

**Ordinary Language and External Validity:
Specifying Concepts in the Study of Ethnicity***

James D. Fearon and David D. Laitin
Department of Political Science
Stanford University
9/28/00

ABSTRACT

Do greater cultural differences between ethnic groups increase the probability of ethnic violence between them? Almost invariably, the questions of interest to social scientists are posed in terms of concepts rooted in ordinary language. In this case, for instance, an empirical inquiry would require a statement of the meaning of “cultural differences,” “ethnic group,” and “ethnic violence.” We argue against the common social scientific practice of merely stipulating definitions (e.g., “By ‘ethnic group,’ we mean ...). Instead, before jumping to stipulation, social scientists should analyze ordinary language usage in order to explicate the range of current meanings. Contrary to some interpretivist arguments against the possibility of cross-cultural analysis and generalization, this approach can facilitate quantitative studies addressing social scientific questions. We illustrate and defend these claims with ordinary language analyses of the meaning of “ethnic group,” “ethnic violence,” “dominant ethnic group,” and “cultural differences.”

*This is a revised version of a paper presented at the 2000 Annual Meetings of the American Political Science Association, Washington, D.C., September 2000.

I. Introduction

Consider the following hypothesis: Members of an ethnic or religious minority are more likely to be engaged in fighting with the state, the greater the cultural divide between the minority and the dominant ethnic or religious group in the state.

This hypothesis is extremely common both in the media and in the analyses of state officials and policy-makers. Foreign policy analysts tend to explain internal conflicts as the consequence of the long-standing cultural differences that distinguish ethnic and religious minorities from majorities.¹ In the academy, modernist theories of nationalism explain separatist nationalist movements as the result of cultural barriers to upward economic mobility in the form of discrimination by dominant cultural groups. These barriers are said to be more likely and more severe the greater the preexisting cultural differences between groups.²

To test this hypothesis systematically, we would have to be able to answer a variety of questions, among them: What is an ethnic group? What is a dominant ethnic group in a country? How does one measure “cultural difference” or distance? If the outcome variable is “ethnic violence,” we will have to be able to say what this is as well. Further, the answers to these questions will have to be posed in a way that allows us to classify groups and measure degrees of cultural difference and ethnic violence across cases, in a broad range of countries.

These are not simple tasks. Consider the problem of identifying the dominant ethnic group in a country (putting aside, for the moment, the question of what an ethnic group is). Suppose we answer, “the dominant ethnic group is the ethnic group with the largest share of total country population.” Applying this rule rigorously will lead us to code Hutus as the dominant ethnic group in Burundi, and make other classifications that seem similarly bizarre. Suppose we try again, saying, “the dominant ethnic group in a country is the ethnic group to which the head of state belongs.” But then African Americans in the U.S. would be coded as dominant if Colin Powell were elected president. Faced with problems of this sort, social scientists have generally had two types of reaction.

1. The quantoid, or Humpty Dumpty reaction, holds that such confusions arise due to the inherent imprecision and unscientific nature of ordinary language concepts such as “ethnic group” and “dominant group.” The proper, scientific response is just to *stipulate* exactly what one means by one’s concept.³ Thus, if a particular definition of a concept generates unintuitive codings (as with the Colin Powell example above), this does not necessarily mean that there is anything wrong with the definition. The social scientist’s task is to discover causal relationships between things in the social world, and it is no matter if these things do not map precisely onto ordinary language notions. Indeed, they should not be expected to because ordinary language conceptions are an unscientific mess.
2. The interpretivist reaction is to claim that difficulties like those encountered above show that the whole project has to be abandoned – there is just no way to define a concept of “ethnic group” or “dominant group” that can be validly applied across a range of cases in quite different cultural and political settings. Instead, all we can do is ask about the meaning of “ethnic group” or “dominant group” in particular places; generalizing such concepts is both impossible and pernicious. We may be able to say what “ethnic group” means to people in a particular country, but because this will be quite different from the meaning in another country, any project of testing hypotheses about ethnic groups across countries is hopeless and misleading.

In this paper we develop a position at odds with both the quantoid and interpretivist reactions.⁴ Against the former, we argue that rather than dismissing the ordinary language meanings of concepts used in social science inquiry, social scientists should devote serious effort to explicating their ordinary language meanings (where “ordinary language” refers here to the speech community from which the theory or hypothesis to be tested emerges). Against the latter, we argue that rooting concepts in an analysis of their meaning in a particular language does not invalidate efforts to test hypotheses using these concepts across cultures.

Though we briefly offer some general arguments on behalf of the approach in section 2, we develop these general positions more by way of example than abstract argument. In sections 3-6, we offer ordinary language analyses of the meaning of “ethnic group,” “ethnic violence,” “cultural distance,” and “dominant group,” with a view to testing a hypothesis like that posed at the outset.

II. Defining social science concepts

It is inconceivable that a social science could base itself entirely on concepts that bore no relation to the concepts that the subjects of social science themselves use in explaining their own actions. If people’s actions are influenced by the social concepts they hold, then explicating their actions requires references to these concepts. In addition, the questions that social scientists seek to answer – for instance, questions about war, democracy, or well-being – are always posed at first in everyday language.

If so, then stipulating or legislating the meaning of a social science concept without first explicating the range of its ordinary language meanings is a dangerous practice.⁵ To see why, note that it would be inconceivable for a student of democracy to declare, “By *democracy* I mean any stable system of government.” Stipulated meanings of social science concepts that are not neologisms have to bear *some* relationship to what the reader already understands by the term -- else what is the point of using a particular term? But this implies that without an explication of the term’s current meanings, neither author nor reader can have a clear sense of how the author has interpreted and modified the concept. Moreover, attempts to legislate the definition of a concept that has a strong base in ordinary language are likely to fail anyway, since both author and reader are apt to slip unconsciously back towards the everyday meaning.

There is a second argument in favor of explicating before stipulating. On examination, everyday understandings of important social science concepts such as ethnic group, democracy, identity, and rationality prove to have complex and interesting underlying structures.⁶ This claim is best demonstrated by way of the examples discussed in subsequent sections.⁷ We hope to show that the common view in political science that ordinary language meanings are suspect for being messy, imprecise, unscientific -- fit only for a good whipping by disciplinary social scientists -- is wrong.

How does one explicate the “ordinary language meaning” of a term? Here is a recipe.

1. Think of a set of cases to which the concept in question would clearly apply (that is, “almost everyone would agree that (say) Britain is a democracy”, or “... Armenians are an ethnic group,” or “... whites are the dominant racial group in the United States.”).
2. Propose a definition that would “cover” these cases. This entails finding a principle or principles that govern(s) the attribution. The nature of these principles is often not at all obvious even though we must know them implicitly since we often have strong intuitions about what is and what is not an X (e.g., democracy, ethnic group).
3. Ask if a mechanical application of this definition would force one to admit other cases under the concept that seem intuitively odd or wrong. For instance, defining “dominant ethnic group” as plurality ethnic group leads to Hutus being classified as the dominant group in Burundi. In considering possible cases, one should consider not only actual, real-world cases, but also hypothetical ones that would meet the definition but seem intuitively wrong (for example, “if Colin Powell were president...”).
4. Go back to step 2, trying to formulate principles that will cover all the cases now in view. If possible, avoid covering the cases by proliferating “or”s, especially if the added condition is intended merely to characterize a single idiosyncratic case rather than capture a principle.

5. Continue until either (a) one can't think of any more exceptions to intuition, which is unlikely, or (b) one is considering cases for which it is implausible to argue that any competent member of the researcher's language community could have a strong intuition about whether this is an X. At this point it may be plausible to say that one has pushed the intuitions behind standard usage as far as they can go.

The goal is to formulate a simple statement of the concept's meaning that maximizes the number of cases it gets right in light of prior intuitions about proper usage. For any interesting concept there are likely to be cases that "escape" even a good statement of the implicit principles that govern the classification of most cases. For purposes of a cross-sectional study, the natural way to handle such problem cases is to do the analysis both with a strict application of the coding rule derived from this procedure, and with ad hoc adjustments that reclassify the (hopefully few) cases that "escape." If the results differ significantly, then they are not robust to conceptual imprecision and this is potentially a problem.

Following this recipe may reveal that a concept X has two or more distinct meanings (or clusters of meanings) in ordinary language.⁸ Depending on the problem at hand, *at this point* it may make perfect sense for the researcher to stipulate that by X, he or she means this or that particular meaning. By explicating before stipulating, the researcher avoids the confusion likely to result if X is stipulated without prior explication, in which case the researcher may well claim too much or produce misleading "results" concerning the impact or explanation of X.

A central presumption in the recipe above is that if we know how to use a concept like "ethnic group," "power," or "rational" in sentences, and can make judgements about valid usage, then there must be some implicit principles that govern these judgements. A natural question arises: Who is the "we"? Isn't it true that the same word can have different meanings for different groups of speakers, even if they all ostensibly speak the same language? Isn't it clear that groups of social scientists evolve meanings for words particular to their own research community and tradition?⁹ Finally, returning to the "interpretivist objection" stated above, isn't it clear that an ordinary language analysis of the meaning of democracy may produce different results depending on whether the language in question is American English versus, say, Wolof or Swahili (see Schaffer 1998 for Wolof; or Scotton 1965 for Swahili)?

That the answer to each question is clearly "yes" neither lowers the value of explicating before stipulating nor undermines the project of testing hypotheses that refer to culturally diverse units of analysis. First, it is hard to see how one could uncover and make clear variation in the meaning of a concept across speech communities *except* by ordinary language analysis along the lines of the recipe given above. Among social scientists, the failure to do so probably accounts for a fair amount of fruitless disagreement among scholars working on similar problems but in different research traditions. Second, ordinary language analysis is an empirical method, a method of inducing and deducing the meaning of a concept from the data provided by usage. The "database" for this method is remarkably public and accessible. Empirical objections can be raised by any competent speaker of the language, for example, in the form of

comments such as “here is an example of a valid usage of the concept X that is not covered by your formulation of its meaning.” Thus, even if there is always the danger that *my* ordinary language analysis of X will produce results that do not fully reflect the meaning of X in *your* specific (possibly academic) speech community, my analysis is easily refuted, corrected, or deepened by means of counterexamples.¹⁰

On the interpretivist objection: It simply does not follow that if (say) democracy means something different in France than in the United States, it is invalid to ask about the cross-national determinants of democracy. It means only that the phenomenon studied under the heading “democracy” might differ depending on the nationality of the researcher. Suppose, contrary to fact, that in France “democracy” means a system of rule by cafe owners. This would not make an inquiry that asked why fair, contested elections are used to fill the highest political offices in some countries but not others any less possible or valuable. No doubt, the gloss on “democracy” in the last sentence does not reflect all that one might recover from an ordinary language analysis of “democracy” as used in English. But it probably does capture one of, or perhaps the main underlying principle that governs ordinary language attributions of “democracy” in both English and French.

For quantitative analyses, we should not ignore the underlying structure of conceptual understanding as reflected in ordinary language. Nor do we need to abandon quantitative analysis as an enterprise because the underlying structure of our concepts is complex and at times contradictory. Rather, ordinary language analysis is a useful tool for conceptual specification that can make quantitative analysis better able to address the questions that drive our own research programs. This case is best made by examples, to which we now turn.

III. Ethnic Groups

What is an ethnic group? The standard approach in the literatures of political science, anthropology (until recently), and sociology has been to list characteristics or beliefs that a set of people must share for them to constitute an “ethnic group.” Such definitions typically begin by saying either “An ethnic group is a group that has [list of characteristics],” or “By ‘ethnic group,’ I mean a group that has [list of characteristics].” Anthony Smith's definition (1986, chap. 2; Hutchinson and Smith 1996, 6) in terms of shared characteristics is one of the most frequently cited. He says that an “ethnic community” is

a named human population with myths of common ancestry, shared historical memories, one or more elements of common culture, a link with a homeland and a sense of solidarity.

Other lists might include a common language or common religion. For instance, Abner Cohen (1969, 4) says that, “an ethnic group is an informal interest group whose members are distinct from the members of other groups within the same society in that they share a measure of ... ‘compulsory institutions’ like kinship and religion, and can communicate among themselves relatively easily.”

Definitions in terms of shared traits or beliefs tend to stumble over the same problem. Interpreted literally, they imply the exclusion of cases that intuition says are ethnic groups, and the inclusion of cases that intuition would like to rule out.¹¹ For example, Jews may speak many different first languages, may or may not practice the Jewish religion, and do not all share a common culture. But nonetheless they are typically described as an ethnic group. Families share “a myth of common descent” but are not considered ethnic groups, and an arbitrarily selected set of professors might share a common culture, “compulsory institutions,” be of the same religion, and speak the same language but no one would call this set an ethnic group.

The problem arises even for Smith's relatively hedgeable definition, which has as one of its conditions “a link with a homeland.” This excludes Roma and can exclude other nomadic groups that intuition and usage say are “ethnic.” And it includes families and other intuitively wrong cases like Yankees or white Southerners in the United States. Perhaps Smith would say that for New Englanders the requisite “sense of solidarity” is not present, or that the “myths of common ancestry” are not strong enough. But how do we judge how much “solidarity” makes a group ethnic, or how strong the “myths of common ancestry” need to be? If the answer is “we know it when we see it,” then the definition is really serving as a front for unarticulated intuitions.¹²

It might be asked why it is such a problem if a few cases “get away,” or slip into a particular category unnoticed but undesired. The nature of the problem this poses depends on what assumptions one makes about the sort of things ethnic groups are. Smith and many others, for example, seem to be saying that ethnic groups exist objectively, “out there in the world,” akin to a natural category like gold, electrons, or perhaps rocks or trees. This is suggested by the claim that ethnic groups are “as old as the historical record” and “have been present in every period and continent and have played an important role in all societies” (Hutchinson and Smith 1996, 3). In this approach to definition -- call it “objectivist” -- the fundamental idea is that there exist objectively discernible things that are ethnic groups, and the problem is to identify the right list of essential characteristics that distinguish these things from others.

For analysts taking this approach, intuitively wrong inclusions and exclusions are a problem because they tend to undermine the claim that what we have here is a natural category, written into the world by Nature for us to read out. Faced with problem cases, there are three possible defenses. First, the analyst might respond by adding or subtracting shared characteristics from the definition, or otherwise fiddling with it to make the problem cases go away.¹³ But this is clearly *ad hoc*, and does not boost confidence in the claim that the category is natural or scientific. Second, one might respond that while there will always be a few exceptions, our intuition tells us what is what in the end. But then what use is the definition? We would do better to interrogate our intuitions.

Third, one might abandon or shelve the objectivist premise, and pose the definition instead as “When I say ‘ethnic group,’ I mean ... [list of shared characteristics or beliefs].” As discussed above, this approach, which follows that of Lewis Carroll's Humpty Dumpty, might be called *legislation*. Relying upon Smith's definition and

confronted with the case of the Roma, the legislator can only respond, "Well, for my analysis the Roma are just *not* an ethnic group," or "Your intuitions are confused, wrong, or unscientific; scientifically speaking, the Roma are in fact not an ethnic group." But then if the analyst reaches seemingly strong or counterintuitive conclusions, this may be because she has simply defined a key term in a peculiar way (that is, peculiar to the understanding implicit in popular usage). It *could* be that the analyst has discovered the most scientifically useful or productive way to define what an ethnic group is. But it also could be that our intuitions should be more carefully consulted, and this bit of legislation is ill-conceived.

If we look to everyday usage, it is quickly evident that efforts to define "ethnic group" in terms of common traits or shared beliefs and myths are unlikely to work. Consider the question of how we decide in practice whether an individual is a member of a particular ethnic group. In deciding a person's ethnicity, *we do not need to know anything* about his or her cultural habits, mother tongue, religion, or beliefs of any sort. Rather, we simply need to know about parentage. In ordinary usage, to ask "What is her ethnicity?" is to ask about what ethnic group her parents (or other close ancestors) were assigned to. If both parents were coded as members of the same ethnic group, then there is no ambiguity about the ethnicity of their children. In the case of mixed marriages, arbitrary (and political) conventions that may vary from place to place are employed.¹⁴ And likewise, all that is necessary to be counted as a member of an ethnic group is to be able to have accepted the claim to be immediately descended from other members of the group.

It is useful here to distinguish between two things that in general define and distinguish social categories like ethnicity: *membership rules*, which are the implicit or explicit rules that we use in deciding who counts as a member of a category, and the *content* of a social category, which are the qualities, attributes, or obligations typically associated with members of a category. The point is that qualities associated with the members of a category (and by which members may self-consciously differentiate themselves as a group from others) need have nothing to do with the criteria that decide membership. This is true of many social categories besides ethnicity (for example, many occupational roles), but ethnicity provides a strikingly clear instance. If an American WASP were to convert to Armenian Orthodoxy and adopt "typically Armenian" mannerisms, few would say that he was "ethnically Armenian." What matters, as we use the concept, is immediate descent and how one's parents or grandparents were coded. Similarly, many Americans who cannot perform a Jewish ritual and don't speak Hebrew still consider themselves and are considered by others as ethnically Jewish because that is the way their parents and grandparents were coded.

So we might then try to capture the sense of ordinary usage as follows: We call a group "ethnic" if it is larger than a family and membership in the group is decided by a descent rule. In fact, this simple and short definition covers a multitude of cases correctly, and is serviceable for much theoretical inquiry. The study of ethnic groups is essentially the study of groups larger than families that reckon membership by the ascriptive criterion of descent.

An immediate benefit of this approach is that it solves the puzzle of why the conflicts in Northern Ireland and Bosnia are often called "ethnic" rather than "religious," even though religion is the main cultural attribute or characteristic that differentiates individuals on both sides. We call these conflicts "ethnic" because the basis of group membership is not religious conviction but rather immediate descent. Converting to Orthodox Christianity would not make a Bosnian Muslim able to keep his house in Srebrenica. What Serbs care about in distinguishing enemies and victims is not religion, not culture or customs, but descent.

Likewise, the observation allows us to make sense of regional categories like "Southerner" or "Bavarian," which may meet Smith's and other definitions in terms of shared characteristics of individuals but are rejected as ethnic groups by intuition. Intuition rejects them as ethnic categories because -- or precisely to the extent that -- the membership rule is not based on descent. What makes one a member of a regional category are the requirements of having lived in the region for some period of time and of having adopted in some measure the cultural content associated with the region (such as dialect, dress, or cuisine).

To be true to ordinary usage, however, our first cut does not go deep enough. "Membership reckoned by descent" is probably necessary for us to count a group as "ethnic," but everyday speech draws further distinctions between various types of groups that all have ascriptive membership criteria. In addition to ethnic groups, we speak of races, castes, clans, tribes, and aristocracies, for instance.¹⁵ It is significant that usage may lump these together with "ethnic group," or have trouble with just what the distinctions are (especially for race and clan). Witness, for example, *The New York Times's* descriptions of the violence in Somalia, which began speaking of it principally as "clan" conflict, but more recently lists it as an example of post-Cold War "ethnic violence." Likewise, race and ethnicity may be either equated or distinguished in everyday discourse in the United States. And while many scholars of ethnic conflict distinguish between ethnic groups and castes, Donald Horowitz sees castes as an example of a "ranked" system of ethnic groups (1985, 53n164). These examples make two points. First, clans, castes, ethnic groups, and races bear a conceptual family resemblance to each other due to the common feature of ascriptive rules for deciding membership. Second, unambiguous intuitions about what counts as "ethnic" begin to break down when we push beyond this central feature.

Is there any logic at all to these distinctions? The distinction does not lie in the fact that ethnic groups are characterized by self-consciously different cultures, because the same can equally be said of British classes or Indian castes. Rather, for castes, aristocracies, and (some) classes, the difference would seem to be that in these cases the social category depends on, or has meaning, only by reference to a larger system of linked categories. That is, an aristocracy cannot exist -- conceptually -- unless commoners exist, and the same applies for classes and castes. By contrast, while it is an empirical fact that ethnic groups "understand themselves" through contrasts with other ethnic groups,¹⁶ the existence of an ethnic category does not depend conceptually on the existence of any particular other ethnic category. For instance, the ethnic category Sinhalese is *conceptually autonomous* in the sense that, hypothetically, there could be a

Sinhalese ethnic group regardless of how any other groups were constituted, but there cannot be Brahmans if there are no (or have never been) other castes who are placed socially in reference to Brahmans. The idea of a caste society with just one caste is conceptually incoherent, while the idea of an autonomous ethnic group is not.

This does not rule out the possibility of what Horowitz (1985, 21-36) calls “ranked ethnic groups,” such as Hutu and Tutsi in Rwanda and Burundi, where ethnic lines roughly coincide with a class division. We believe the logic embedded in current usage would hold these to be ethnic groups to the extent that Hutus and Tutsis can be imagined as autonomous societies – and indeed they became so in the late colonial period -- rather than categories constituted (conceptually rather than in practice) by their functional relationships.

So we obtain another condition that differentiates ethnic groups from castes, aristocracies, and some clans: We call a group “ethnic” if the membership is reckoned by descent and the members imagine their group as conceptually autonomous. That is, the idea or concept of the group does not depend on the existence of other groups.

There remains the problem of finding a sense, if there is any, to ordinary language distinctions between ethnic groups, races, tribes, and clans or other “subethnic” units.¹⁷ To some extent, we think in terms of set/subset relations here. Ordinary language will allow clans to be subsets of an ethnic group, and ethnic groups to be subsets of a race, but intuition rebels at the idea of a race that is a subset of an ethnic group, or an ethnic group that is a subset of a clan.¹⁸

Note that nothing in the working definition posed above precludes an individual from being a member of *more than one* group that reckons membership by descent and is conceptually autonomous. As the examples of race and clan show, this is not uncommon. Indeed, the fact of multiple ascriptive affiliations is at the root of the phenomenon known as “situational ethnicity” (Young 1976) -- the fact that individuals often define their ethnic affiliation differently depending on political and social circumstances. Multiple ascriptive affiliations make situational ethnicity possible, both conceptually and practically. Conceptually, the common element of a person's different “situational ethnicities” is the ascriptive rule of membership for each of the categories. This is what makes them situational ethnicities rather than just situational identities.

Further, such multiple ascriptive categories are often ordered in a set-inclusive manner. In these cases, we may find ourselves without strong intuitions about where the designation “ethnic group” should apply, or with strong intuitions that are hard to explain. For these cases, when there is an ordinary language decision on which level is “ethnicity” (e.g., Somalis rather than Somali clans), this is purely a matter of convention, and we think one will not find any rigorous conceptual logic to it.

Nonetheless, bloody struggles often occur over the question of *what the convention will be* in particular cases. The question of where in a hierarchy of ascriptive categories the designation “ethnic group” should fall can be the subject of intense political contestation. In a large part, this is because current global norms and practice

view ethnic groupness as conferring eligibility, under certain circumstances, for the privileged condition of nation-statehood. For example, the Turkish government and many citizens wish to classify Kurds as “mountain Turks,” against the wishes of many Kurds, and the Greek government and most citizens refuse to recognize any Macedonian ethnic group whatsoever (preferring the term, “Slavophone Greeks”).

To recapitulate, an analysis of what we count as an “ethnic group” in practice yields the following rough statement: In ordinary language, a group can be counted as “ethnic” if it is larger than a family, membership is reckoned by immediate descent, and it is conceptually autonomous. In addition, individuals may be coded as members of several such groups, in which case conventions decide which one is typically referred to as an ethnic group, and usage may allow several such designations depending on the context of discussion.¹⁹ In turn, political argument and contestation can decide the conventions in particular cases.

But this statement still falls short. There is more implicit structure in “ethnic group,” as shown by a final set of important counterexamples. Consider the following cases: (1) The United States (or almost any country) changes its citizenship rule so that only those born of U.S. citizens can be U.S. citizens. Are “U.S. citizens” now an ethnic group? Intuition, we think, says definitely not. (2) Suppose a group of right-wing militia types sets up camp in Idaho or Texas, declaring myths of common descent, reckoning future membership by a descent criterion, and arguing that they constitute an ethnic group. Are they one? Again intuition says No.

Why do we bridle at such examples? Note that it won't work to try to rule them out by introducing a criterion of shared culture, language, religion, memories, or, as indicated in (2), a shared belief in common descent. We could add any of these and it could still be intuitively wrong to describe such collections as ethnic groups. Instead, the problem with these cases has to do with the idea that an ethnic group can *legislate itself* into existence by a positive, deliberate act.

What these examples indicate, then, is that our intuitive notion of what counts as an ethnic group contains within it a normative model of how a proper “ethnic group” should come into being. Our usage supposes that real ethnic groups should have evolved or emerged *naturally*, in some sense or be of long-standing. The natural process need not be understood as closely analogous to natural processes in the physical world; it might be understood as an accretion of innumerable small social decisions, recognitions, and uses of an ethnic category term like Jew or English or Mongolian. But one cannot compel people to believe in a newly invented ethnic category in one fell swoop. A group is counted as “ethnic” only if it meets the previously stated conditions and, in addition, if it is accepted or granted that the ethnic category arose from some kind of natural (or better, naturalized) process.

What will intuition count as an acceptable natural process? Again, politics enters in force. Claims to ethnic groupness are typically backed by (often contested) claims to a “natural history” as a group. Empirically, the groups we count as “ethnic” can emerge in many different ways, ranging from decentralized, bottom-up processes of social

categorization to specific, top-down acts of legislation. For an example of the latter, note the Stalin-era separation of Kirghiz as a group distinct from Kazakh. Catastrophic historical events such as war and famine can have similar effects. Consider the emergence of a Palestinian ethnic group as distinct from the generic "Arabs" during and after the 1948 Arab-Israeli war. Whether we count a group as "ethnic" depends on a negotiation and agreement on the group's history and emergence as such. Perhaps the legislated Kirghiz or the regional-to-ethnic Palestinian categories did not seem "real" at first, but after time to acclimate and develop social relations conditioned on the category, "Kirghiz" and "Palestinian" meet the condition of emerging from naturalized process. Hypothetical cases like the militia group in Idaho fail because *at first*, at any rate, there is no way to naturalize their emergence.

To conclude: As the term is used, "ethnic group" refers to a group larger than a family for which membership is reckoned primarily by descent, is conceptually autonomous, and has a conventionally recognized "natural history" as a group. Our analysis shows the term to be intrinsically paradoxical, in that the idea that a group must be primordially given to be "ethnic" (i.e., the last condition) is built into its meaning, even though what is recognized as primordially given (and thus "ethnic") is clearly a matter of political negotiation and convention in many cases. Even so, to say that what is an ethnic group is partly a matter of convention is not to say that the conventions are arbitrary or that they follow no internal logic. We therefore ought not preclude coding groups on the basis of these conventions for the purpose of testing a hypothesis like that posed at the start of the paper.

IV. Ethnic violence

We now have a statement of the meaning and some of the normative presumptions of "ethnic group" as the term is currently used. What, then, is ethnic violence? This question is almost never posed in studies of ethnic violence, presumably because it is assumed to have a very straightforward answer. For instance, how about "ethnic violence is just violence between ethnic groups"? This natural first cut, however, hides crucial features of the concept by attributing agency and motivation to ethnic groups. The way the term "ethnic violence" is used in practice follows a more detailed and interesting logic, which can be elucidated by considering two counterexamples.

First, suppose a white American is mugged by an African American. We are reluctant to describe such incidents as "ethnic violence" even though in literal terms it is violence between members of two different ethnic groups.

Thus, whether we call violence "ethnic" depends on something about our assessment of the participants' motivations, intentions, or mental states. The problem is to say what this is. For a first cut, we might argue that a mugging is not "ethnic violence" in so far as the mugger is motivated by the desire to get money rather than by the desire to harm an ethnic other -- mugging is an "economic crime" rather than a "hate crime." But it isn't right to say that we term violence "ethnic" only if the motivation is hatred or dislike of ethnic others. This is sufficient but not necessary. For instance, if the members of one group attack those of another in hope of gaining access to land, jobs or other material

goods, we still typically call this ethnic violence. Or consider the case of a soldier in a separatist army who plants a bomb that kills a number of ethnic others. The soldier may not have been motivated by hatred or dislike of the other group, but even if this is so most would still call it ethnic violence. Finally, we can turn a mugging into "ethnic violence" as follows: Suppose one guy, who in fact wants to steal money, says to another, Hey let's go beat up a black, and his accomplice agrees simply because he is worried about looking like a wimp if he says no. We are much more likely to accept this as a case of "ethnic violence" than we would the simple mugging described above.

These examples suggest either that our usage is arbitrary and inconsistent or that we employ a more complex set of decision rules. We would argue for the latter. We term violence "ethnic" if it involves members of different ethnic groups and either

1. it is motivated by hatred or dislike of ethnic others in general;
2. the criterion for selecting victims is ethnicity, meaning that members of one's own group are exempted and members of the other group are eligible (as it were);
3. it is committed with the idea of being *on behalf of* or *in the name of* an ethnic group, or is committed against those who claim to represent or act on behalf of an ethnic group (and because of this status).

These rules cover the examples just given, and we think they work more generally (though see caveats and the second counterexample below). While any particular incident can satisfy more than one of these conditions, they pick out three classes of cases.

The first condition covers hate crimes, some cases of police brutality, many riots, and incidents between individuals where a "non-ethnic" dispute escalates to a higher level of violence because of interethnic dislike or animosity. The second condition covers, among other things, the case of a riot or attack on ethnic others with an economic motivation. Even if the main motive is to gain land or jobs, we call such events ethnic violence because of the selection criterion for victims -- one is "eligible" to be a victim only by virtue of one's ethnicity. The second condition can also cover acts of terrorism and government measures of population control and discipline that might be described as "ethnic violence," such as the internment of Japanese-Americans in the U.S. during World War II.

The third condition mainly covers cases of violence involving autonomy, separatist, and other organizations that claim (credibly or not) to act politically on behalf of an ethnic group.²⁰ The first clause in this condition -- "*on behalf of* or *in the name of* an ethnic group" -- covers cases of insurgents attacking agents of the state as self-appointed representatives of an ethnic group. Consider Sri Lanka, where since 1983 guerrilla armies based in the Jaffna Peninsula have militarily challenged the right of Sri Lanka to govern their section of the country. Much of the violence is perpetrated by young thugs, who may be engaged in thuggery for purposes of loot or salary, or perhaps because they were kidnapped by the guerrilla army and forced to fight in one of its bands. The leaders of this army, however, justify their violence as being perpetrated in the name of the Tamil

nation, with the goal of creating a Tamil state within a country now dominated by Sinhalese. We would claim that this is generally sufficient in standard usage for a case to be described as “ethnic violence.”

The second clause in the third condition -- “or committed against those who represent or act on behalf of an ethnic group” -- covers cases of state agents attacking members of mobilized ethnic groups, which are often described as “ethnic violence.” For example, the U.S. press may refer to British army attacks on I.R.A. agents as part of “ethnic violence,” and likewise Turkish army attacks on the P.K.K. But note that such designations are politically contested and contestable. The U.S. press would almost surely not refer to F.B.I. attacks on Black Panthers as instances of “ethnic violence,” even though this is structurally quite close to the I.R.A. and P.K.K. cases.

A reason for the ambiguity and political contestation over these cases is that the state's justification for such an attack is often *to put down or end* “ethnic violence,” and thus that the state's agents are acting on behalf of all citizens. From the vantage of the state, it cannot itself be engaging in ethnic violence. But since those attacked claim to represent the ethnic group, from their vantage it can be seen as violence against the group as such, thus meeting condition (2) and probably (1) as well. In particular contexts, then, to take a strong position on whether state violence against autonomy and separatist organizations is “ethnic violence” is to take a position in a political struggle. We don't wish to legislate arbitrarily about these cases, many of which are massacres by powerful states against entire populations. As in the case “ethnic group,” an ordinary language analysis is useful for bringing to light the political conventions and contests embedded in the concept's meaning.

Because motives and intentions are the private information of those who have them and are revealed imperfectly by actions, there can be room for argument and disagreement about whether specific incidents are cases of ethnic violence. (In the U.S., a recent example concerns the dispute over whether the C.I.A. deliberately sought to introduce crack cocaine into black neighborhoods in Los Angeles as a policy of “racial genocide.”) Brubaker and Laitin (1998) have suggested that this implies another condition for something to be “ethnic violence.” Namely, there are arbitrary social conventions about how different sorts of incidents should be coded, and these will determine what is or is not (counted as) ethnic violence. They note that there is an elite and media convention in the U.S. not to describe street crime by African Americans against whites as “racial” in nature, though (they suggest) it might be described otherwise. In northeastern Estonia, Russians burned down an Estonian secondary school; but elites on both sides of the ethnic divide, seeking to preserve peace, agreed that the violence was “mafia” related. Popular talk about the incident was on the horrors of the mafiosi, not of the potential for the spiraling of ethnic violence.

There is a valuable insight here about the degrees of latitude in the social coding of everyday violent incidents between members of different ethnic groups. In some places, like Tito's Yugoslavia, there was a deliberate state policy of publicly describing interethnic incidents and some nationalist dissent in as non-inflammatory a way as possible (a policy that ended with Milosevic's rise to power). But this analysis doesn't go

deep enough into the logic of existing social conventions for coding "ethnic violence." The media conventions Brubaker and Laitin refer to are about interpreting motivations, but not about what would count as ethnic violence given knowledge of motives and intentions. For example, if it were discovered that in a particular incident the major motive for the mugging was not money but to harm an ethnic other, then few would say that this was in truth *not* a case of ethnic or racial violence, even if it was reported otherwise in the press. We need to distinguish between the motives and intentions which may help to determine whether something is "ethnic violence" and the socially or politically generated rules for inferring or attributing motives from the observable evidence of actions and words.

Returning to the main argument, recall the simple but inadequate definition of ethnic violence as violence between members of different ethnic groups. Here is the second counterexample: By this definition, the Franco-Prussian war of 1870-71 would count as ethnic violence, as would almost every interstate war over the past two centuries (at least), even though usage virtually never counts interstate wars as "ethnic." Moreover, we would still hesitate to call interstate violence "ethnic" if one or more of the three conditions given above were met. For instance, even in a war where each state can be pictured, more or less, as the executive committee of an ethnic group, we are unlikely to use the term "ethnic violence" (Germany vs. Poland in 1939? Japan vs. China in the 30s?).

The example reveals an additional implicit rule we use in making attributions of "ethnic violence." It seems that violence is "ethnic" *only if at least one of the groups involved is not fully in possession of an internationally recognized state apparatus.*

Thus, our concept of "ethnic violence" appears to be premised on a particular conception of the organization and operation of the international system. In particular, whether something is coded as "ethnic violence" depends in part on the organizational means available to the participants and the international recognition accorded them. By convention, states may be party to ethnic violence when they deal with *substate* actors (though whether the states are *themselves* engaged in ethnic violence is contested). But when they engage in violence with each other, they are judged to be *above* the realm of ethnic violence. Ethnic groups that gain recognized title to a state apparatus (a sufficient condition for being called a "nation") seem to shed their ethnicity when acting "as a state" in the high realm of interstate politics.²¹ This convention reflects the general acceptance of a normative hierarchy of collective actors in world politics that puts states (or nation-states) at the top and distinguishes these actors *in kind* from others like ethnic groups. This is so even when the "promotion" to statehood -- international recognition -- is purely a matter of collective negotiation and agreement among states.²²

In sum, in ordinary language "ethnic violence" refers to violence meeting conditions 1-3 above and in which at least one protagonist is not a state. The analysis shows that the concept covers an extraordinary variety of events, implying, in turn, that an empirical inquiry like that suggested by the initial hypothesis ought prudently to pick out a subset of all possible events of "ethnic violence" for the dependent variable. By explicating before stipulating, one avoids mistaking partial measures of "ethnic violence"

for the whole concept, and one can situate particular forms of ethnic violence under study (for example, ethnic war) sensibly within the larger terrain of the concept.

V. Cultural Differences

The initial hypothesis demands a measure of cultural difference, which may seem a daunting task inasmuch as anthropologists have no general guidelines on how to specify “culture.” In ordinary language, however, the concept of cultural difference is not so mysterious, meaning something like “they do (and think about) things differently over there.” When we visit a foreign country or an ethnically distinct community in our own country, we are apt to remark that the culture we are witnessing is different from our own. The way people relate to one another, the way they act in public and in private, the perspectives on the world they tend to take, and the norms they uphold seem shared among them and distinct from others. In sum, culture in ordinary language refers to values, norms, beliefs, and practices that distinguish one group from another.

Further, the everyday notion of “cultural difference” carries with it a presumption that cultural beliefs and practices are related to each other in a nonarbitrary way. “That is so French!” expresses the belief that diverse practices and actions -- cuisine, propensity to build barricades, habits of watching professors discuss books on TV -- accord with a common logic. “Culture” in the term “cultural differences” might be taken to refer either to (1) the practices, beliefs and norms of a group, or perhaps their “sum”, or (2) the underlying system or logic presumed to tie these diverse practices and beliefs together. Popular understanding is partial to the second understanding, in light of our sense that the cultural traits of a group are related in a nonarbitrary, or even coherent, manner.

Specified either way, we face a difficult problem: “Cultural difference” seems resistant to measurement. Italians may feel comfortable speaking with their noses centimeters apart; Germans are more comfortable chatting with noses perhaps a meter apart. This is the kind of thing we notice in observing different cultures, but it would be hard to sustain the idea that nose distances in chatting are a good indicator of cultural difference, under either interpretation. If Iroquois chat with their noses at Italian-length distances apart, we would be reluctant to infer that Italian culture is closer to Iroquois than to German.

Language, however, is intuitively a more central aspect of culture than nose distance norms, since a language is often thought to encode a whole worldview. This is especially the case when referring to an ethnic group’s ancestral language. Even under conditions where only a few speak that language, people from a particular ethnic background retain manners of speech and perhaps styles of thought in their adoptive language that reflect norms and practices of their ancestral language. Also, even a vague memory of an ancestral language indicates a storehouse of cultural practices and beliefs that can be mined by political entrepreneurs seeking to mobilize support along cultural lines. In light of this intuition, we offer the following proposal: For a measure of the degree of “cultural difference” between two groups, use a measure of the structural distance between their ancestral languages.

This coding scheme makes many intuitively plausible assessments. For example, French and Italian are both Romance languages; German is not. A measure that relied on structural differences between historic languages would code cultural difference between the first two as smaller than between French and German or Italian and German, which arguably corresponds to the intuitions of travelers and observers of these cultures.²³ And the measure would code French, German, and Italian as far closer to each other than to the Iroquois, which again corresponds with intuition. These examples suggest the possibility of using the branching-off point between any two languages as an indicator of the cultural distance between the groups associated with them. In the specific measure we have developed, two groups are coded as more culturally similar the more linguistic family classifications they share in common (according to the comprehensive coding of language families in Grimes 1992).

Not surprisingly, this measure is imperfect. In the first place, for some ethnic groups there is no obvious choice for “ancestral language.” African-Americans have no one ancestral language because their ancestors came from diverse linguistic communities. The rule gives us no help in deciding whether the historic language of Jews is Hebrew, Yiddish, or, more plausibly, no language at all. Thus we might amend our rule such that a group’s language is the historic language associated with the group only if the group had a historic language associated with its name. Alternatively, in the case where there are multiple ancestral languages, one might code “cultural difference from dominant group” by looking at the mean difference between the dominant group’s language and the full set of ancestral languages of the minority.

Second, why use ancestral languages rather than languages currently spoken by the groups’ members? Coding on language currently used would have the virtue of potentially picking up cultural assimilation that reduces cultural differences. Should German be used to code the cultural difference between German-Americans and the dominant group in the United States, even though for several generations most German-Americans have been monolingual in English? Wouldn’t using Native Americans languages for Native Americans or African languages for African Americans overstate the degree of cultural difference between dominant group and minority, again due to assimilation reflected in current language use? Using ancestral language differences, Irish Catholics in Northern Ireland would be coded just as culturally distant from the dominant group in the U.K. as would recently arrived Pakistanis, since both speak Indo-European languages that diverge from English at the first branch after Indo-European.²⁴ Such examples raise the concern that using ancestral language would lead us to exaggerate cultural difference in cases where there has been significant assimilation over time.

On the other hand, using language currently used has its own problems – in our view even more serious. Using current language would lead to underestimating differences when a group’s first language has changed over time but many other practices and beliefs have not, or have changed but remained distinct (perhaps a common state of affairs). Coding Native Americans or African-Americans as culturally identical to the dominant group in the U.S., or Catholics in Northern Ireland as identical to the dominant English in the U.K, seems intuitively wrong as well. Second, there is the problem of

deciding how to code a particular group's current language. Do African-Americans speak standard American English or Ebonics? Coding the former might overestimate similarity with WASP culture; coding the latter might underestimate it. In many cases of minority groups, not all members speak the same first language. Many Kurds in Turkey speak Turkish as their first language; many others speak Kurdish as their primary language. While problems of this sort would arise for *any* measure of "cultural difference," they are especially acute in measuring current practice.

From a social scientific point of view, one might well reject the presumption implicit in ordinary usage that cultural practices and beliefs emerge from, or are defined by, an unobservable, underlying logic or system. Arguably, this is a theory to be tested and explored rather than taken for granted.²⁵ Taking this approach, a researcher trying to test our initial hypothesis might say something like this: Whatever "culture" is, people think language is an important aspect of it, so I am interested in testing whether differences in language are associated with higher or lower propensities for political conflict. This approach would probably incline the researcher towards coding language currently used.

Alternatively, the researcher might reject the presumption of cultural logics but argue either that the presence of a distinct ancestral language provides political entrepreneurs with an important resource for group mobilization, or that current language use often contains remnants of ancestral language speech patterns. In either case the better measure would be ancestral languages. We can see arguments for either measure depending on more specific versions of our initial hypothesis. However, in so far as ordinary language presumes in favor of "cultural differences" as manifestations of common underlying cultural logics, the measure of difference between ancestral languages probably better catches the intuition that stands behind the hypothesis that we laid out at the start of this paper. It is therefore the more defensible measure in testing that hypothesis.

VI. Dominant Group

The final concepts in the hypothesis posed at the start of the paper that require specification are the "minority" and the "dominant ethnic group" in the state. Here we will discuss only how to determine the dominant group, taking all other ethnic groups as "minorities."

The natural specification of "dominant," especially when juxtaposed to the "minority," is that of the predominant or majority group. This first-cut definition "works" for many cases, yielding (for instance) Estonians as dominant in Estonia, Han Chinese in China, and Russians in the old Soviet Union. But problems arise for cases where there is no majority group and for cases where the majority are dominated by a numerically weak but socially, politically, and economically powerful minority.

In cases where there is no clear majority, we tend to think that the dominant group is that group *controlling state power and access to economic welfare and advancement*. In many African countries no ethnic group constitutes a majority. In these countries

leaders who gain control of the state apparatus typically surround themselves with ethnic kin both for protection and to dispense patronage. Trade licenses, cabinet ministries, and land go disproportionately to those groups whose leaders control state power.

If the ethnicity of leadership of the country is the key to our intuitions about dominance, why do we need the first rule on predominance? The answer is that when there is a majority group in a country we often think of it as dominant even if its members do not control state power. In Canada, the past five Prime Ministers and a significant percentage of the state service are Quebecois, yet we refer to this situation as the minority challenging the dominant group through its greater solidarity, rather than seeing this situation as evidence that the Quebecois are the dominant ethnic group in Canada. Going back to our hypothetical, we could imagine Colin Powell becoming President of the United States and even appointing many African Americans to leading positions in the Cabinet and Judiciary. Suppose we coded the dominant ethnic group in the US as WASPs, and suppose they fell below a majority of the population. Yet even still, we would see Powell's presidency as an example of a minority group making inroads in WASP-dominant society rather than as evidence that African Americans had replaced WASPs as the dominant group. In these two cases, dominance reflects the predominance of members of an ethnic group in a society irrespective of whether they control the commanding heights of the polity.

More difficult cases are those of the Alawites in Syria and the Tutsis in Burundi. In both of these cases minority ethnic groups have long had a grip on power, and the overwhelmingly majority groups (about 85% Sunnis in Syria and Hutus in Burundi) have suffered from the control by a highly militarized minority. Here ordinary language gives us no intuition as to whether to think of the minority groups as the dominant ethnic/religious groups ruling over the majority or the majority groups as the dominant ethnic groups yet under the control of a powerful minority.

Even in heterogeneous states with no majority ethnic group, in some cases it would go against intuitions to ascribe dominance to the ethnic group of the leader. Consider Nigeria. Power went originally to the Hausas who constitute just under a majority of the country. Out of power due to a coup that put an Igbo officer as head of state, military leaders, mostly Hausas from the North, plotted a counter-coup in July 1966, and this led to the accession to power of Yakubu Gowon, a northerner, but a Christian from a tiny ethnic group, the Angas (who constitute 100,000 out of Nigeria's population of over 100 million). It would be counter-intuitive and against popular understanding to say that dominance moved from the Igbos to the Angas. Most observers considered that Gowon's incumbency was made possible by virtue of the fact that he could not bestow (many) benefits on his own group, and had to represent the interests of the Hausas to whom his rule owed its longevity.

This example suggests an amendment to the rule that focuses on the leadership of the country as an indicator of ethnic dominance. If the leader of the country is from a group that represents a tiny fraction of the population (in our coding, less than five percent of the population), the dominant group becomes the plurality group of the country, but only if the plurality represents a group with a substantial percentage of the

population, perhaps 20 percent. So for Nigeria, the dominant group in the Gowon era becomes Hausa.

What about Tanzania? The largest ethnic group (the Sukuma) constitute only twelve percent of the population, hardly a substantial percentage. Ninety-five percent of the population are fluent Swahili speakers but only five percent are classified ethnically as Swahili. Suppose none of Tanzania's leaders wanted to or tried to divert significant resources to his ethnic group. Then it would be natural to say that Tanzania has no dominant ethnic group. This suggests yet another amendment to our rule, namely that if there is no majority group, no leadership group with more than a small percentage of the population, and no plurality group that has a substantial percentage of the population, then we would say that the country has no dominant ethnic group.

There will still be cases (beyond the earlier examples of Alawites and Tutsis) where our coding rules are in some tension with intuitions. In the Zaire of President Mobutu, his ethnic group, the Ngbandi hover around 5 percent of the population, yet he bestowed such largesse on his group, and dominated so completely over the plurality Bakongo, that it seems wrong to code Bakongo as the dominant group during Mobutu's rule. This case has a similar character to that of the Alawites and Tutsis.

Despite cases that are ambiguous and coding rules not perfectly in accord with our own intuitions, our rules get most cases "correct". To summarize, we first look to see if a country has an ethnic group that constitutes the majority of the population. If it is, we would code that group as the dominant ethnic group in the country. If there is no majority group, we consider the ethnic group of the country's political leader to be the dominant ethnic group. If the country's leader comes from a tiny ethnic group himself, then the dominant ethnic group is the plurality group. If the country's leader comes from a tiny ethnic group and the plurality group is also very small as a percentage of the country's population, we would code that country as not having a dominant ethnic group.

VII. Conclusion

Standard practice in political science and social science more generally regards the everyday meaning of key concepts as inherently vague, confused, and thus unfit for Science. From this view some deny the possibility of social science in any exacting sense. Others respond by arguing that social scientists should stipulate or legislate precise definitions in an effort to override and dispel the fuzziness of ordinary language. We have argued against both responses. We advocate the explication of the ordinary language meanings of such terms as "democracy," "ethnic group," or "power" *before* stipulating. The project of testing a hypothesis is undermined if it is not clear how the researcher's concepts map onto the ordinary language terms in which the hypothesis is originally expressed. The problem becomes acute when researchers begin to give real-world examples of their statistical findings and inevitably slip into ordinary language meanings of the variables. With such slippage, the narratives and the statistical models are incommensurate. This plea against early stipulation is not to undermine cross-sectional statistical analysis. By way of examples we hope to have shown that the intuitions that govern ordinary language meanings of important social science concepts

are not nearly as hopeless as they seem, and may contain very rich and interesting implicit structure. Elucidating this structure can greatly enrich subsequent empirical or theoretical analysis and is worthwhile in its own right.

Is ethnic violence more likely between more culturally distinct ethnic groups? Though not the focus of our paper, we owe the reader at least a hint of what one finds if one carries out the project suggested above. Ted Gurr and his collaborators collected data on 278 “minorities at risk” in 115 countries that, with a few exceptions, readily qualify as “ethnic groups” by our analysis. They also include an 8-point scale measuring levels of minority rebellion against the state apparatus, certainly one important dimension in the multifaceted concept of ethnic violence as explicated above. Using the measure of cultural difference proposed in section 5, we find its bivariate correlation with the maximum rebellion score for each group from 1980 on to be *negative* at $-.15$ (significant at $p = .01$).²⁶ In a regression that controls also for country per capita income, the estimated coefficient on the measure of cultural difference remains negative though it is statistically indistinguishable from zero. This finding certainly does not rule out any relationship between degree of cultural differences and ethnic violence. However, it does suggest that if any relationship does exist it is contingent on factors that remain to be identified. The simple and highly influential hypothesis stated at the outset of the paper is not supported.

References

- Austin, J. L.. 1962. How to do things with words. Cambridge: Harvard University Press.
- Barth, Fredrik, ed. 1969. Ethnic Groups and Boundaries. Boston: Little Brown.
- Brass, Paul. 1997. Theft on an Idol. Princeton: Princeton University Press.
- Brubaker, Rogers and David Laitin. 1998. Ethnic and Nationalist Violence. Annual Review of Sociology 24: 423-52.
- Cohen, Abner. 1969. Custom and Politics in Urban Africa. Berkeley: University of California Press.
- Collier, David and James Mahon. 1993. Conceptual "Stretching" Revisited: Adapting Categories in Comparative Analysis. American Political Science Review, 87, 4: 845-855.
- Deutsch, Karl. 1953. Nationalism and Social Communication. Cambridge: MIT Press.
- Fearon, James D. 1999. What is Identity (as We Now Use the Word)? Mimeo, Stanford University.
- Gellner, Ernest. 1983. Nations and Nationalism. Oxford: Blackwell.
- Gellner, Ernest. 1994. Encounters with Nationalism. Oxford: Blackwell.
- Grimes, Barbara ed. 1992. Ethnologue: Languages of the World 13th edition. Dallas, Summer Institute of Linguistics.
- Gurr, Ted Robert. 1993a. Minorities at Risk: A Global View of Ethnopolitical Conflict. Washington, DC: United States Institute of Peace Press.
- Hechter, Michael. 1975. Internal Colonialism. Berkeley: University of California Press.
- Horowitz, Donald. 1985. Ethnic Groups in Conflict. Berkeley: University of California Press.
- Huntington, Samuel. 1996. The Clash of Civilizations. New York: Simon and Schuster
- Hutchinson, John and Anthony D. Smith, eds. 1996. Ethnicity. Oxford: Oxford University Press.

- Laitin, David. 1986. *Hegemony and Culture*. Chicago: University of Chicago Press.
- Laitin, David. 1998. *Identity in Formation*. Ithaca: Cornell University Press.
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press.
- Pitkin, Hanna 1978. *The Concept of Representation*. Berkeley: University of California Press.
- Schaffer, Frederic C. 1998. *Democracy in Translation*. Ithaca: Cornell University Press.
- Scotton, Carol. 1965. "Some Swahili Political Words" *Journal of Modern African Studies*. 3, 4:
- Smith, Anthony D. 1986. *The Ethnic Origins of Nations*. Oxford: Blackwell.
- Smith, Anthony D. 1999. *Myths and Memories of the Nation*. Oxford: Oxford University Press.
- Young, M. Crawford 1976. *The Politics of Cultural Pluralism*. Madison: University of Wisconsin.

ENDNOTES

¹ Huntington 1996.

² See Gellner 1983, Deutsch 1953, Anderson 1983, Hechter 1975.

³ As Humpty Dumpty said, "When *I* use a word, it means just what I choose it to mean – neither more nor less."

⁴ Our position is in the spirit of Collier and Mahon 1993. In that paper, they assume that researchers know the underlying structure of concepts. For example, their requirement that researchers avoid the "inappropriate rejection" of categories implies that we already have intuitions of appropriateness. Here we seek to develop a procedure for honing those intuitions.

⁵ Legislating definitions may make sense for things or concepts that are being named for the first time, such as names of species, or which have only a rather technical sense, such as "international regimes" or "proportional representation." But pure neologisms are rare in social science, especially among its central concepts.

⁶ Austin's ordinary language analyses (1962) remain the most subtle demonstrations of this point; see also Pitkin 1978 on the concept of representation.

⁷ For another example, see Fearon 1999 on "identity."

⁸ For an example, see Fearon 1999 on "identity." Pitkin 1978 elucidates two clusters of meaning for representation.

⁹ For instance, what rational choice theorists mean by "rationality" is typically narrower than the everyday notion.

¹⁰ Nexis-Lexis and similar search engines provide a valuable "high tech" way to test whether the results of one's ordinary language analysis in fact capture contemporary usage. See, for example, Laitin 1998 on "identity".

¹¹ See Collier and Mahon 1993 for a general discussion of this problem for political science categories; see also Lakoff 1987. The same problem of embarrassing inclusions and exclusions arises in attempts to define "nations" in terms of shared characteristics.

¹² Making "a sense of solidarity" a condition for a group to be ethnic has other bad consequences. For example, this means that subsequent statements to the effect that "ethnic ties are very powerful," are now true by definition rather than being empirical claims.

¹³ This is hard to do without generating new problem cases.

¹⁴ Until recently, the (racist) convention in U.S. has been essentially that "one drop of black blood" makes a person "black." The discussion of a "multiracial category" on the census suggests that this rule may be entering a period of open dispute.

¹⁵ British classes are another example.

¹⁶ Since Barth 1969, anthropologists have shown repeatedly that how people think about their ethnic group depends on drawing boundaries with respect to other groups.

¹⁷ For clans whose conceptual existence depends on a particular clan system, the above remarks apply. According to Laitin, however, Somali clans, at least, are conceptually autonomous in the sense described.

¹⁸ Will usage allow an ethnic group whose members belonged to two different races, or a clan whose members belonged to two different ethnic groups? Hispanics in the U.S. are sometimes described as an ethnic group, but "can be of any race." We know of no cases of ethnically heterogeneous clans.

¹⁹ As with Somali clans. Other examples come from Pakistan and India, where “ethnic conflict” may be used to refer to either Hindu/Muslim or (in Pakistan) Punjabi/Pushtun/Sindi/Baluchi disputes.

²⁰ Also covered here are cases of individuals who take it on themselves to commit violence “for” the group, independent of any organizational ties.

²¹ Put differently, we make an implicit distinction between acting on behalf of a recognized state that is controlled by or identified with an ethnic group, and acting on behalf of an ethnic group that does not control a state.

²² In a partial exception to this rule of usage, there are cases where we may call violence prosecuted by states on both sides “ethnic,” such as the conflict between Armenia and Azerbaijan, arguably the Serb-Croat war of 1991-92, and the conflict between India and Pakistan over Kashmir (were it to heat up to interstate hostilities). But these exceptions really prove the rule, since in each case the interstate conflict evolved out of a conflict in which one side was an ethnic group that did not possess a recognized state. These are cases where the conflicts emerged out of fighting between non-state actors (such as paramilitaries) and thus “from the ground up” rather than from diplomacy between capital cities, in the manner paradigmatic for interstate conflict.

²³ Keep in mind the specific project here: to evaluate the cross-cultural hypothesis about ethnic conflict posed at the start of the paper. The measure of cultural difference proposed here is obviously crude; it would be very little help if one’s question were “How are the French different from the Germans?” for example.

²⁴ Gaelic is Indo-European, Celtic, while Urdu is Indo-European, Indo-Iranian; see Grimes 1992.

²⁵ Indeed, Laitin (1986) tested this notion of congruence and found it wanting.

²⁶ The specific measure of cultural distance here is one divided by one plus the number of common language families as coded from Grimes (1992); other functional forms we have tried are equally unrelated. Note that to code cultural distance we needed to code “dominant group” ourselves, since it is not specified in the MAR data even though many variables depend implicitly on a comparison of the minority to a “dominant group.”