A theory of ethnic group boundaries

SUN-KI CHAI
Department of Sociology,
University of Arizona, Tucson, AZ 85721

ABSTRACT. Despite widespread agreement that ethnic boundaries are malleable rather than fixed, theories of ethnicity and collective action have been unable to adequately explain why individuals choose to mobilise collectively within particular boundaries rather than others. The boundaries of groups engaging in ethnic collective action are always taken for granted at some level rather than problematised. This leads to an undesirable reification of ethnic groups as actors.

The theory presented in this article integrates a social psychological view of motivation with a rational choice view of action to provide a systematic way of predicting the boundary location of ethnic groups that begin to mobilise in societies undergoing modernising structural change. It first focuses on the link between cooperation and altruism in small communities. It then predicts how altruistic preferences, in conjunction with structural factors and rational behaviour, will generate boundaries for larger-scale ethnic collective action that transcends yet incorporates such communities. The theory's predictions are then applied to explain the location of group boundaries in four very prominent cases of ethnicity 'creation' and collective action in this century.

The issue of ethnic boundaries is one that has occupied social scientists at least since Barth's (1969) seminal work. Since then, a long-standing debate has continued over the extent to which ethnic boundaries are 'primordial' (tied to ascriptive characteristics) or 'circumstantial' (determined by economic and political structure).

However, much of recent writing correctly recognises that both types of factors are significant: some sort of ascriptive commonality is after all necessary for a group to be ethnic in any meaningful sense, but the salience and level of inclusiveness of different ascriptive characteristics in determining ethnic boundaries varies according to differences in circumstance.

However, this growing consensus has not brought theorists any closer to understanding the precise manner in which primordial and circumstantial factors interact to determine the boundaries of ethnic groups. Hence systematic, predictive theories about the location of boundaries have not been forthcoming. One facet of this problem is that the boundary concept

The author would like to thank Ted R. Gurr, Dwight Hahn, Michael Hechter, Linda Molm, Susan Olzak, Robert Packenham, the late Aaron Wildavsky and various commenters at Nations and Nationalism for their suggestions on this article.
has not been defined in a way that permits ready operationalisation and would allow such theories to be tested against empirical evidence. In this article, I define ethnic boundaries as the ascriptive criteria that determine membership in particular groups organised for collective action. By defining ethnic boundaries in this manner, it is possible to infer from the historical record where ethnic boundaries were located at particular places and points in time.

This is not to say that alternative conceptualisations of ethnicity and hence boundaries are not possible. Ethnic collective action (also often referred to as ethnic mobilisation) can be distinguished from ethnic group awareness, which is simply the cognition that common ascriptive characteristics are shared among a set of individuals, without any salience necessarily being placed on these commonalities. It can also be distinguished from ethnic group solidarity, which is its broadest definition indicates that social resources of some type (affectual, cultural, organisational) are shared among a set of individuals with common ascriptive characteristics. Nonetheless, it can be argued that boundaries for collective action are the most significant ones from the point of view of social science, since these are the boundaries that have the most direct impact in the social, political and economic arena. An awareness of shared ascriptive characteristics is certainly a precondition to other forms of ethnicity, but can exist simultaneously over a vast number of possible boundary criteria and will be of little consequence unless it is acted on in some manner. Furthermore, existing definitions of solidarity place the boundaries of solidarity close to the boundaries of collective action, since levels of solidarity will determine a group’s potential for engaging in collective action (Hechter 1987: ch. 3).

In the first section of this article, I examine existing theories of ethnicity and collective action, arguing that all in one way or another take ethnic boundaries for granted rather than explaining them. I then present a new theory to account for the origin and location of the boundaries of large-scale ethnic groups in societies undergoing modernising structural change. The first part of the theory analyses altruism that results from pre-adult cooperative interactions in small, stable and relatively self-sufficient communities, using a model based upon a number of findings in social psychology, including dissonance theory, network exchange theory and the 'contact hypothesis'. The second part of the theory links the initial mobilisation of ethnic groups to the migration of individuals from such communities to larger metropoles and shows how altruism interacts with economic and political structural conditions to determine the boundaries for collective action among rational actors. The theory, while drawing insights from earlier analyses of ethnicity and collective action, will augment them in a way that allows it to determine the location of ethnic boundaries. These predictions will then be used to explain the location of boundaries in four of the most notable cases of ethnic group 'creation' and collective action that have occurred during this century: the

A theory of ethnic boundaries

Igbo in Nigeria, the Luba-Kasai in Zaire, the Malay and the Muhajir in Pakistan.

Existing theories of ethnic collective action

While the variety of social science concerning ethnicity is beyond enumeration, I will focus in this section on those theories that attempt to make general, systematic predictions about ethnic collective action. These theories can be placed into two categories: the first implicitly accepts the assumption that individuals act on the basis of self-interested preferences, i.e. that they care only about their own welfare, while the second assumes that individuals have altruistic or discriminatory preferences which predispose them to act on behalf of a particular ascriptive-defined group.

The first category includes the most prominent sociological theories of ethnic collective action, including the ‘cultural division of labour’ theory, the ‘competition’ theory and the ‘split labour markets’ theory. Each of these theories posits that ethnic collective action will occur when ethnic boundaries interact with a particular set of economic structural factors, although they disagree upon which structural factors are the most important.

Hechter’s cultural division of labour theory focuses on differences in general economic status (Hechter 1971) or in occupational specialisation (Hechter 1978) between ethnic groups as the main causes of ethnic collective action, while Bonacich’s split labour markets theory focuses on occupational overlap and differences in wages between ethnic groups (Bonacich 1972; 1979). Nagel and Olzak’s competition theory, on the other hand, focuses on modernisation and the breakdown of pre-existing economic and political barriers between ethnic groups (Nagel and Olzak 1982; Olzak and Nagel 1986). Banton’s theory of ethnic competition focuses on the monopolisation of particular economic factors by an ethnic group (Banton 1983: ch. 6), while Rogowski’s theory of ‘pillarisation’ focuses on economic self-sufficiency within an ethnic group (Rogowski and Wasserspring 1971; Rogowski 1974; 1985).

Hence, all these theories accept the notion of ethnic boundaries which exist prior to the effect of structural factors; structural factors in turn determine their salience for collective action. Hence, they do not attempt to analyse collective action when structural factors do not align with ethnicity. This raises the question of whether structural factors are the sole determinants of collective action, while ethnicity may simply be aligned coincidently with it but not have any independent causal force. Among the above authors, Banton does not address the issue, while Bonacich unequivocally places causal weight entirely upon structural economic factors (Bonacich 1979: 19-20). The latter position is somewhat problematic, given that it leaves unclear why collective action should ever occur.
under the banner of ethnicity rather than pure economic interest or why members of different ethnic groups who have common economic interests cannot unite. Bonacich does discuss the possibility of false consciousness, but this begs the question of where such consciousness comes from (Bonacich 1979: 34 5).

The other authors, however, accept a causal role for ethnicity, attributing it to various factors that can collectively be referred to as 'ethnic resources'. Olzak, for instance, focuses on the organisational factors such as 'inter-connected networks, institutions and information systems' built up over time through stable interactions (Olzak 1985: 80). Hechter, on the other hand, focuses on ease of communication and its effect on the cost of collective action, writing: 'cultural differences between individuals generally impose barriers to communication between them. Language is the most obvious of these cultural difference, but it is by no means the only one. The interpretation of meanings across boundaries is invariably problematic'.

Rogowski focuses on yet another factor, which he calls stigmata, i.e. human characteristics that are easily identifiable and changeable only at high cost. Stigmate, he asserts, make it more difficult for members of a faction to engage in non-cooperative behaviour without being detected, hence they facilitate the application of intragroup sanctions (Rogowski 1974: 71, 83; 6; Rogowski and Wasserspring 1971: 19–25).

All these different types of ethnic resources can plausibly play a role in collective action, but they have not been operationalised in a way that is readily measurable or been theorised in a way that allows one to predict their levels within particular groups and their precise effect on individual incentives. Because of this, it is difficult to use them to predict the boundaries of ethnic groups or to explain how ethnicity affects collective action in particular cases. This is reflected in the fact that, in their case studies, all these authors end up discussing existing ethnic groups (e.g. Native Americans, Scots, Tamils) as if their boundaries were static, rather than explaining where their boundaries came from.

The above theories do have unquestionable assets: each can be used to predict variations in levels of collective action along pre-existing ethnic boundaries within a broad range of socio-economic and political environments. Because of this, they have brought a new, higher level of deductive rigor to the social science literature on ethnicity. However, it is also true that they pay relatively little attention to problematising ethnicity itself.

It is conceivable that some of these problems could be resolved by a broader and more precise specification of the ethnic resource and more detailed reasoning about how it affects individual-level incentives, but there are inherent limits in such an approach. The ways in which culture, language, physical appearance, existing organisations and so forth can conceivably promote cooperation are so numerous and varied that any attempt to pack them into a single concept would render it unmeasurable, except tautologically.

Furthermore, no conception of ethnicity based on resources and economic self-interest can be consistent with the apparent motivations that drive individuals participating in many types of ethnic collective action. According to Epstein, the underlying issue is 'the powerful emotional charge that appears to surround or to underlie so much of the ethnic behaviour, and it is this affective dimension of the problem that seems to be lacking in so many recent attempts to handle it' (Epstein 1978: xi). As Bell notes in his often-quoted statement, ethnicity is a powerful determinant of collective action because it can 'combine interest with an affective tie' (Bell 1971: 169). Or, as Rothschild suggests, ethnicity's great appeal 'lies precisely in its ability to combine emotional sustenance with calculated strategy' (Rothschild 1981: 61). Finally, as Nagata puts it: 'Ethnic identity then is a unique blend of affective, expressive and basic ties, sentiments and loyalties with (sometimes blatantly) instrumental, calculated, political interests, and the latter are explained and given meaning by the former' (Nagata 1981: 112).

Such comments about ethnically linked motivations imply that individuals have preferences that go beyond the conventional assumptions of economic self-interest. Without taking such preferences into account, theories of ethnic collective action can never explain why individuals will take part in collective action despite opportunities for 'free riding' on the contributions of others (Hechter 1988a: 18), even when such actions involve high levels of personal risk (Connor 1987: 204; 1972: 342).

This brings us to another category of theory, found primarily in economics, which attempts to incorporate such preferences directly into the utility calculations of individuals, i.e. into the perceived costs and benefits that drive their behaviour. The earliest and perhaps best-known work in this field was done by Becker, who assumed a 'taste for discrimination' among individuals, which is manifested as a 'discrimination coefficient' that imposes costs to transactions with individuals of a disliked ascriptive group (Becker 1957). Theory-building along these lines has been continued by others such as Sowell, who posits that discrimination against undesired groups will be most pronounced where wealth-maximization incentives are low (Sowell 1975: 165–7); Klitgaard notes that the known presence in a population of individuals with discriminatory tastes may cause even individuals without such tastes to discriminate (Klitgaard 1976: 9–11); and Schelling, who shows how even slightly discriminatory preferences can lead to highly segregated patterns of residence (Schelling 1978: ch. 4).

Another line of theory-building in this category has focused on altruistic preferences, i.e. the ways in which individuals incorporate the welfare of particular others into their own utility calculations. This is arguably a more precise approach than discriminatory preferences, since it distinguishes between different varieties of interaction (i.e. cooperative vs. conflictual) with a particular group. The earliest work in this area was done by Johnson, who posited a 'taste for nationalism' among government elites linked to having members of one's own ethnic group hold jobs or own
property (Johnson 1965: 176; 1967). Similarly, Seers posited that elites act according to ‘the urge to promote the presumed interests of a group with culture coherence, probably showing at least a degree of linguistic and ethnic homogeneity’ (Seers 1983: 9). Work by Rabushka and Shepsle, while not explicitly talking about altruistic preferences, implicitly incorporated such assumptions by treating ethnic groups as monolithic entities with unified preferences (Rabushka and Shepsle 1972). Hechter himself has also discussed the significance of altruistic preferences for ethnic collective action in some of his recent work (Hechter 1988a; 1993). Beyond this work on ethnicity, there has been a growing literature in economics that attempts to model the effects of altruistic preferences on collective action in general (Phelps 1975; Collard 1978; Boulding 1981; Margolis 1982).

The common problem with these theories is that, while they model the possible effects of discriminatory or altruistic preferences on ethnic collective action, they cannot account for the preferences themselves. Because of this, each type of theory must take preferences as exogenously determined, and cannot predict location of boundaries for ethnic collective action.

Hence, neither theories of self-interested or of discriminatory/altruistic ethnic collective action can account for the boundaries of ethnic groups. The preceding discussion indicates that recognition of preferences that go beyond self-interest is an essential element of any theory that attempts to explain the effects of ethnicity on collective action. However, such preferences need to be accounted for rather than simply assumed, and their effects on the boundaries of ethnic groups need to be specified.

A theory of ethnic group collective action

The theory of ethnicity presented here will attempt to address these issues by accounting endogenously for altruistic preferences and showing how they interact with structural factors to determine the boundaries of ethnic groups engaging in collective action. It focuses in particular on societies that are beginning to experience modernising structural changes such as industrialisation, commercialisation and large-scale political consolidation. In such societies, the majority of the population will still reside in small, sub-ethnic communities, but many individuals from such communities will eventually migrate to larger population centres in pursuit of work and education. This ‘transitional period’ of structural change has been a major focus for analysis both by nineteenth-century social theorists (Tönnies 1957; pt. 4, sect. 4; Durkheim 1984: book II, ch. 2) and by theorists of the postwar ‘modernisation’ approach (Lerner 1958; Rigg 1964; Pye 1962; ch. 2; Deutsch 1961). Despite widespread debates over the specific predictions made by such theories, there is no questioning that the transition from small communities to larger population centres is one of the most significant effects of long-term structural change.

The formation of altruism in communities of origin

The first part of the theory will attempt to explain the nature of the altruistic preferences felt by individuals in such societies. I will assert that such preferences will generally be directed primarily at the members of each individual’s community of origin. This accords with the ideas of the ‘classical’ social theorists, who posited that strong affectual ties are a distinctive characteristic of relationships between individuals in such ‘traditional’ communities.

This assertion not only follows from classical social theory, however, but also from a number of findings in social psychology. By definition, patterns of cooperation for each individual in small, stable and relatively self-sufficient communities will revolve around a relatively small group of individuals. Community members will have to engage in common activities such as agriculture, hunting and foraging, as well as possibly defence, in order to ensure their mutual survival. Furthermore, given relatively low levels of technology and a relatively simple division of labour, we can expect that such cooperation will more likely be among individuals of equal or similar social status than would be the case in highly industrialised or commercialised areas. Obviously, as in most general social theories, the communities depicted here are ideal types and reality is somewhat more complicated. Nonetheless, as long as these characteristics are predominant ones in communities that have not yet experienced modernising structural changes, the implications that follow are still valid.

From these types of cooperative interactions, one can expect individuals in such communities to generally develop positive preferences towards other community members. Support for this inference can be derived indirectly from the ‘contact hypothesis’ in research on race relations, which links the reduction of prejudice between individuals to ‘equal-status contact . . . in the pursuit of common goals’ (Allport 1954: 281). Since its original promulgation, the hypothesis has been investigated empirically in a wide range of cases and applications (Sherif 1961; 1978). Though the hypothesis was developed originally to explain reductions in prejudice, its logic implies that a sufficient level of cooperative activities will lead to ‘negative prejudice’, i.e. altruistic preferences. This interpretation is supported by empirical research which shows that participation in cooperative activities can lead not only to the reduction of prejudice, but to actual friendship and mutual sympathy (Sherif 1978).

The link between cooperation and altruistic preferences can also be justified by findings in a more general social psychological paradigm, cognitive dissonance theory. Although there are many versions of dissonance theory in circulation, they are generally consistent with the proposition that individuals will adjust their preferences in order to minimise their clashes with their past actions. Since cooperative actions by definition collectively benefit a group of individuals, this would imply that
A theory of ethnic boundaries

dissonance-reduction for those participating in cooperative action would involve the acquisition of altruistic preferences towards the other members of the group. Altruistic preferences increase the subjective benefits attached to cooperative action, hence minimising any dissonance-creating doubts about whether the cooperation was worthwhile.

Moreover, the link between cooperation and altruism is consistent with another large body of psychological literature, reinforcement theory. One of the key components of a reinforcement theory is the assumption that stimuli that are associated with rewarding outcomes will themselves become the basis for action, even in the absence of the original reward. This in turn implies that action that benefits an individual, as long as it benefits the group as well, will reinforce the group’s benefit as a basis for further action, even in the absence of personal reward (Lawler and Yoon 1993). In sum, there are a number of strands of psychological thought that confirm the links between cooperation and altruism.

Of course, community interactions can also involve conflict as well as cooperation. However, it is also more than probable that an individual’s interactions with other community members will on aggregate have a positive net effect on the welfare of the other community members, i.e., cooperative interactions will predominate. If this were not true, then community members would be able to gain a higher level of welfare through solitary life than through community life, and would hence exit the community. Furthermore, the pressure of the environment in such communities is often high enough that any failure to cooperate could quite possibly lead to mutual starvation. Given all this, we can plausibly assume that individuals will generally have altruistic preferences towards the other members of their communities.

Such a Gemeinschaft altruism can be considered ‘primordial’, but the group of individuals towards whom the individual will feel altruistic will be much too small to be considered an ethnic group. Furthermore, while members of a community will share ascriptive characteristics such as race, religion, language, customs and so forth, individuals will not need at this point to demarcate the boundaries of their altruism based on these characteristics. Their sense of altruism will be attached to a concrete group of individuals, not to any abstractly defined ‘imagined community’. The creation of such a community will await migration to large population centres.

**The expansion of boundaries in population centres**

As noted, the focus of the theory will be on societies undergoing modernising structural changes such as industrialisation, commercialisation and large-scale political integration (with its accompanying bureaucratisation and expansion in education). The causes behind these structural changes can be multifaceted; the most obvious cause is endogenous technological advancement in production methods and transportation, but exogenous forces such as invasion and colonisation can also clearly play a major role.

These structural changes in turn generate large population centres for the conduct of production, administrative and education activities, as well as incentives for individuals to migrate from the home communities into these population centres. Once individuals migrate into such areas, they will face competition for scarce resources such as jobs, placement in educational institutions, residential space and government services, as posited by the competition theorists.

In such competition, a disproportionate share of the resources available will go to groups of individuals who organise effectively and act cooperatively. Hence individuals will have an incentive to join factions whose members act together in the battle for economic and political goods (Bates 1974: 460-1; Abernethy 1969: 106-7). These incentives will exist across a wide variety of arenas. Within businesses and bureaucracies, there will be incentives for individuals to form factions whose members mutually favour one another in hiring and promotion decisions. Where political institutionalisation is strong, there will also be incentives for individuals to form lobbying groups or political parties that work to pursue shared interests. Where institutionalisation is weak, organising groups for collective violence may be another way to pursue such interests.

In the process of forming factions, the decisions faced by individuals are two-fold: the first is what sort of group to join and the second is whether to act cooperatively or to ‘free-ride’ on the cooperation of other members once one has joined such a group. In such situations, one can expect rational individuals to join the type of group that provides them with the highest level of utility, i.e., the maximal fulfillment of their preferences. This utility in turn will be related to the size of each potential group, the costs of organising it, and the amount of resources that will have to be expended (for monitoring and sanctioning) to ensure that individuals within it act cooperatively.

It can then be posited that the total utility gained from membership within a group will be an increasing function of the size of the group, up to some ideal point that is somewhere about half the size of the entire population. This can be justified by pointing to the fact that a group with only a few members will be overwhelmed by larger groups, while a group that is sufficiently large to defeat all other groups will dominate access to resources and have no incentive to include more members, who will simply increase the number of individuals among whom such resources must be distributed. Clearly, the exact optimal size of such groups will depend on the capabilities of group members. In any reasonably large population centre, however, the optimal size will be considerably larger than the number of migrants from any single community of origin.
It can also be posited that a group will not be able to act effectively unless group members share interests in particular political or economic outcomes that can be gained through collective action. In the absence of such a shared interest, actions which help some members of a prospective group will hurt other members, and there will obviously be little incentive for members who will be hurt to cooperate in collective action. The most obvious way in which a group of individuals can have a shared interest in a particular outcome is through a common position in the division of labour. Such a common position will give all group members an interest in either raising the economic status of the occupational roles that they commonly hold or in opening up opportunities for entry into other, more prestigious occupations. This implies, in line with the cultural division of labour theory, that a common position in the division of labour will be a strong factor in determining the boundaries of collective action.

Even where shared interests exist, however, collective action will not occur if group members attempt to 'free-ride' on the cooperative behaviour of other members rather than cooperate themselves. In order to prevent this, private benefits will generally have to be provided to group members who cooperate in collective action and punishments to those who do not. However, such a system of 'selective incentives' will require the expenditure of resources for monitoring members and for administering rewards and sanctions. It is reasonable to say that the amount of resources required for this will be positively related to the amount of selective incentives that must be provided to make members act cooperatively. This suggests that factors that raise the utility of cooperation for group members prior to the application of selective incentives will increase the 'efficiency' of a group as a collective enterprise.

It is at this point that altruism comes into play. One factor that can clearly lower the cost of cooperation within a group is altruism among a subset of group members. For any individual, the marginal utility of cooperation vis-à-vis non-cooperation, prior to selective incentives, will be the intrinsic cost of cooperative action minus one's own benefit from the increase in group welfare that occurs. Likewise, the total utility of cooperation as opposed to non-membership will simply be a function of one's own benefit from total group welfare.

What this indicates is that a higher aggregate altruism towards group members (which is proportionate to the number of community members in the group) will increase the total utility of cooperation in two ways: it will directly increase the utility each member receives from the benefits obtained by others in the group and it will reduce the resources required to ensure cooperation by group members. These two effects will be mutually reinforcing (a simple formal proof of this can be found in the appendix). This in turn implies that the total utility for each individual for group membership will be an increasing function of the number of other members who share the same community of origin. This in turn implies that, aeternis puribus, it will be more efficient to organise groups around criteria that envelop communities of origin rather than those that cross-cut them.

It can also be posited that the costs of organising a group will be prohibitively high unless there is a relatively simple set of criteria that can be used to define the group's membership. This can be justified by noting that any large-scale group will have to organise individuals who basically have no acquaintance with one another. Because of this, there must be some criteria by which individuals can distinguish members from non-members. Furthermore, these criteria must be simple and clear-cut enough so that individuals can estimate the prospective size and membership of such a group in order to rationally make their own decision on whether to join. The only criteria that fill such requirements, lead to groups of sufficient size and encompass entire communities of origin are ascriptive ones such as race, religion, language, customs or region of origin. This, rather than any inherent emotional drawing power of any of these types of factors, can be used to account for the initial mobilisation of large-scale ethnic groups based upon ascriptive criteria: given existing patterns of altruism towards communities of origin, they are efficient ways of organising individuals for collective action.

The implications of this discussion can be summarised as follows:

Hypothesis 1: Large-scale ethnic groups will first form in societies where the majority of the population still resides in small, stable and relatively self-sufficient communities, but where modernising structural change has proceeded to the point where many individuals will migrate from these communities to larger population centres in their lifetime.

Hypothesis 2: Where such groups are formed, their boundaries will be based upon a relatively simple set of criteria that meets the following conditions:

2.1. It generates a group of sufficient size to comprise a substantial portion of the population within the population centre, but not significantly more than half the entire population.

2.2. It aligns with a common position in the population centre's division of labour.

2.3. It encompasses communities of origin rather than cross-cutting them.

Where no set of criteria meet these conditions, we should not expect large-scale ethnic groups to form, even in the presence of modernising structural change. Where more than one boundary meets the conditions, we can expect instability in the boundaries of groups as competing political and social entrepreneurs attempt to organise groups along alternative 'cross-cutting' ethnic cleavages. What this indicates, as a number of authors have noted, is that the creation of ethnic groups may be the product of entrepreneurs who are attempting to 'build' groups (Ranger 1983: 252-4; Vail 1991: 11), but it also indicates that in order to be successful such
entrepreneurs must shape their ideas around the existing raw material of altruism towards communities of origin and shared interests caused by a common position in the division of labour.

Once ethnic groups are formed in the population centre, group members may go back to their communities of origin, attempting to enlist their non-migrant home community members in their groups. The success of such efforts will be greatly promoted by the existence of a democratic system of government, which provides strong possibilities for rural residents to organise in ways that will be effective at the national level. It will also depend on whether newly organised ethnic groups comprise a sufficient proportion of the population for effective action at the national level as well as in the population centres where they originally arose.

The preceding discussion has focused on the initial mobilisation of large-scale ethnic groups for collective action. It does not assert that members of such groups will, at that point, feel any altruism towards the ethnic group as a whole; altruism will still be directed primarily at their communities of origin. Once ethnic groups become established, however, the logic of the argument implies that a broader ethnic altruism will begin to develop. Once an individual has begun acting cooperating with other members of the larger ethnic group in pursuit of shared objectives, the link between cooperation and altruism will generate altruistic preferences towards the other ethnic group members with whom the individual is cooperating. Among children of migrants, a new generation may grow up outside of local communities and in a milieu where cooperative interaction takes place primarily within the boundaries of large-scale ethnic groups; hence such ethnicity will become such an individual’s initial focus for altruistic preferences.

Once patterns of altruism are transferred, moreover, they will promote further cooperation within the ethnic group by increasing the marginal utility from cooperative behaviour and thereby decreasing the resources required to ensure cooperation, hence improving the group’s efficiency as an organisational unit. This will in turn hamper cooperation along alternative boundaries, since this would place group members who are altruistic towards one another in conflictual situations. Those outside the group will be discriminated against by those within, hence providing them with further incentives to form their own ethnic groups for collective action. Hence, the boundaries of ethnic groups will to a certain extent be ‘cemented’ into place by the development of ethnic altruism. This implies that once large-scale ethnic groups have become the basis for collective action, dramatic changes in their boundaries will be rare, though incremental changes may occur as a result of further structural transformations. Changes in the types and levels of shared circumstantial characteristics within these groups will hence be likely to be manifested in rising and falling levels of collective action within marginally shifting boundaries rather than by major changes in the location of the boundaries.

**Cases of ethnicity creation in the twentieth century**

The hypotheses outlined in the previous section can be applied to historical cases in order to compare the usefulness their predictions about the timing and boundaries of ethnic group collective action against actual outcomes. The cases examined below comprise perhaps the four most notable examples of ‘ethnicity creation’ that have occurred during this century. Each has been prominently featured in the literature on race and ethnicity, and they do not in any way constitute unusual or special cases in the formation of ethnicity. Although four cases cannot constitute a definitive test of the theory, the prominence of these cases makes them an important indicator of its usefulness.

In each case, a group of individuals began to engage in significant collective action along large-scale ascriptively defined boundaries in the presence of economic and political structural change, where no previous record of such collective action had existed. In the process, the boundaries of collective action became the basis for institutionalised ethnic groups that have remained prominent since their formation.

**Pan-Igbo ethnicity in Nigeria**

Prior to colonialism, a sense of common identity was absent among the Igbo-speaking peoples of what became Eastern Nigeria. Political organisation usually did not extend beyond the village level, and what larger subjective distinctions existed were between Igbo subgroups such as the Nri and Aro, or between mainlanders (who were simply called Igbo) and the riverine Otu (Nnoli 1989: 22; Olorunsola 1972: 7-9).

As Hypothesis 1 predicts, large-scale ethnicity became a salient political factor only after colonialisation generated political consolidation and rapid migration into large urban areas such as Lagos and Ibadan in the western region of Nigeria; Kano, Zaria, Kaduna and Jos in the north; and Port Harcourt, Onitsha, Aba and Enugu in the east. Each of these cities experienced rapid increases in population during the colonial and post-colonial era, and the ethnic composition of migrant populations has not been homogeneous. Because of generally poorer agricultural conditions and inheritance customs based upon primogeniture, easterners (predominantly Igbo-speakers) exceeded westerners and northerners, among natives of the three main regions, in the number of individuals who migrated out of their major home regions in search of jobs.

Hypothesis 2.1 posits that the criteria for the mobilisation of large-scale ethnic groups for collective action in a metropolis would need to be relatively simple, yet encompass a significant proportion of the population. For the cities listed above, the most obvious of such criteria were linguistic descent (Igbo, Yoruba, Hausa), religion (Muslim, Christian, ‘animist’), or major region. On a national scale, each of these criteria divide the population in
such a way that each group formed is significant in size but not dominant. However, Hypothesis 2.3 would rule out religion-based boundaries, because they tended to exercise communities of origin. Igbo speakers, though largely Christian, were often divided within villages between Catholics and Protestants. Likewise, Yoruba-speaking villages were similarly divided between Christians and Moslems (Laitin 1985: 289; 1986).

Secondly, linguistic boundaries aligned with positions in the division of labour, a factor that is emphasised by Hypothesis 2.2. Though they initially fell behind other groups, Igbo-speakers were after the 1930s more prone than others to take advantage of educational and entrepreneurial opportunities, and neighbouring villages were caught up in very intense competitions to ‘get up’ economically and politically (Abernethy 1969: 68-9). This, along with their propensity to migrate, often brought Igbo speakers into communities where they were not only a minority, but were viewed as relatively privileged and upwardly mobile compared to non-Igbo speakers. On the other hand, purely regional boundaries would have diluted this alignment, since it would have grouped the Igbo with less upwardly mobile eastern groups such as the Ibibio.

As Hypothesis 2.1 predicts, pan-Igbo groups first became significant within the major urban areas outside their home eastern region in which Igbo-speaking migrants formed a significant but minority portion of the population (Nnoli 1989: 25-8). There, pan-Igbo political collectivity was common, usually in the form of cultural/political union activity (Abernethy 1969: 110-11), often in opposition to Yoruba collective action in the west and Hausa-Fulani collective action in the north. In fact, urban conflicts between Igbo groups on the one hand and Yoruba or Hausa groups on the other have been largely responsible for major outbreaks of ethnic violence in Nigeria during the colonial and post-colonial era.

Eventually, this type of urban pan-ethnic collective action was reflected in the formation of political parties and the struggles between them, and gradually came to assume a national significance. As noted earlier, this was enabled by the fact that national population was split up so that Igbo, Yoruba and Hausa formed a significant but not dominant proportion. In the 1959 election results just prior to independence, the National Council of Nigeria and the Cameroons captured every seat in Igbo majority areas, while the Action Group was nearly as dominant in Yoruba majority areas. In the north, the Hausa-Fulani-based Northern People’s Congress won the over three-quarters of the seats available (Post 1963: 358–75). By the 1960s, the strength of pan-Igbo collective action had spread to the extent that the Biafran separatist war was fought in the name of Igbo self-determination and was largely supported by the Igbo population, while it was opposed by the non-Igbo-speaking residents of the seceded eastern region (Young 1983: 204–11).

Although this section has focused on the mobilisation of pan-Igbo ethnicity, surveys of evidence show that the theory can provide insights into group collective action among the other two main ethnic groups of Nigeria. The Yoruba were the most urbanised group in West Africa prior to colonisation and, unlike the Igbo, shared a sense of common origin prior to the colonial period, tracing their ancestry back to the mythical Oduduwa. However, they did not consider themselves a single ethnic group, and in fact did not have any common term with which they used to refer to themselves. The term ‘Yoruba’ originally referred only to the residents of the Kingdom of Oyo, and gradually came to be adopted more inclusively only after colonisation and greater contact with groups not sharing the common descent myth and language (Nnoli 1989: 23). This is in accordance with Hypothesis 2.1, since such contact would create the need to band into larger groups than those based upon village or ancestral town.

Prior to colonisation, Hausa-Fulani identity likewise was directed primarily at the existing large emirates rather than at a common community. As Hypothesis 2.1 would predict, the mobilisation of a pan-Hausa ethnic group was originally triggered among migrants to southern cities, where they were only able to form a viable political force when unified (Young 1976: 278). Although religion played a stronger role as a unifying force in the north than it did elsewhere (incorporating subgroups that were not in the Hausa language family), a sense of Hausa ethnic identity is still stronger in Nigeria, where Hausa comprise a large but minority portion of the population, than in Niger, where they comprise a dominant majority (Miles and Rochefort 1991).

The Luba-Kasai of Zaire

‘Luba-Kasai’ refers to a group of people originating from Kasai who emigrated to various other areas of the Belgian Congo in the early twentieth century. Luba refers to their common language family, although it encompasses a variety of dialects. The Luba-Kasai are often referred to simply as ‘Luba’, although there are separate ethnic groups, such as the Lulua (sometimes spelled Lulua) and the Luba-Shaba, who also speak Luba dialects.

The mobilisation of the Luba-Kasai as a politically salient ethnic group is regarded as perhaps the most extreme case of ‘un-natural’ ethnic collective action in Africa, since it was not based on a single clear ascriptive criterion, but rather on a combination of linguistic and geographical criteria. Furthermore, prior to the period of migration, there was little awareness of any clear physical or cultural differences between Luba-Kasai and Lulua, much less any reason to make the boundary between the two the basis for political collective action (Turner 1972: 217-8; Roosens 1989: 117-8).

The Luba-Kasai identity began to form during the period of migration, when large numbers of Luba-speakers from the southern regions of Kasai migrated into different areas of Kasai as well as into Shaba. As migrants with little claim to land, they had greater need to learn new skills in order to
A theory of ethnic boundaries

Katanga, otherwise known as Conakat, led by Moise Tshombe, which sought to combat Luba-Kasai domination of the labour force and of local politics (Young 1965: 490–6)

As ethnic tension polarised, this conflict exploded into a number of skirmishes. ‘Authentic Shaban’ nationalism eventually led to the Katangan secession crisis of the early 1960s (Young 1983: 201–4), during which Luba-Kasai had to be temporarily evacuated from Lumumbashi by the UN in order to protect them from ethnic-based attacks (Young 1976: 181). Back in Kasai, the increasing saliency of Luba-Kasai ethnicity was reflected in the attempt by Albert Kalonji to take advantage of the Katanga secession in order to declare an independent Luba-Kasai republic in Southern Kasai (Young 1976: 188; Roosens 1989: 121).

Pan-Malay ethnicity in Malaysia

Prior to the onset of colonialism and urbanisation, the residents of the Malay peninsula could not be said to have held a high degree of ethnic identification. While for the most part they shared a common religion (Islam), loyalties remained directed at a parochial level (Nagata 1981: 97–100; Horowitz 1971: 127)

As predicted by Hypothesis 1, a politicised pan-Malay ethnicity only began stirring with British colonialisation and the beginnings of migration into the Straits Settlements, the urban centres of colonial rule. Within the settlements (Penang, Malacca and Singapore), natives of the peninsula mixed with Indians, Javanese and other Dutch East Indies natives, Thais, Burmese, Arabs and especially with large numbers of Chinese who had been imported as a labour force by the British (Ongkili 1985: 5–8). Indeed, the presence of the immigrants caused the indigenous elites to worry that they would be overwhelmed and dominated. This subsequently led to a great expansion in political activism among them (Roff 1967: 207–10; Nagata 1979: 47–8; Young 1976: 121–5). One of the preconditions for this activism was a clear definition of the boundary of their ethnic group, and this triggered considerable debate among intellectuals about the criteria for ‘Malayness’ (Roff 1967: 242–5).

Hypotheses 2.1 and 2.3 predict that the chosen boundaries of ethnicity would form a significant but not dominant portion of the population and would encompass communities of origin. Given this, the main criteria that could be used to construct such boundaries were colour, religion and territory/country of origin. Furthermore, Hypothesis 2.2 predicts that the chosen boundaries would align with a clear boundary in the division of labour. In the Settlements, the most prominent feature of the division of labour was the relative economic pre-eminence of the Chinese, who had disproportionately taken advantage of education and entrepreneurial opportunities, gradually taking a place of ascendancy in the urban division.
of labour. Hence, the hypotheses would predict that the boundaries of Malayness would be drawn in a way that would isolate the Chinese.

Furthermore, the Chinese formed nearly half of the population in each of the Straits Settlements. Because of this, Hypothesis 2.1 would predict that the chosen boundary of Malayness would be inclusive, bringing together the majority of the non-Chinese groups. Religion was the simple characteristic which aligned most clearly with the Chinese/non-Chinese split, and this indeed became the basis for the Malay group inclusion in the Settlements, encompassing not only Muslims with roots in the Malay group inclusion in the Settlements, encompassing not only Muslims with roots in the Malay peninsula but also Muslims of Javanese descent (Jawi peranakan) as well as Indians and Arab Muslims (Nagata 1981: 105–7). The one exception to this was the exclusion of the relatively small number of Chinese Muslims, despite their willingness to adopt the Malay language and customs. As Hypothesis 2.2 predicts: ‘(the Chinese) economic dominance in the country at large . . . alienates them from the rest of the Muslim community’ (Nagata 1981: 107), and inclusion of the Chinese Muslims as Malays would have diluted the shared interests caused by the Malay’s common position in the division of labour.

On the other hand, the boundary of a pan-Malay ethnicity has not been uniformly inclusive in all parts of the peninsula. In Kelantan, where Chinese form a very small portion of the population, the definition of Malay has been focused more on territory of origin and less inclusive, clearly excluding Muslims of Arab Descent (Nagata 1981: 108). This follows from Hypothesis 2.1, since more inclusive boundaries in Kelantan would form a coalition that was much larger than that needed for political domination, hence diluting the per-individual benefits that could be gained from such domination.

In the post-independence state of Malaysia, Kelantan is an exception with regard to its proportion of Chinese. Hence one would expect that nation-wide boundary definitions would follow along the broader lines of the Straits Settlements rather than the narrower ones of Kelantan. This has indeed been the case, and the administrative category of bumiputra, which is the basis for the affirmative-action programmes that have been in place since the early 1970s, includes not only non-Chinese Muslims, but also non-Muslim peripheral ‘tribes’ (orang asli) (Tan 1982: 39–40, 50). This is generally thought of as a way in which the government seeks to maximize combined pan-Malay strength (Provencher 1987: 107–9).

At any rate, the strength of pan-Malay solidarity has been shown by the domination of the Malay vote by the United Malays National Organization (UMNO) in the post-independence era. Even the only exception to this domination, the Parti Islam (PI), supports the previous analysis because the only state where it has been able to gain power over local government has been Kelantan, where Malays, broadly defined, form an overwhelming majority of the population. Patterns of electoral politics,

like those of ethnic definition, follow along the lines predicted by the hypotheses.27

The Muhajir of Pakistan

‘Muhajir’ (meaning refugee in Urdu) is a generic term to refer to the Muslims who fled from India to Pakistan during the tumult following the partition of the two countries in the aftermath of British rule, or from Bangladesh following its separation from Pakistan. As indicated by the recent genesis of their name, there was no reason for Muhajirs-to-be to feel any sense of common identity or to engage in collective political action against other groups during the era of the British Raj. Instead, they saw themselves primarily in terms of their Muslim identity on the one hand, or their local regional identities on the other. Although the vast majority were Urdu-speaking, they included Gujarati-speakers as well (Kennedy 1991: 939).

Moreover, while most came from the Hindi heartland of Uttar Pradesh, others originated from other areas and had Punjabi or Pathan names (Ahmed 1990: 32).

In post-independence Pakistan, however, the Muhajirs have become known as a coherent and politically active ethnic group, despite the fact that their very existence dates back only to the Partition. This is reflected both in the massive success of the Muhajir Qaumi Movement (MQM) in sweeping Muhajir votes in the large urban areas (Karachi and Hyderabad) where they reside in significant numbers (Ahmed 1990: 32–3; McDonald 1988). It is also reflected in the massive ethnic violence that has been perpetrated in those cities in support of the Muhajir cause, often in response to anti-Muhajir attacks by Pathan refugees from Afghanistan and by native Sindhis (Kennedy 1991: 948–54; Phillips 1988a, 1988b).

According to Hypothesis 1, the rise of large-scale ethnic groups should coincide with large-scale migration into urban metropolises. This is indeed what has occurred during the post-independence period. For example, the population of Karachi, Pakistan’s largest city, has swelled from 400,000 to over 7,000,000 in the past four decades. This has coincided with the rise of Muhajir ethnic solidarity, as well as efforts to suppress the Muhajirs on the part of other ethnic groups.

However, urbanisation alone does not explain why ethnic divisions in Pakistan took place primarily along lines that divided refugee from non-refugee rather than along religious lines (between Sunni and Shia), since both lines meet Hypothesis 2.3’s requirement for encompassing communities of origin. The other hypotheses provide two explanations for the initial boundary. First, Hypothesis 2.1’s requirement for sufficient size is more closely met by ethnic divide based refugee status than one based upon religion. Although Shi’ite Muslims make up about 15 per cent of Pakistan’s total population (Qureshi 1989: 109), they comprise a smaller percentage in the major cities of Sindh, where the Muhajir ethnicity is the strongest. The
two other major groups in those cities, Sindhi and Pathan (the latter comprised in large part of refugees from Afghanistan), are both entirely Sunni, while the Muhajir are split between a Sunni majority and a Shia minority (Harrison 1986: 269). On the other hand, the refugee category comprises a bare majority in both Karachi and Hyderabad (Kennedy 1991: 940).

Furthermore, as Hypothesis 2.2 would imply, a strong Muhajir identity has been encouraged by the fact that this boundary aligns with a particular position in the division of labour. Muhajirs on the whole have been more entrepreneurial and economically successful than other elements of the Pakistani population (Wright 1974). In urban areas, they have replaced the largely Hindu Sindhi middle class, which fled to India after the partition (Harrison 1991: 282). Furthermore, with Jinnah and Liaqat Ali Khan as their most notable representatives, they made up a disproportionately large share of the early Muslim League elite in the immediate post-independence government (Wright 1991). This was particularly resented by the Sindhis, who perceived Muhajir domination as the primary cause for the administrative severing of Karachi from the rest of Sindh (Harrison 1991: 280–1).

This combination of factors caused the Muhajirs to stand out from the rest of Pakistani society, and made them the focal point of resentments over the political and economic travails that have plagued post-independence Pakistan. This has led to changes in policies over time that have favoured Sindhis and Punjabis at the expense of Muhajirs (Wright 1991), which in turn has caused the Muhajirs to huddle together and respond in kind, a response exemplified by the popularity of the MQM.

Appendix

The relationship between altruism towards others and utility gained within a group.

We can model the utility function of an individual as \( u_t = u_i + \omega u_p \), where \( u_i \) represents the individual’s welfare, \( u_p \) represents the aggregate welfare of the other members of the individual’s community of origin and \( \omega \) represents the weight of altruism towards the community, with \( \omega > 0 \).

The marginal utility of cooperation for an individual within a large group as compared to free-riding will be \( u_m = b/n - c + \omega bk/n = (\omega k + 1) b/n - c \) where \( c > 0 \) is the exogenous cost of cooperation, \( b > c \) is the increase in the aggregate group welfare that results from an individual’s cooperation, \( n \) is the total size of the group and \( k \) is the number of members of one’s community of origin included within the group (not including the individual); \( b/n \) is the net utility an individual gains from personal benefit due to cooperative action, while \( \omega bk/n \) is gained from the benefit to fellow members of an individual’s community within the group. If \( u_m \) is less than zero, then the group must provide selective incentives exceeding the negative of this amount in order for cooperation to occur.

Given provision of sufficient resources \( r \) per individual to enforce cooperation through selective incentives, the total utility for each individual from participation in such a group if all members cooperate will be

\[
 u_i = (\omega k + 1)(b - r) - c 
\]

By definition, \( \omega k + 1 > 0 \). Furthermore, \( \partial(\omega k + 1)/\partial k = \omega > 0 \). Since \( r \) is a monotonically increasing function of \( -u_m \), the selective incentives that must be provided, \( \partial r/\partial u_m < 0 \). Because \( \partial u_m/\partial k = \omega b/n > 0 \), \( \partial r/\partial k < 0 \). This implies that \( \partial(b - r)/\partial k > 0 \), and that the effects of \( k \) on the two factors are mutually reinforcing.

Notes

1 For an overview of this debate, see Scott 1990, Stack 1986a, McKay 1982.
3 Thanks to Susan Oltzak for pointing out the need to make distinctions among different facets of ethnicity.
4 This does not imply, of course, that a group with high solidarity will always engage in
collective action, but rather that collective action, if it does occur, will tend to do so within high solidarity group boundaries.

5. Preference is the usual term used in rational choice theory to refer to the desires and goals that motivate behavior. This and other terms from rational choice theory are used here in order to maintain consistency with existing analysis along those lines. This does not imply acceptance of all the conventional assumptions of rational choice theories; the incorporation of altruism into the model is a movement away from such assumptions.

6. For mutual criticisms along these lines, see Olzak 1983: 360 and Hechter 1988b: 108, as well as Connors 1984.


8. The paradigmatic explication can be found in Tonnis 1957: 42-4.

9. Originally proposed by Festinger 1957.

10. Recent definitions of dissonance emphasize behavior that threatens self-image, variance of behavior from internalized norms, and/or the perception of personal responsibility for undesirable outcomes. See Cooper and Fazio 1984.

11. Conversely, dissonance-reduction in cases of conflict would involve the acquisition of negative altruistic preferences. See Davis and Jones 1960.


13. This also, indirectly, became the basis for one attempt by a prominent sociologist to devise a theory of value formation. See Homans 1961: 53-5.

14. To use Anderson's term; see Anderson 1983. The distinction also corresponds closely to that made by Patterson between 'existential' and 'ethnocentric' solidarity. See Patterson 1977: 43-4.

15. Though such centres are generally urban, they can also include non-urban settings such as plantations. See Horowitz 1971: 127-31.


17. This reasoning is analogous to Riker's theories about minimum winning coalitions. This points to the fact that, even within less structured contexts than legislatures, maximum returns for members in collective action will be achieved by groups that are just strong enough to prevail over other groups. The key is the all-or-nothing nature of the prize and its reach. See Riker 1962.

18. Debates on the selective incentives and the problems of collective action have arisen in large part from Olson 1965.

19. For a related discussion, see Becker 1983.


21. One exception to this will place in the case of 'remigration' of members of an established ethnic group to a new metropole. In such situations, even the ethnic group may be too small to be an effective political and economic faction, and a 'super-ethnic' group may form. The most notable case of such a group is the local 'identity' in Hawaii, which encompasses Japanese, Filipinos, Part-Hawaiians, Chinese, Koreans, Portuguese, and Samoans but excludes non-Portuguese Caucasians.

22. For a discussion of migration patterns, see Green 1974. For a discussion of the reasons for Igbo migration, see Nzirim 1963: 30-4.

23. It should be made clear that the eventual boundaries of the Igbo group were based upon descent from traditional Igbo language speakers rather than actual linguistic capability, since many second-generation Igbo in the West and North could not speak the language fluently.

24. For example, a survey in the early 1970s showed that a sample Igbo community of 1878 people consisted of 52.3 Roman Catholics and 46.5 Protestants. See Iro 1981: 185.

25. The causes of migration were linked to the vulnerability of their region to slave trade, as well as to the better opportunities available in other regions. See Young 1976: 176.

26. For a large-category breakdown over the years, see Ratnam 1965: 2.

27. The formation of the UMNO splinter group, Semangat '86, is an exception to this trend. However, its political doctrines are practically identical to that of UMNO, suggesting that its formation reflects a failure of UMNO internal organisation rather than any fragmentation of Pan-Malay solidarity. Furthermore, it has been unable to challenge the dominance of UMNO among Malays, and in this turn seems to be leading to a slow reconciliation process between the two parties.


References


A theory of ethnic boundaries


Lerner, Daniel. 1938. 'The Passing of Traditional Society: Modernizing the Middle East'. Glencoe, Ill.: Free Press.


A theory of ethnic boundaries

for, showing how the presence of the specific ethnicity translates into a specific form of social organization. This process is called ethnic identity formation. 

In this way, ethnic identity is seen as a social construct that is shaped by different factors such as history, culture, religion, and politics. These factors interact with each other to create and sustain ethnic identities. 

The concept of ethnic identity has been influential in the study of social inequalities and social exclusion. It helps explain how certain groups are marginalized and excluded from society based on their ethnicity.

For example, in the study of immigration and integration, ethnic identity plays a crucial role in determining an individual's ability to integrate into the host society. Those with a strong ethnic identity may find it challenging to assimilate into a new culture, while those with a weaker ethnic identity may find it easier to integrate.

In conclusion, ethnic identity is a complex and multifaceted concept that is shaped by various factors. Understanding the role of ethnic identity in social organization is crucial for addressing social inequalities and promoting social cohesion.