Predicting Ethnic Boundaries

Sun-Ki Chai

It is increasingly accepted within the social sciences that ethnic boundaries are not fixed, but contingent and socially constructed. As a result, predicting the location of ethnic boundaries across time and space has become a crucial but unresolved issue, with some scholars arguing that the study of ethnic boundaries and predictive social science are fundamentally incompatible. This paper attempts to show otherwise, presenting perhaps the first general, predictive theory of ethnic boundary formation, one that combines a coherence-based model of identity with a rational choice model of action. It then tests the theory's predictions, focusing in particular on the size of populations generated by alternative boundary criteria. Analysis is performed using multiple datasets containing information about ethnic groups around the world, as well as the countries in which they reside.

Perhaps the best-known innovation in contemporary theories of ethnicity is the idea that ethnic boundaries are not predetermined by biology or custom, but malleable and responsive to changes in the surrounding social environment. Ethnicity is seen to be socially constructed, and the nature of this construction is seen to vary from situation to situation.[1] This in turn has linked to a more general concern with the contruction of social boundaries in general, including class, gender, and profession.[2] Nonetheless, little work has been done to generate a positive theory of ethnic boundary formation, one that can predict what sorts of ethnic identities will be formed under different circumstances, and how these identities will affect collective action. This paper presents such a theory, then tests parts of it against multiple cross-national ethnic datasets.

Existing Literature

Part of the problem in devising a positive theory of boundaries has been the difficulty in coming up with a consensually-held definition of 'ethnicity.' However, the recent tendency in international social science, one that is adopted here, has been to define ethnicity broadly as encompassing all ascriptively-based group boundaries, including those based upon race, religion, language, and/or region. Characteristics such as religion, language, and region are viewed as ascriptive rather than cultural because they are typically defined for purposes of ethnicity to refer not to the current practice or location of individuals, but rather to their lineage, the customs of their ancestors. Hence, while individuals may adopt new languages, lapse or change religions, or alter their places of residence, it is much more difficult for them to change their ethnicity, even when ethnicity is based upon such characteristics.

Some scholars argue that viewing race as a kind of ethnicity downplays the extent of racial discrimination in U.S. society by overgeneralizing from white immigrant experiences. However, while this may reflect the past use of these concepts in the U.S. context, there is nothing in defining race as a type of ethnicity that forces social scientists to treat all bases for ethnicity as having equal salience. Indeed, the focus in this article will be on showing how certain ascriptive characteristics tend to be more salient than others according to each set of circumstances. Overall, a more encompassing definition of ethnicity has been gaining ground due to the recognition that all ascriptive categories, including

race, are defined in a way that is socially rather than biologically determined.[3]

Sociological theories of ethnicity have customarily been divided between two approaches, the 'circumstantial' and the 'primordial'.[4] Circumstantial theories emphasize the instrumental basis for ethnicity, particularly the role of self-interested rational action, while primordial theories emphasize identity based upon affective ties. Each approach, however, has been seen as incomplete in explaining the problem of ethnic boundaries, i.e. the criteria that determine membership within active ethnic groups.

The established body of literature in the primordial approach has usually been characterized as emphasizing the role of ancient history or kinship as the basis for affective, emotional attachments, which in turn form the basis for ethnic boundaries. Among the more prominent of these writings are Shils' examination of how the expansion of family ties acts as the basis for larger-scale ethnic relationships, and Geertz's extension of the notion of 'ineffability' to describe ties based upon a wider range of ascriptive factors such as religion, language, and region.[5] Somewhat more recent versions include Isaac's presentation of the need for 'basic group identity' as a human universal and Stack's examination of the 'primordial challenge' in the modern world.[6]

These writings are often criticized as ignoring the possibility of change in ethnic boundaries, and the plentiful evidence they shift over time in response to political and economic factors. From events of recent history, it has become clear that the salience and even definition of categories, from 'Latino' and 'Asian', to 'Hutu' and 'Tutsi', and to 'Croat' and 'Serb', has been far from stable and uncontested. In many parts of the developing world, the most prominent ethnic groups are of rather recent vintage, emerging as salient social entities only in the twentieth century. Moreover, primordial theories are also criticized for ignoring the fact that the particular ascriptive factor salient to ethnic boundaries varies according to situation. In some places, skin color forms the basis for ethnic divisions, in other parts, language, religion, or some combination thereof.

In part, these criticisms of primordial theories are misplaced. None of the above authors claim that kinship and ancient history are the only causes of ethnic boundaries. Indeed, they highlight the role of social change in generating ethnicity out of more parochial attachments. However, it is nonetheless true that these theories do not specify why certain ascriptive characteristics are chosen over others as the basis for ethnic boundaries, or why a particular set of boundaries may become more or less salient over time.[7]

Perhaps because of this, the circumstantial approach has predominated in recent decades over the primordial approach in the study of ethnicity. In particular, there have been a wide range of prominent theories that focus on the calculation of economic interest and its role in the rise and fall of collective action within a particular set of ethnic boundaries. Prominent examples among these include Hechter's theory of internal colonialism. Nagel and Olzak's competition theory, Bonacich's theory of split labor markets, Rogowski's theory of pillarization, and Banton's theory of ethnic monopolies.[8] Each of these theories isolates certain economic characteristics, and argues that presence of these characteristics within a given ethnic group determines the perception of mutual self-interest among group members, and hence is the key to predicting its changing level of collective action at different points in time. Hechter's theory centers around income and occupational concentration, Nagel and Olzak's on occupational overlap with other groups, and Bonacich's on occupational overlap and combined with income differences. Banton's theory focuses on the monopolization by a group of certain economic goods, and Rogowski's on economic self-sufficiency.[9]

One strength shared by these theories is that they provide falsifiable predictions about the activity level of particular ethnic groups. However, rather than examining alternative sets of boundary criteria in order to predict which will become salient for collective action, they focus on a particular set of ascripitive criteria, which they deem as defining the ethnic group, and seek to predict when it will become more or less salient for those enclosed by the criteria. In other words, rather than seeking to account for the location of ethnic boundaries, they take a fixed location and seek to account for variations in its importance for organization-building and collective action. This inability to account for the location of boundaries is inherent in their focus on material self-interest as the main determinant of salience. Indeed, the logic of material self-interest would seem to imply that boundaries for groups engaged in collective action ought to be based on economic characteristics such as income level and occupation, rather than ascriptive characteristics such as race, language, religion, and region. Hence, these theories seem to view ethnicity as, at most, an incidental appendage to economically-based interest-group formation rather than as a causal factor in its own right. Nor do these theories explain the strong role that identity linked to ascriptive ties plays in the justification that ethnic groups put forward for their activity, as well as the related fact that individuals often seem to genuinely sacrifice their own

interests and in extreme cases, their own lives for the sake of their ethnic groups.

Given the respective criticisms directed against the two approaches, efforts have been made in recent years to devise a new approach that combines aspects of both primordial and circumstantial analysis. Such an approach is sometimes called 'constructionist' in the sense that it explicitly recognizes the independent role of ascriptively-based affective attachments in determining ethnic group boundaries, yet also discusses the way in which such attachments are negotiated and mobilized in a dynamic fashion in reaction to political and economic incentives.[10]

Much of the constructionist literature focuses on case studies of ethnic groups within the U.S. Among the most prominent works is Alba's analysis of the declining salience of European national background in the behavior of Americans, and its replacement by a more encompassing white ethnicity.[11] Another is Waters' examination of individual choice in the development of American ethnicities. In it, she argues that white Americans, by virtue of their socioeconomic position, possess a wider range of options regarding ethnic identification than non-whites.[12]

Constructionism is also well-represented in the literature on Asian-American ethnicity. Espiritu's pioneering work describes how the creation of a pan-Asian identity encompassing different national origin groups took place through a process of common struggle for mutually desired political goals.[13] More recently, Okamoto has investigated this process quantitatively, testing a set of propositions on when a pan-ethnic boundary will be more salient than national origin-based ethnicity amoung Asian-Americans.[14]

Work on other American minority groups also focuses on the dynamic process of ethnicity creation. In her study of native American groups, Nagel has examined how U.S. government policy and increased opportunities for activism created incentives for 'red power' rather than assimilation.[15] Cornell focuses, among other things, on movement by native Americans into urban areas, where tribal groups were too small and weak to be politically effective.[16]

There is a largely separate literature investigating ethnic group construction globally, specifically with regards to nationalism. This literature is often tied to notions of the 'invention of tradition'.[17] Perhaps the best known work in this field is Anderson's examination of the role of capitalism and mass communications in the facilitation of 'imagined communities', i.e. groups whose common identity was not based on face-to-face interactions.[18]

Gellner, similarly, focuses on the rise of capitalism and the modern state, which in turn generates the need for a unifying language and culture for political units.[19] Their work echoes some ideas from Deutsch's earlier analysis of the role of social mobilization and communication technology in promoting nationalism.[20]

These constructionist works lean more to the circumstantialist side of the spectrum. Indeed, their main concesion to the importance of ascriptive commonality is the practical importance of common language as a facilitating factor in the formation of national groups. On the other hand, Smith's work on 'ethnosymbolism' leans more toward the primordial; it views nationalism as an activation of historic, non-instrumental ethnic community (*ethnic*) attachments.[21] Both literatures link nationalism to some existing conception of ethnicity, without emphasizing the multiplicity of boundary criteria from which ethnicity (and thus nationalism) can be chosen, much less how they are chosen.

Some constructionists, focusing on national origin-based forms of ethnicity in the U.S., view ethnic collective action as having little significance for economic or political outcomes. Drawing on the work of Herbert Gans, Alba argues that national background has been reduced to a kind of 'symbolic ethnicity', a form of 'vestigial attachment'.[22] Like Alba, Waters focuses on symbolic ethnicity, viewing ethnic identification primarily as a psychological device through which individuals come to terms with the contradictory cultural impulses of modern American society.[23]

On the other hand, those who focus on racially-based forms of ethnicity in the U.S. generally view it as having practical as well as symbolic consequences. Omi and Winant focus in their work on the 'political contestation over racial meaning', one which is central to the organization of state institutions.[24] Nagel discusses the 'political construction of ethnicity', in which officially recognized ethnic categories shape patterns of ethnic mobilization.[25]

While most work within the constructionist approach consists of empirical case studies, theoretical frameworks have been suggested as well. Cornell and Hartman describe numerous 'sites' of ethnic group construction, such as politics, labor markets, residential space, social institutions, culture, and daily experience.[26] Nagel lists external forces such as immigration, resource competition, and political access, which influence internal processes of cultural redefinition.[27] Waters identifies the factors shaping ethnicity choice as knowledge about ancestors, surname, physical appearance, and the relative status rankings of groups.[28] Fenton argues that ethnic boundaries

are created when cross-cutting ties lose influence, there is a general breakdown of civic ties, and/or the state acts in an ethnically-influenced way.[29] In the more general constructionist approach to boundaries, Tilly has identified five types of mechanisms precipitating the creation of boundaries: encounter, imposition, borrowing, conversation, and incentive shift.[30]

The constructionist approach has also produced theoretical work that focuses on the relative strength of various primordial and circumstantial factors in determining ethnic boundaries. Cornell and Spickard link three main variables to the constitution of ethnic groups: interests, institutions, and culture.[31] Cornell argues that groups with existing strong institutional or cultural ties will be less influenced by circumstances than those joined purely by self-interest. Lopez and Espiritu, on the other hand, argue that 'structural' ascriptive similarity (race, class, generation and geography) is far more important than 'cultural' ascriptive similarity (language and religion) in generating pan-ethnic development among groups of disparate national origin.[32]

Finally, there is theoretical work that discusses the relationship between primordial and circumstantial factors. McKay presents a 'matrix model' which provides labels for different cases of ethnic collective action based upon the relative strength of circumstantial and primordial forces,[33] while Scott adapts a model of George Spicer to argue that ethnic solidarity and mobilization arise when groups possessing enduring sources of identity face opposition from outside forces.[34]

While this theoretical literature is rich and insightful, it has not provided us with a general, predictive theory of ethnic boundaries. The various works discussed above list and analyze variables important in the creation and maintenance of ethnicity, recognizing how both primordidial and circumstantial factors aid the formation and action of groups. However, they do not attempt to delineate the interaction of these variables in a fashion suitable for making falsifiable predictions about boundary location. In particular, the causal role assigned to ascriptive ties is typically taken for granted rather than explained. The theories implicitly assume that such ties provide something 'extra' that boosts group solidarity, yet do not attempt to show which ascriptive characteristics will be selected in determining ethnic boundaries in different sets of circumstances, or how boundaries will change over time.

Hence, perhaps somewhat surprisingly, the plethora of work that has appeared on ethnic boundaries has yet to produce any theory that attempts to predict boundaries in a fairly general way. As noted already, the literature is valuable and informative for a variety of other reasons. However, for those who believe that predictive theory building is a key component of the social science endevour, this has placed a limit on the perceived value of its insights.

The rest of this paper attempts to build upon and extend earlier work by presenting and testing a positive theory of ethnic boundaries. The theory focuses on how altruistic, other-regarding preferences are built up over time and transformed into ethnic identities during periods of disequilibrating social change. Moreover, it analyzes the way in which conditions during periods of change affect the way in which this tranformation takes place, shaping the ethnic group boundaries that are generated. Finally, it attempts to show which types of ethnicity will tend to facilitate ethnic collective action under each set of conditions. In doing so, it predicts the type of ascriptive marker (e.g. physicial appearance, language, religion, region of origin) that will be most salient in defining the ethnic identities that people assume and the boundaries of the groups that will mobilize. This work builds upon and elaborates earlier work by the author on a positive theory of ethnic boundaries,[35] refining the propositions that were presented, and operationalizing for empirical testing.

The latter part of the paper is devoted to a statistical test of the theory using cross-national data on ethnic groups around the world. While it turns out that cross-national ethnicity data is very difficult to find, it is still possible to test the theory's predictions about the relationship between population and the ascriptive characteristics that are chosen for ethnic boundaries. In one part of the empirical test, the predictions of the theory on the salience of language and religion, relative to other ascriptive factors, in determining ethnic boundaries will be tested against data on the linguistic and religious distinctiveness of ethnic groups in different parts of the world. In another part of the test, the theory's predictions for the relative success of different ethnic groups in collective action are tested against various indicators of group organization, mobilization, and participation in conflict.

The Theory

This positive theory of ethnic boundaries is based upon two major planks: a coherence-based model of identity formation and a rational choice model of action. The first is important in determining the boundaries towards which identity is directed, and the second in showing how such identity affects action.

A theoretical combination of rationality and identity may seem at first glance somewhat self-contradictory, given that rational choice models are usually associated with the assumption of self-interest, and identity is typically seen as an antithetical to self-interest. However, the assumption that people are purely self-interested in their goals is not inherently tied to rational choice. Instead, this perceived attachment comes from the frequent association of rational choice with conventional assumptions taken from microeconomic theory. However, mainstream microeconomics employs only one kind, and the most narrow, of rational choice models. Indeed, if one views things more flexibly, the assumption that action is rational, i.e. directed towards particular goals, implies not that goals are always self-interested, but rather raises the question of what role identity plays in defining those goals.[36]

The coherence model of identity is an attempt to create a general, predictive model of preference and belief formation.[37] The basic assumption of the model is that individuals will retrospectively adjust their preferences and beliefs in order to eliminate 'expected regret' over actions that are believed to be optimal, but for which certainty is absent due to incomplete information. Where observed stochasticity in the environment makes it impossible for the individual to eliminate the beliefs that are leading to her lack of certainty, then the model implies that her preferences will change to place greater value on consequences that are more reliably linked to those actions. This in turn will cause individuals to view their actions in a teleological fashion, that is, as fulfilling a clear and predictable destiny. The model is consistent with a choice-theoretic model of action. However, unlike the conventional rational choice model, it implies that identity not only influences action, but is itself a product of action.

Among the implications of the coherence model is that individuals engaging in cooperative actions that generate uncertain outcomes will tend to develop altruistic preferences for others who benefit from their actions. Though the initial motivation for cooperation may be based upon mutual self-interest, over time cooperating individuals will tend to intrinsically value the benefits of cooperation for their partners as well as for themselves. This implication arises from the basic assumptions of the model and the fact that when an individual engages in cooperation within a group towards a mutually-desired good, her action entails both opportunity costs (the time, effort, and other resources devoted to her action) and benefits (the enhanced provision of the good) for herself, but only benefits and no costs for the other members of the group.[38] Hence, by increasing the value that she places on the provision of the good to the other members of the group rather than herself, she increases the subjective benefit/cost ratio of her actions and hence can reduce expected regret. If such interactions are repeated among the same or similar group of individuals over time, then each member of the group will internalize a group identity, i.e. incorporate other group members into her own sense of self. A strong identity will allow him or her to achieve coherence by removing the possibility that her actions will turn out to have been less than worthwhile given her values, even when the exact balance between costs and benefits to herself is not certain.

With regards to ethnicity, this model fits well with existing empirical findings. Perhaps the most notable predictive finding in social psychology on ethnic identity is the so-called 'contact hypothesis'. Briefly stated, the contact hypothesis asserts that positive feelings of identity among members of a population will be generated by repeated cooperative interactions. A wide range of experimental results have show that cooperation can create sustained fellow feeling, even among quite heterogeneous sets of people. In particular, it has been shown that sustained cooperation across racial lines tends to generate non-transient forms of identity that transcend race.[39]

In addition to these well-known results, there exist findings in sociology from experiments on group behavior showing that cooperative exchange tends to lead to an intrinsic commitment to others.[40] Finally, there exist findings in comparative political science showing that collective action to attain common goals can create group consciousness that persists beyond the interactions themselves.[41]

Given this, there seem to be some preliminary grounds for attempting to use the coherence model to predict the location of ethnic group boundaries. However, it still leaves some questions unanswered. Why should group boundaries be ascriptive? Why should individuals cooperate along lines drawn from ascriptive characteristics? Why should the ascriptive criteria for boundaries vary from place to place and across time?

In order to understand this, we need to analyze the interactive nature of the relationship between action and identity in the context of ascription. One of sociology's enduring principles is the notion that individuals grow up with a strong attachment to a 'primary group', with whom they come into frequent and wide-ranging cooperative interaction early in life.[42] In virtually every society in the world, the most important of these primary

groups is an individual's own family, either nuclear or extended. In terms of coherence theory, the tendency for an individual's most salient early identity to be centered around the family can be explained by the fact that the family is typically the group that an individual cooperates with most frequently early in life, and that these interactions tend to range over all realms of experience. As with other forms of cooperation, this cooperation can be explained from a rational point of view as being initially a matter of the individual's self interest within the confined environments of childhood. However, over time, as cooperative interaction takes place repeatedly within the primary group, altruistic preferences towards members of the group will be acquired.

A family, however, even an extended family, is not an ethnic group. Besides the difference in size, an individual's conception of her family is as a group of specific individuals rather than an abstract 'imagined community'.[43] Indeed, we can make a distinction between two types of group boundaries. A boundary that is defined by enumeration of the members within a group can be called a *concrete boundary*, while one that is defined by specifying necessary and sufficient attributes for group inclusion rather than each member individually can be called an *abstract boundary*. In general, the larger a group is, the harder it is to sustain concrete boundaries, since doing so will increasingly tax the time and cognitive resources of members to become acquainted with and remember each other individual member. Primary groups are defined by concrete boundaries, while ethnic groups are defined by abstract ones.

Because of this, the development of ethnic, as opposed to family, identity tends to occur as individuals venture out from primary groups and enter larger arenas of social interaction in which face-to-face acquaintance with all whom they are cooperating and competing is impossible. No attempt is made in this paper to claim that there is a completely uniform process whereby this occurs. Indeed the timing and circumstances of a individual's movement into larger groups will depend in large part on the specific structural conditions of the society in which the individual is living. In pre-industrial societies, such movements may not occur at all and interaction may remain restricted to primary groups within an extended family or village. In societies undergoing modernizing structural changes such as urbanization, industrialization, commercialization and political integration, individuals may first encounter large-scale interaction in urban metropoles where they migrate in search of employment once they reach adulthood. Finally, in highly urbanized societies, individuals may encounter such interaction from a much earlier age, and may do so in an environment where large-scale groups already have an established organizational presence.

Nonetheless, via the combined application of rational choice and coherence theories, we can posit certain regularities of group-formation in the latter two circumstances. In general, where the scale of interaction is very large and the breadth of possible cooperation diffuse, it reasonable to posit that it is to the advantage of individuals to form large alliances with others in order to get ahead in the competition for high-paying jobs, advantageous residential arrangements, access to political patronage, and other sources of social status. Solitary individuals and even small groups will be at a great disadvantage in competing for resources with larger groups. At the same time, groups that are overly large, i.e. larger than they need to be to dominate the others, will tend to fracture over the division of spoils amongst group members.[44]

But exactly how large should such groups be? Among the most throroughly empirically substantiated propositions within rational choice theories of political and economic competition is the proposition that actors in competition for benefits will tend to form 'minimum winning coalitions' that encompass approximately half of the total power resources in the given arena of interaction.[45] For the analysis of minimum winning coalitions in democracies, it is generally assumed that power resources are distributed evenly across a society's population, hence group size alone is seen to count. Coalitions that are too small will not be able too contend for power, while those that are too large will fragement over the division of spoils, since a subset of an excessively large coalition may still be large enough to gain power on its own. The larger the arena of interaction, the larger the rationally-sized groups will tend to be, *ceteris paribus*. It follows from these findings that when competition over resources takes place within a sufficiently large population, the groups that rational actors will form will tend to be large as well, and based upon abstract rather than concrete boundaries.

This relationship between group size and ethnic boundaries is relevant to two very recent works by Chandra and Posner. In Chandra's work, she argues that ethnic parties succeed in part when the 'ethnic group they seek to mobilize is larger than the threshold of winning or leverage imposed by the electoral system'.[46] This is not explicitly an argument about ethnic boundaries, but its logic would imply that ascriptively-defined boundaries need to be of sufficient size in order to become salient as the basis for ethnic political mobilization.

Posner's work argues that the Chewa and Tumbuka ethnic groups are adversaries in Malawi but not in Zambia because, in the latter country, their group size is too small to be used as a basis for mobilization.[47] Both works hence argue that an adequately large group size is necessary for ethnic mobilization, though they do not focus on the opposite problem of boundaries that are too large and hence tend to fragment.

However, in addition to examining group size, explaining ethinc group formation requires a certain twist away from the usual rational choice examination of group formation. While conventional rational choice theory can help us to understand group size, it does not provide us with sufficient insight into the origins of group boundaries. Indeed, as an illustration, if power resources were distributed equally in a population of size n, the number of different ways in which groups of optimal size can be formed is approximately $_nC_{n/2} = n!/(n/2)!^2$. Among all such possibilities, how would large numbers of individuals converge upon a mutually agreed-upon set of criteria for allying themselves with other individuals? Furthermore, why will individuals often form groups based upon ascriptive characteristics such as race, religion, language, and region of origin?

Here coherence theory can provide us with some answers. An identification with a primary group can be conceptualized in choice-theoretical terms as incorporation of the utility of group members into an individual's own utility function. Given such 'altrusitic' incorporation, it has been shown that individual action will not be purely opportunistic and entirely based upon self-interest.[48] More specifically, each individual's sense of primary group identification will cause them to define their goals in terms of the welfare of other members of their groups as well as their own. This in turn will make it rational for an individual to prefer cooperative action within a large group that encompasses his or her primary group over that within a similar-sized group that does not, since this means that the benefits that the individual's cooperation has for other group members will be reflected in the individual's own utility.

This rational attraction to large groups that encompass primary groups can be used to explain, from amongst the countless ways in which large groups can be organized, the pervasiveness of an ascriptive basis for cooperation in urbanizing or urbanized societies. Criteria for boundaries of large groups will have to incorporate existing kin, yet at the same time be broad enough generate sufficient size, as well as providing a social marker that can be verified with relative ease when necessary. Nearly be definition, the only characteristics that meet such requirements are ascriptive ones such as race, language, religion, and region of birth, the characteristics that usually go under the heading of 'ethnic'.

Hence both ascriptive commonality and an appropriate group size encompassing that commonality will be seen as necessary conditions for a particular boundary to become salient for group formation. However, if multiple boundaries meet such requirements, another factor that can also affect the formation of a group is whether or not a potential ethnic group boundary encompasses individuals sharing common interests based upon specific economic or cultural characerics. For instance, such shared interests can come from a common economic position, be it based upon income level,[49] occupation,[50] or control of a production sector.[51] A common political position may be created by a history of past discrimination along the boundary in question, or simply by the coincidence of the boundary with existing administrative units.[52] A common position of this kind creates a clearer policy agenda for the potential group formed by the boundary, hence increasing the expectation that collective action will result in immediate gains for group members.

This leaves us with three major propositions regarding ethnic group formation:

Proposition 1: The boundaries of a large-scale group will be based on an ascriptive attribute that surrounds rather than cross-cuts existing primary group boundaries and can be used as an extension of cosanguinity, i.e. one of the following: race, language, religion, or region of birth, or a combination of those attributes.

Proposition 2: The boundaries will also encompass an appropriate share of the power resources in the main arena of social interaction so that a 'minimum winning coalition' is formed.

Proposition 3: Where multiple potential boundaries for large scale groups meet the ascriptive and power share requirements, the one chosen will be one that brings together individuals sharing a particular common position within the economic and political structure.

To summarize, rational choice theory shows why groups of individuals have an incentive to form large groups with abstract boundaries in urbanizing or urbanized societies and predicts the approximate size of such groups in terms of power share, while indicating that groups will also tend to form around common structural positions. The coherence theory shows why such groups are not based upon arbitrary aggregations that meet minimum winning coalition requirments, but are limited

to groups that are based on boundaries described by specific ascriptive attributes. As has been noted by a number of scholars, this is ethnicity in a nutshell – the ability to combine strong identification with members of a community with rational maximization of one's goals.[53]

Data and Hypotheses

A wide-ranging quantitative analysis is helpful in testing and refining this theory. One problem for testing the theory cross-sectionally is that doing so requires unusually detailed information on social conditions in different countries in order to determine the ascriptive marker that will best fulfill the above requirements for boundary criteria. Ideally, in order to make predictions about which boundary will be chosen, a dataset should have population, socioeconomic, and political indicators for each country that are broken down according to different ascriptive characteristics such as language, religion, race, and region, as well as according to various combinations of these characteristics. Furthermore, in order to test the accuracy of predictions, it should have information on the characteristics of existing recognized ethnic groups within each country in terms of the ascriptive characteristics that set them off from the rest of the population, their socioeconomic and political position, and their level of solidarity and mobilization.

No existing dataset contains such information. Indeed, uniformly coded cross-national datasets on ethnic groups of any kind are quite rare, especially those that provide a reasonable sample across different regions of the world. Often, the main objective of most of these datasets is to provide accurate population data on ethnic groups, and with little or no data about their other characteristics.[54] Furthermore, where a greater variety of data is provided, data on large ethnic groups is typically mixed in with data on small communities that only inhabit a single settlement, and the focus is on variables that are not directly relevant for our purposes, such as the presence or absence of specific family practices or rituals.

However, information drawn from different various sources does allow us to make tentative predictions, even if the data fall short along some dimensions compared to what would be ideal. Country data were obtained for two ascriptive variables: religion and language. Data on the religious population breakdown of different countries were obtained from the CIA 1997 World Factbook. The World Factbook has 147 countries for which it carries data identifying the percentage of the total population that is comprised by the largest religious affiliation. Data on linguistic affiliations were taken from the Ethnologue dataset produced by the Summer Institute of Linguistics under the guidance of the cultural anthropologist Kenneth Pike. Ethnologue is widely considered the most complete cross national catalog of languages available. It covers 4406 distinct languages across 236 countries in its 1998 version, and provides population breakdowns for each country according to language.

The Minorities at Risk (MAR) dataset assembled by Ted R. Gurr and associates at the University of Maryland is undoubted the most detailed source of worldwide information on ethnic groups, describing numerous characteristics of over 250 recognized minority groups. This information focuses on the characteristics that distinguish them from other groups, such as their linguistic and religious distinctiveness, along with their relative population, as well as economic and political status, and it contains indicators of collective action such as organizational cohesion, mobilization and participation in conflict. It reflects subjective codings based on secondary qualitative material rather than measures taken from direct observation. Nonetheless, given the great difficulty of generating such observations in a uniform fashion across such a wide breadth of groups, it provides the closest thing available to a cross-national catalog on the characteristics of ethnic groups.

Unfortunately, no cross-national dataset with adequate coverage exists on racial or regional affiliations within different countries, nor the economic and political characteristics of different religious or linguistic affiliations, much less combinations of those affiliations. Moreover, the CIA data is incomplete in that it often specifies only the population proportion of the largest religious affiliation in each country, listing only the names of smaller affiliations or simply omitting them. While MAR identifies whether or not a recognized ethnic group is religiously or linguistically distinct, it does not in general describe the name of the religion or language of the group. Most importantly, none of the data available provides economic or political indicators for each religious and linguistic affiliation. Hence it is not possible to test whether the religion or language affiliation would meet the requirement of power share optimality.[55]

However, if we take the step of allowing population size to stand in for power resource share and examine only single-dimensional boundary criteria based on religion or language, there is a way to indirectly test the proposition that a type of affiliation whose membership meets the power resource share requirement will be the more likely to form the basis for ethnic groups compared

to those whose share is suboptimal. Admittedly, population size is not an ideal indicator, but given that it is certainly one major component of a potential group's power resource share, its use does have some validity. Using World Factbook and Ethnologue data, we can determine for each ethnic group the population percentage comprised by the largest religious or linguistic affiliation in the country where the ethnic group is located. The logic of the theory implies that the boundary described by the largest affiliation, if it defines an ethnic group boundary, will also affect the ethnic boundaries for all groups in a country, not simply the one associated with that affiliation. If the largest affiliation is too small to be optimal as the basis for ethnic boundary criteria, then all smaller affiliations will be too small as well. If the largest affiliation is too large, then chosen ethnic criteria must divide the affiliation into subgroups, in which case they cannot be of the same nature as that defining the affiliation. Hence, it follows that where the boundary described by the largest religious or linguistic affiliation is suboptimal, this affiliation will be suboptimal across all potential groups using such affiliations as ethnic markers. Thus, the closer the largest linguistic affiliation comes to meeting the power resource share requirement, the more likely that ethnic groups will be based upon linguistic distinctions. A similar logic applies for religion affiliation as well. This generates the following hypotheses:

Hypothesis 1: For each society, ethnic groups will be more likely to be based upon religion where the size of the largest religious affiliation in that society is relatively closer to the optimal minimum winning coalition size of half the population.

Hypothesis 2: For each society, ethnic groups will be more likely to be based upon language where the size of the largest linguistic affiliation in that society is relatively closer to the optimal minimum winning coalition size of half the population.

In other words, while religion and language divide populations into categories in every society, the extent to which these categories will become salient for ethnicity will depend in part on how close the subpopulations they contain are to half of the population. Ideally, of course, one would prefer to choose indicators for power resources to reflect per capita levels of income and political participation for the subpopulations described by each ascriptive marker (i.e. potential ethnic boundary marker), but given that no such measures exist, size becomes the best-available indicator.

The MAR dataset contains indicators that can be used to determine racial and regional distinctiveness of each recognized ethnic group in addition to their religious and linguistic distinctiveness, though there is no information on which ascriptive characteristics are shared with other groups. Nonetheless, there is some reason to operationalize distinctiveness so that the influence of a particular ascriptive characteristic in determining ethnic group boundaries is interpreted as being at least to some extent a zero-sum proposition. It thus makes sense to operationalize distinctiveness variables in relative terms as well as absolute terms.

Hypothesis 3: In each society, the relative strength of a religious or linguistic marker in determining ethnic boundaries vis-a-vis other possible ascriptive markers will be statistically related to the closeness to optimal size of the corresponding religious or linguistic affiliation.

As mentioned, there is no way to use data from the MAR, Ethnologue, and CIA datasets to determine levels of income or other per capita power resources among individuals who share a particular ascriptive marker. However, MAR does have data for whether or not a recognized ethnic group is economically or politically disadvantaged. Hence one can indirectly extend the operationalization of power resource share beyond mere population size by examining a logical implication of the theory – that ideal group size, in terms of members, will be larger for groups whose members have less per capita power than for groups whose members have greater power. This is because, if per capita power resources matter, disadvantaged groups will require more members to build winning coalitions than other groups. This in turn implies that:

Hypothesis 4: Recognized ethnic groups with greater economic and political resources will tend to be smaller than groups with fewer resources.

One possible operationalization of the model is to view 'groupness' as an interval variable whose strength is represented by the extent to which individuals within a particular set of boundaries form active organizations or tend to engage in collective action. This provides an alternative way of examining the importance of the group share of resources requirement, again using group population as an indicator because of the data limitations mentioned above.

It also provides perhaps the only practical way, given available data, to examine the requirement of common structural position listed in proposition 3, since economic or political disadvantage can be used as an indicator of a

shared low position within the stratification system of a group's society. However, interpretation of any results for the structural position requirement would have to be qualified, since such disadvantage is only one kind of structural commonality, and groups that are not disadvantaged may still share a distinctive economic or political position.

We can thus obtain the following hypotheses:

Hypothesis 5: Recognized ethnic groups whose size more closely matches the optimal size will tend to be more active in protest, rebellion, and organization.

Hypothesis 6: Recognized ethnic groups which are politically or economically more disadvantaged will tend to be more active in protest, rebellion, and organization.

Empirical Analysis

All tested models took ethnic groups as their unit of analysis. The first type of model tested (see Table 1, regressions 1–4, dependent variables correspond to columns) was a multivariate logistic regression, designed to test hypotheses 1 and 2. The first dependent variable was a dummy indicator of religious distinctiveness. More specifically, it was based upon the variable BELIEF in MAR, which indicates whether the group is seen to have distinct religious beliefs and practices from the other groups in the country. The dependent variable for religion is a dummy variable and was coded here as 1 if the group was coded as being from a clearly distinct religious tradition (i.e. coded as 'different religion'), and 0 otherwise. The second dependent variable is likewise a measure of linguistic distinctiveness based upon the variable LANG in MAR. It was coded here as 1 if the group's language or dialect was seen as clearly distinct from that of the rest of the population, and 0 otherwise.

The two independent variables found in each regression were indicators of the population optimality of the corresponding potential marker within the country where the ethnic group resides. The theory implies that population optimality is non-monotonic vis-a-vis proportion of population first rising, then falling as a marker boundary becomes too large to form the basis for a solitary group. The most straightforward way to model such a relationship is by including both the proportion of the population within a country comprised by the largest affiliation linked to a potential ethnic marker, as well as the squared term of this proportion.

Table 1 *Results of univariate logistic regressions*

	(1) Religious distinct	(2) Religious distinct	(3) Linguistic distinct	(4) Linguistic distinct
Intercept	−5.69*	−7.58*	−2.17*	−1.84*
	1.81	23.75	1.24	1.40
Proportion of population	16.99*	23.75*	6.16	5.72
	5.42	7.34	3.80	4.42
Proportion of population^2	−12.39*	−16.83*	−4.39	−3.05
	0.43	5.34	2.70	3.24
Economic Disadvantage × Proportion of the population^2		0.12		−0.02
		0.16		0.12
Political Disadvantage × Proportion of the population^2		0.45		0.01
		0.20		0.16
Democracy × Proportion of the population^2		0.02		−0.06
		0.06		0.05
Nat. Income × Proportion of the population^2		−0.00001		0.000004
		0.00003		0.00003
Nat. Industry × Proportion of the population^2		−0.049		−0.003
		0.028		0.02
n	264	177	297	206
Concordant	60.3	71.1	54.5	64.9
Discordant	36.1	28.6	42.4	33.7
Tied	3.6	0.3	3.1	0.4

Notes: For each regressor, the top number represents parameter estimates and the bottom standard errors.
*P < 0.05, two-tailed tests.

The model would imply a positive coefficient for the linear term and a negative one for the squared term, forming a parabola whose maximum is at about one half of the population.

In further regressions, these two independent variables were accompanied by interactions between proportion of population squared with measures of economic and political disadvantage of these groups, taken from MAR, as well as additional country-level information on income level and level of industrialization taken from the World Factbook and on level of democracy from MAR. Such interactions can provide some indication of whether the appropriate level for an religious or linguistic affiliation to be selected as an ethnic marker is raised or lowered by the average income or level of power of a member of the group defined by such a marker. Given that the logic of the theory suggests that optimality should reflect the proportion of power held within the boundary defined by the marker as much as the proportion of population, we would expect a positive coefficient on the interaction term with economic and political disadvantage, indicating that optimal size is higher for markers that surround disadvantaged subpopulations rather than those surrounding advantaged subpopulations. The indicators for country income, industrialization, and democracy were chosen to test the notion that more narrowly defined groups tend to rise under wealthy, industrialized, and democratic systems that under poor, agrarian, and non-democratic ones, due to greater openness in such societies. If this is the case, one would expect a negative coefficient on these interaction terms.

In the results of these regressions, the marker optimality variables had the expected signs, i.e. positive for the linear terms and negative for the squared ones, across all four equations. Moreover, the coefficients for religious distinctiveness were significant at the $P < 0.05$ across both the linear and square terms across both regressions for religion – the only coefficients to reach significance in any of the regressions. None of the interaction terms were significant, and indeed all but the terms for political disadvantage changed signs across the regression for religion and that for language.

One interesting thing to note is that the linear term was of slightly larger magnitude than the squared one, but with a fairly stable ratio, between 1.3 and 1.5 times in the first three equations, and slightly less than 2 in the fourth. The optimal proportion of the population indicated by the equation $y = \alpha_1 p + \alpha_2 p^2$ in which p stands for proportion of population, with $\alpha_1 > 0$ and $\alpha_2 < 0$, will be the point at which dy/dp reaches a maximum,

which in turn will be the solution to the first-order equation $\alpha_1 + 2\alpha_2 p = 0$, or $p^\star = \alpha_1/(2\alpha_2)$. Given α_1 between 1.3 and 1.5 times α_2, this implies an optimal proportion of between 0.65 or 0.75 of the total population, slightly larger than predicted by the theory.

The second type of model was based upon ordinary least squares regressions (see Table 2, regressions 5 and 6), and designed to test hypothesis 3. The first dependent variable was simply an index formed by taking the dummy variable for religious distinctiveness, multiplying by three, then subtracting the sum of dummy variables for the group's linguistic, racial, and regional distinctiveness, taken from MAR. Such an index controls in a rough fashion for the fact that different potential ethnic markers may cluster together *a priori*, and that population optimality for one potential marker may overlap with that for another. In other words, the index measures not simply whether religious affiliation will be selected as a marker, but whether it will be selected above other potential markers for ethnicity. A similar index was constructed for linguistic distinctiveness, multiplying by three, then subtracting the sum for religious, racial, and regional distinctiveness. As with the logistic regressions, the signs for both the linear and the squared term had their expected sign. However, in these regressions the terms were significant for both religion and language at the $P < 0.05$ level. Moreover, the ratios between the magnitudes of the linear and squared terms were again in the 1.3 to 1.5 range for both regressions.

Three additional regressions (also on Table 2, regressions 7–9), designed to test hypothesis 4, took a group's proportion of the population as their dependent variable, examining which sorts of economic and political factors might cause average ethnic group size to be smaller or larger. In this case, the proportion of the population in question was that of the ethnic group itself, not the largest group defined by a religious or linguistic marker in its country. In particular, this was another way to test whether relative economic or political resource advantages would affect the size of the boundary around which an ethnic group will form, independent of the specific marker used to describe it. The presence of political and economic 'distinctiveness' are coded as ordinal variables in MAR, though somewhat confusingly different scales are used for indicating disadvantaged as opposed to advantaged groups. Furthermore, very few groups-were marked as 'advantaged'. Because of this, a simple dummy variable was created to indicate those groups that MAR coded as disadvantaged economically. There was no clear pattern across these results, except for the fact that democracy seemed to

Table 2 *Results of OLS regressions*

	(5) Religious distinct	(6) Linguistic distinct	(7) Proportion population	(8) Proportion population	(9) Proportion population	(10) Protest	(11) Rebellion	(12) Organization
Intercept	−6.03*	−2.71*	0.19*	0.13*	0.19*	−0.81	−2.16	0.26
	1.39	1.01	0.03	0.01	0.03	1.55	2.28	2.18
Proportion of	17.73*	9.32*				7.64	13.94*	11.02
population	4.20	3.14				4.63	6.82	6.50
Proportion of	−13.06*	−6.98*				−4.95	−9.06	−8.36
population^2	2.96	2.27				3.26	4.80	4.58
Country			0.0000005		0.0000006	0.00003*	−0.00003	0.00003
income			0.000001		0.000001	0.00002	0.00002	0.00002
Country			−0.001		−0.001	−0.008	−0.01	−0.01
industry			0.0008		0.0009	0.01	0.02	0.02
Democracy			−0.009*		−0.009*	0.07*	−0.15*	0.04
			0.002		0.002	0.03	0.04	0.04
Economic				−0.01*	−0.008	0.04	0.08	0.16
disadvantage				0.006	0.006	0.08	0.11	0.11
Political				0.007	0.008	0.04	−0.15	−0.21
disadvantage				0.007	0.007	0.09	0.14	0.13
n	230	260	229	241	209	173	175	176
R^2	0.0857	0.0361	0.0768	0.0242	0.0846	0.0942	0.1035	0.0621

Notes: For each regressor, the top number represents parameter estimates and the bottom standard errors.
*$P < 0.05$, two-tailed tests.

lower the average size of an ethnic group, providing some very limited support for the hypotheses about system openness and ethnic group size.

The final three regressions (10–12), designed to test hypotheses 5 and 6, took as their dependent variables measures of political protest, rebellion, and organization, respectively, each taken from MAR. As mentioned, theories of ethnic mobilization and solidarity have focused primarily on an ethnic group's economic and political position to account for its degree of activity. In addition to such factors, these equations attempted to investigate the effect of population optimality. Again, the proportion of the population in question was that of the ethnic group itself, not that of any religious or linguistic affiliation, and the calculations were independent of the specific marker used to define the group. The hypothesis here was that, even if the percentage of the population described by a chosen ethnic marker is close enough to the optimal to allow the group to survive as a recognized and well-defined entity, those groups that are closer yet to the optimum will tend to be more active than those who are relatively further away. Moreover, we should expect the signs on the linear and squared terms to be similar to those found in the previous regressions.

Interestingly, despite the change in the source of the population proportion variable from regressions 1–4, the resulting coefficients not only maintained the same signs as those in the earlier regressions, but also similar magnitudes relative to one another that they had in the earlier ones, again at the 1.3 to 1.5 range, suggesting an optimal proportion of between 0.65 and 0.75 of the population. Though they were not statistically significant in this case (except the linear term vis-a-vis rebellion), the conformity of their signs and relative magnitudes across the board with those of the earlier ones suggests that there is an underlying pattern behind this series of results. Again, the variables for the economic and political characteristics of the group and of their country were not significant and exhibited no clear-cut pattern, failing to confirm hypothesis 6. However, the interpretation of the results remain ambiguous for the proposition regarding common structural position, since there was no way to obtain data for the other ways in which groups might be structurally distinct.

Hence the results of the regressions for the theory are somewhat mixed, but they largely support the idea that the ascriptive markers that will be chosen for an individual's ethnic identity will be those that best conform to the optimal group size given the overall population. Furthermore, they support the idea that the closer the size of a group is to the optimal size, the more effective the group will be in pursuing collective action. The fact that the optimal proportion of the population is slightly larger than predicted by the theory is interesting, and suggests that certain factors, such as international competition or diseconomies of scale in the exercise of power, may inhibit a group comprising over half the population from being an oversized coalition and therefore fragmenting.

Conclusion

As with all general, predictive theories, an emphasis on parsimony was placed on the theory of ethnic boundaries presented here. This meant that there was of course no way to incorporate all significant factors that might have influence on the formation of ethnic boundaries. Indeed, the typical goal for such a theory is to tie together a number of important factors is an original fashion, and furthermore to generate nontrivial, testable predictions that are not disconfirmed by the data. Focusing on such limited goals was especially important for this topic, where there has been little previous effort aimed at general prediction and where comprehensive data is very difficult to obtain.

Nonetheless, there are ways in which the theory could be extended in the future. Most notably, while the theory focuses on a single set of ethnic boundaries within a single arena of social interaction, it is not inconsistent with the possibility that individuals may simultaneously carry with them multiple levels of identity that are salient for collective action, each operating at a different arena. One natural extension of the theory presented here is the notion that such multiple identities will not cross-cut one another. Even if ethnic collective action is initially built upon self-interest and an attachment only to the subset of the ethnic group comprising the primary group, the logic of the coherence model implies that as repeated collective action occurs within ethnic boundaries, a broader attachment will result. Once this occurs, it will set conditions on the boundaries of other groups with which the member identifies and acts collectively. Hence, where the different arenas of interaction involve magnifications in scale, such as provincial, country, regional, and world arenas, the model would imply that identities at highers level of interaction will encompass rather than overlap with those at lower levels, just as the initial formation of ethnic identity encompasses parochial attachments to a primary group. The labels that are typically placed on identities at such levels could be pan-ethnic, national, and even civilizational rather than ethnic, but similar principles of boundary formation will apply.

An explicit set of propositions regarding the way in which these various layers of identity interact with one another would have implications on the issue of ethnic boundary change. Though individuals may possess multiple identities, the relative saliences of these identities need not be equal. Saliences may change over time, and particular identities may supercede instead of supplement others. Again, the logic of the theory does provide material from which propositions can be generated, since the rational choice model suggests that the salience of a particular identity 'layer' will depend in part on the incentives to act collectively at the corresponding arena of interaction. This in turn could quite plausibly be linked to the amount of zero-sum goods being contested in that arena. For example, as the primary arena at which goods are allocated moves from the national arena to the regional or international, one would expect subnational ethnic boundaries to become less salient relative to larger boundaries.

Extensions of the theory aside, the results presented in this paper seem to justify future empirical analysis which will uncover in more detail the ways in which key variables of the theory relate to ethnic identity and collective action outcomes. Furthermore, they provide clear evidence that ideas about ethnic identity formation and social construction need not be seen as contradictory to the methodology of positive social science, and that clear patterns exist in the way in which structural factors affect the types of identities that are formed. However tentatively, this paper puts forward a positive theory of ethnic identity formation that predicts the outcome of social construction, and subjects it to empirical testing using mainstream statistical techniques. The door remains open to further work in this direction.

Notes

1. Barth (1969); DeVos (1975); Glazer and Moynihan (1975); Young (1976); Rothschild (1981); Keyes (1981); Horowitz (1985); Yinger (1985); Kasfir (1986); Brass (1991).

2. Lamont and Molnar (2002); Tilly (2004). The Symbolic Boundaries Research Network of the Theory Section of the American Sociological Assocation is another example of this concern (see <http://www.people.virginia.edu/˜bb3v/symbound/>).

3. See for instance discussions in Fenton (1999: 3–4) and Spickard and Burroughs (2000: 3–5).

4. Scott (1990); Stack (1988); Douglass (1988); McKay (1982).

5. Shils (1957); Geertz (1963).

6. Isaacs (1975); Stack (1988).

7. Evolutionary theory is typically seen as one type of 'primordial' approach to ethnic boundaries. As such it can explain why ascriptive connections may affect behavior (through genetic ties and motives derived from inclusive fitness) but cannot readily explain why different types of ascriptively-based boundaries will be chosen over others for identity among different populations, or for that matter why non-biological ascriptive characteristics such as religion or language may be the basis for ethnic boundaries. See Chai (2001); 63–65 for a more complete discussion of these points.

8. Hechter (1971); Hechter (1975, 1978); Nagel and Olzak (1982); Olzak and Nagel (1986); Olzak (1990); Bonacich (1972, 1979); Rogowski and Wasserspring (1971); Rogowski (1974); Banton (1980, 1983).

9. For a longer review of these theories, see Chai (2001, ch. V).

10. Cornell (1996, 1998, ch. IV); Nagel (1994); Spickard and Burroughs (2000). See also Cerulo (1997).

11. Alba (1990); see also Gans (1979).

12. Waters (1990).

13. Espiritu (1997).

14. Okamoto (2000, 2003).

15. Nagel (1994, 1995, 1996); Nagel and Snipp (1996).

16. Cornell (1988).

17. Hobsbawm and Ranger (1983); Vail (1991).

18. Anderson (1983).

19. Gellner (1969, 1983).

20. Deutsch (1953).

21. Smith (1986, 1998, 1999, 2001)

22. Alba (1990: 306)

23. Waters (1990, 148–50).

24. Omi and Winant (1994: 68–70, 76–78).

25. Nagel (1986: 100–102).

26. Cornell and Hartman (1998, ch. VI).

27. Nagel (1994).

28. Waters (1990, ch. III). For multi-ethnic choices, see Waters (2000).

29. Fenton (1999, ch. III).

30. Tilly (2003, 2004).

31. See Cornell (1996); Cornell and Hartman (1998, ch. IV); Spickard and Burroughs (2000, ch. I); Spickard (1996).

32. Lopez and Espiritu (1990).

33. McKay (1982).

34. Scott (1990).
35. Chai (1996, 2001).
36. Chai (1997).
37. The general model and its implications is discussed in detail in Chai (2001), especially chapter III.
38. For a discussion of this, see Chai (2001: 116–120). The basic insight about the personal costs and benefits to self for cooperative action is an adaption of Olson (1965: 9–16).
39. The classic studies are Allport (1954); Sherif *et al.* (1961). An well-known application to education is Aronson (1978). More recent developments in the theory are discussed in Pettigrew (1998).
40. Lawler and Yoon (1993, 1996); Tallman *et al.* (1991).
41. Huntington and Nelson (1976, ch. IV); Horowitz (1975).
42. Cooley (1909).
43. Anderson (1983).
44. Those who would like specific cases studies of such movements and immigrations can consult an earlier paper by the author (Chai, 1996). Given that the current paper is theoretical and quantitative, there was not sufficient room to summarize cases here. However, it is hoped that the basic proposition that migration into areas of large-scale interaction requires the formation of large-scale groups is sufficiently uncontroversial as a rough generalization for those familiar with social change and development.
45. Riker (1962); Waltz (1979).
46. Chandra (2004: 15).
47. Posner (2004).
48. Collard (1978); Boulding (1973); Margolis (1982).
49. Hechter (1971); Bonacich (1972)
50. Hechter (1978).
51. Banton (1983).
52. Nagel (1986); Brass (1985).
53. Epstein (1978: xi); Bell (1975: 169); Rothschild (1981: 61); Nagata (1981: 112).
54. See discussion in Fearon (2003). See also Fearon and Laitin (2003) for an application of Fearon's own dataset.
55. If such data existed, it would be possible to construct an indicator, for example, for economic power resource share within the boundaries of a particular religious affiliation, by multiplying the per capita income of members of that affiliation with the population size of the affiliation.

Acknowledgements

References

Alba, R. D. (1990). *Ethnic Identity: The Transformation of White America*. New Haven, CT: Yale University Press.

Allport, G. W. (1954). *The Nature of Prejudice*. Cambridge, MA: Addison-Wesley.

Anderson, B. (1983). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.

Aronson, E. (1978). *The Jigsaw Classroom*. Beverly Hills, CA: Sage Publications.

Banton, M. (1980). Ethnic Groups and the Theory of Rational Choice. In Blumer, H. and Duster, T. (Eds.) *Sociological Theories: Race and Colonialism*. Paris: UNESCO.

Banton, M. (1983). *Racial and Ethnic Competition*. Cambridge: Cambridge University Press.

Barth, F. (Ed.) (1969). *Ethnic Groups and Boundaries: The Social Organization of Cultural Difference*. Bergen-Oslo: Universitets Forlaget.

Bell, D. (1975). Ethnicity and Social Change. In Glazer N. and Moynihan D. P. (Eds.) *Ethnicity: Theory and Experience*. Cambridge, MA: Harvard University Press.

Bonacich, E. (1972). A Theory of Ethnic Antagonism: The Split Labor Market. *American Sociological Review*, **37**; 547–59.

Bonacich, E. (1979). The Past, Present and Future of Split Labor Market Theory. *Research in Race and Ethnic Relations*, **1**; 17–64.

Boulding, K. E. (1973). *The Economy of Love and Fear: A Preface to Grants Economics*. Belmont, CA: Wadsworth Publishers.

Brass, P. (Ed.) (1985). Ethnic Groups and the State. In *Ethnic Groups and the State*. London: Croom Helm.

Brass, P. (1991). *Ethnicity and Nationalism: Theory and Comparison*. Newbury Park, CA: Sage Publications.

Cerulo, K. (1997). Identity Construction: New Issues, New Directions. *Annual Review of Sociology*, **23**; 385–409.

Chai, S. (1996). A Theory of Ethnic Group Boundaries. *Nations and Nationalism*, **2(2)**; 281–307.

Chai, S. (1997). Rational Choice and Culture: Clashing Perspectives or Complementary Modes of Analysis?

In Ellis, R. and Thompson, M. (Eds.) *Culture Matters.* Boulder, CO: Westview Press, 45–58.

Chai, S. (2001). *Choosing an Identity: A General Model of Preference and Belief Formation.* Ann Arbor: University of Michigan Press.

Chandra, K. (2004). *Why Ethnic Parties Succeed: Patronage and Ethnic Head Counts in India.* Cambridge: Cambridge University Press.

Collard, D. (1978). *Altruism and Economy: A Study in Non-Selfish Economics.* New York: Oxford University Press.

Cooley, C. H. (1909). *Social Organization.* New York: C. Scribner's Sons.

Cornell, S. (1988). *The Return of the Native: American Indian Political Resurgence.* Oxford: Oxford University Press.

Cornell, S. (1996). The Variable Ties that Bind: Content and Circumstance in Ethnic Processes. *Ethnic and Racial Studies*, **19(2)**; 265–89.

Cornell, S. and Hartmann, D. (1998). *Ethnicity and Race: Making Identities in A Changing World.* Thousand Oaks, CA: Pine Forge Press.

Deutsch, K. (1953). *Nationalism and Social Communication.* Cambridge, MA: MIT Press.

DeVos, G. (1975). Ethnic Pluralism: Conflict and Accomodation. In DeVos G. and Romanucci-Ross L. (Eds.) *Ethnic Identity: Cultural Continuities and Change.* Palo Alto, CA: Mayfield Publishing Company.

Douglass, W. A. (1988). A Critique of Recent Trends in the Analysis of Ethnonationalism. *Ethnic and Racial Studies*, **11(2)**; 192–206.

Epstein. A. L. (1978). *Ethos and Identity.* London: Tavistock Publications.

Espiritu, Y. L. (1997). *Asian American Panethnicity: Bridging Institutions and Identities.* Philadephia: Temple University Press.

Fearon, J. D. (2003). Ethnic and Cultural Diversity by Country. *Journal of Economic Growth*, **8(2)**; 195–222.

Fearon, J. D. and Laitin D. (2003). Ethnicity, Insurgency, and Civil War. *American Political Science Review*, **97(1)**; 75–90.

Fenton, S. (1999). *Ethnicity: Racism, Class, and Culture.* Lanham, MD: Rowman and Littlefield.

Gans, H. J. (1979). Symbolic Ethnicity: The Future of Ethnic Groups and Cultures in America. *Ethnic and Racial Studies*, **2**; 1–20.

Geertz, C. (1963). The Integrative Revolution: Primordial Sentiments and Civil Politics in the New States. In Geertz C. (Ed.) *Old Societies and New States: The Quest for Modernity in Asia and Africa.* New York: Free Press of Glencoe.

Gellner, E. (1969). *Thought and Change.* London: Weidenfeld and Nicholson.

Gellner, E. (1983). *Nations and Nationalism.* Oxford: Blackwell.

Glazer, N. and Moynihan, D. P. (Eds.) (1975). *Ethnicity: Theory and Experience.* Cambridge, MA: Harvard University Press.

Hechter, M. (1971). Towards a Theory of Ethnic Change. *Politics and Society*, **2(1)**.

Hechter, M. (1975). *Internal Colonialism: The Celtic Fringe in British National Development.* London: Routledge and Kegan Paul.

Hechter, M. (1978). Group Formation and the Cultural Division of Labor. *American Journal of Sociology*, **84(2)**; 293–318.

Hobsbawm, E. J. and Ranger, T. (Eds.) (1983). *The Invention of Tradition.* Cambridge: Cambridge University Press.

Horowitz, D. L. (1975). Ethnic Identity. In Glazer N. and Moynihan D. P. (Eds.) *Ethnicity: Theory and Experience.* Cambridge, MA: Harvard University Press.

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

Huntington, S. P. and Nelson, J. M. (1976). *No Easy Choice: Political Participation in Developing Countries.* Cambridge, MA: Harvard University Press.

Isaacs, H. (1975). *Idols of the Tribe. Group Identity and Political Change.* New York: Harper and Row.

Kasfir, N. (1986). Explaining Ethnic Political Participation. In Kohli, A. (Ed.) *The State and Development in the Third World.* Princeton: Princeton University press.

Keyes, C. F. (Ed.) (1981). *Ethnic Change.* Seattle: University of Washington Press.

Lamont, M. and Molnar, V. (2002). The Study of Boundaries in the Social Sciences. *Annual Review of Sociology*, **28**; 167–95.

Lawler, E. J. and Yoon, J. (1993). Power and the Emergence of Commitment Behavior in Negotiated Exchange. *American Sociological Review*, **58(4)**; 465–81.

Lawler, E. J. and Yoon, J. (1996). Commitment in Exchange Relations: Test of a Theory of Relational Cohesion. *American Sociological Review*, **61(1)**; 89–109.

Lopez, D. and Espiritu, Y. L. (1990). Panethnicity in the United States: a Theoretical Framework. *Ethnic and Racial Studies*, **13(2)**; 198–224.

Margolis, H. (1982). *Selfishness, Altruism and Rationality: A Theory of Social Choice.* Cambridge: Cambridge University Press.

McKay, J. (1982). An Exploratory Synthesis of Primordial and Mobilizationist Approaches to Ethnic Phenomena. *Ethnic and Racial Studies*, **5(4)**; 395–420.

Nagata, J. (1981). In Defense of Ethnic Boundaries: The Changing Myths and Charters of Malay Identity. in Keyes C. F. (Ed.) *Ethnic Change.* Seattle: University of Washington Press.

Nagel, J. (1986). The Politization of Ethnicity. In Olzak S. and Nagel J. (Eds.) *Competitive Ethnic Relations*. Orlando: Academic Press.

Nagel, J. (1994). Constructing Ethnicity: Creating and Recreating Ethnic Identity and Culture. *Social Problems*, **41(1)**; 152–76.

Nagel, J. (1995). American Indian Ethnic Renewal: Politics and the Resurgence of Identity. *American Sociological Review*, **60(6)**; 947–65.

Nagel, J. (1996). *American Indian Ethnic Renewal: Red Power and the Resurgence of Identity and Culture*. Oxford: Oxford University Press.

Nagel, J. and Olzak, S. (1982). Ethnic Mobilization in New and Old States: An Extension of the Competition Model. *Social Problems*, **30(2)**; 127–43.

Nagel, J. and Snipp, M. (1993). Ethnic Reorganization: American Indian Socia, Economic, Political and Cultural Strategies for Survival. *Ethnic and Racial Studies*, **16(2)**; 203–35.

Okamoto, D. (2000). *Pan-Asian Ethnic Social Movements*. Dissertation, University of Arizona Sociology Department.

Okamoto, D. (2003). Toward a Theory of Panethnicity: Explaining Asian American Collective Action. *American Sociological Review*, **68(6)**; 811–42.

Olson, M. (1965). *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.

Olzak, S. and Nagel, J. (Eds.) (1986). *Competitive Ethnic Relations*. Orlando: Academic Press.

Olzak, S. (1990). *Dynamics of Ethnic Competition and Conflict*. Stanford: Stanford University Press.

Omi, M. and Winant, H. (1994). *Racial Formation in the United States: From the 1960s to the 1990s*. New York: Routledge.

Pettigrew, T. F. (1998). Intergroup Contact Theory. *Annual Review of Psychology*, **49**; 65–85.

Posner, Daniel N. (2004). *The Political Salience of Cultural Difference: Why Chewas and Tumbukas are Allies in Zambia and Adversaries in Malawi. American Political Science Review*, **98**, 4 (November 2004).

Riker, W. H. (1962). *The Theory of Political Coalitions*. New Haven: Yale University Press.

Rogowski, R. (1974). *Rational Legitimacy*. Princeton: Princeton University Press.

Rogowski, R. and Wasserspring, L. (1971). *Does Political Development Exist? Corporatism in Old and New Societies*. Beverley Hills: Sage Publications.

Rothschild, J. (1981). *Ethnopolitics: A Conceptual Framework*. New York: Columbia University Press.

Scott Jr, G. M. (1990). A Resynthesis of the Primordial and Circumstantial Approaches to Ethnic Group Solidarity: Towards an Explanatory Model. *Ethnic and Racial Studies*. **13(2)**; 147–71.

Sherif, M., Harvey, O. J., White, B. J., Hood, W. R. and Sherif, C. W. (1961). *Intergroup Competition and Cooperation: The Robbers Cave Experiment*. Norman, OK: University Book Exchange.

Shils, E. (1957). Primordial, Personal, Sacred and Civil Ties. *British Journal of Sociology*, **8**; 130–45.

Smith, A. D. (1986). *The Ethnic Origins Of Nations*. Oxford: Blackwell.

Smith, A. D. (1998). *Nationalism And Modernism: A Critical Survey Of Recent Theories Of Nations And Nationalism*. London: Routledge.

Smith, A. D. (1999). *Myths And Memories Of The Nation*. Oxford: Oxford University Press.

Smith, A. D. (2001). *Nationalism: Theory, Ideology, History*. Cambridge: Polity Press.

Spickard, P. R. (1996). *Japanese Americans: The Formation and Transformations of an Ethnic Group*. New York: Twayne.

Spickard, P. and Burroughs, W. J. (Eds.) 2000. *We Are a People: Narrative and Multiplicity in Constructing Ethnic Identity*. Philadelphia: Temple University Press.

Stack Jr, J. P. (Ed.) (1988). *The Primordial Challenge: Ethnicity in the Contemporary World*. New York: Greenwood Press.

Tallman, I., Gray, L. N. and Leik, R. K. (1991). Decision, Dependency and Commitment: An Exchange Based Theory of Group Formation. *Advances in Group Process*, **8**; 227–57.

Tilly, C. (2003). Identities and Boundary Activation. *Symbolic Boundaries: Change (http://www.people. virginia.edu/~bb3v/symbound/conf2003/vforum2.html)*.

Tilly, C. (2004). Social Boundary Mechanisms. *Philosophy of the Social Sciences*, **34(2)**; 211–36.

Vail, L. (Ed.) (1991). *The Creation of Tribalism in Southern Africa*. Berkeley, CA: University of California Press.

Waltz, K. (1979). *Theory of International Politics*. Reading, MA: Addison-Wesley Publishing Co.

Waters, M. C. (1990). *Ethnic Options: Choosing Identities in America*. Berkeley, CA: University of California Press.

Waters, M. C. (2000). Multiple Ethnicities and Identity Choices in the United States. In Spickard P. and Burroughs W. J. (Eds.) Philadelphia: Temple University Press.

Yinger, J. M. (1985). Ethnicity, *Annual Review of Sociology*, **11**; 151–80.

Young, C. (1976). *The Politics of Cultural Pluralism*. Madison, WI: University of Wisconsin Press.

Author's Address

Department of Sociology, University of Hawaii, Honolulu, HI 96822, USA. Tel.: +1 808 9567234; Fax: +1 808 9563707, Email: sunki@hawaii.edu.