MIT, Cambridge, Massachusetts, 1995.

54. Weibull, J.: Testing game theory, (mimeo), Boston University, 2004.
55. Weingast, B.R.: Constructing trust: The political and economic roots of ethnic and regional conflict, In: V. Haufler, K. Soltan and E. Uslaner (eds.), *Where is the New Institutionalism Now*, Michigan University Press, Ann Arbor, 1997.
56. Wilson, W.J.: *The Truly Disadvantaged: The Inner City, the Underclass and Public Policy*, Chicago University Press, Chicago, 1987.
57. Young, P.H.: *Individual Strategy and Social Structure*, Princeton University Press, Princeton, 1998.