

# Is Ethnicity Identifiable? Lessons from an Experiment in South Africa

Adam Harris and Michael Findley

August 20, 2010

Word Count: 10,500

## **Abstract**

Ethnicity is frequently posited as an important factor in civil violence and other political contexts. Despite the significant attention that ethnicity receives, its effect depends on an important, but mostly ignored, assumption: that ethnicity is identifiable within and across groups. There is likely considerable variation in peoples' abilities to identify each other. Certain individuals within groups might be better at identifying others' ethnicities; further, different types of information might aid identification better. We contend in this paper that the strength of an individual's ethnic identity influences her ability to identify others correctly. We consider this argument using an experiment in the Eastern Cape of South Africa in which individuals attempted to identify members of all of the major black ethnic groups. We find that the average individual struggles to identify ethnicity correctly in many conditions. Individuals with a stronger ethnic identity, however, are better at correctly identifying the ethnicity of others relative to the average individual. When we allow deceptive information, individuals with stronger identities were deceived more easily than others. These results have implications for a diverse set of studies relying on the identifiability assumption.

Acknowledgments: We thank James Habyarimana, Macartan Humphreys, Daniel Posner, and Jeremy Weinstein for sharing information about their research including pre- and post-study survey questions and Jamie Druckman, James Kuklinski, and Dan Nielson for helpful comments on this manuscript.

*“One morning, while we were eating before going to mass, they closed the windows and the gates. Then some boys from another school came in the dining hall and circled the tables. I was trembling. The boys shouted, ‘Get up, Tutsis. All the Tutsis stand up.’ There was a boy from my hill at home. We went to primary school together, and he said, ‘You Odette, you sit down, we know you’ve been a Hutu forever.’ Then some other boy came and pulled my hair and said, ‘With this hair, we know you’re a Tutsi.’” (Gourevitch 1998, 66-67).*

## **1 Introduction**

Ethnicity was a central part of the Rwanda genocide. Hutu extremists instigated the genocide and targeted Tutsis along with moderate Hutus. But Hutu killers often could not discern Hutu from Tutsi unless they viewed officially-issued identity cards or obtained other additional information. Even then, some of the best efforts to identify Tutsis failed. The complexity of ethnic identity in Rwanda is not unique; the world is replete with conflicts, such as Kosovo, Chechnya, and Kenya, in which ethnicity plays an important, but often complicated, role. In this paper, we explore this more systematically asking: How well can people discern ethnicity? And what conditions enable or hinder accurate ethnic identification?

Studies of the identifiability of Jews following World War II uncovered just how complicated ethnic identification can be (Rice and Mullen 2003). Despite these early studies, the difficult-to-test assumption – that ethnic divisions and classifications are real and often more easily recognized than other categories – has thus far been generally accepted (Horowitz 1985: 45-47; Chandra 2004, 2006a; Habyarimana et al 2009). More recently, evidence from a study in Uganda corroborates earlier studies and shows that the identifiability assumption does not always hold (Habyarimana et al: 2009). In their study, Ugandans could not easily identify each other and, under some limited circumstances, obtaining more information about others even made correct identification more difficult. They focus on the ability to identify another’s ethnic identity correctly given information primarily about the person being identified. This approach captures an important part of a person’s ability to correctly

identify others; it also leaves open avenues for future research. Among them, understanding better the variation in individuals' abilities to identify others *based on the characteristics of the identifier* is important.

Drawing on social identity theory (Tajfel and Turner 1979; Blascovich 1997), we contend that as individuals identify more strongly with their own group, they not only develop stronger antipathies for other groups, but they also develop a better ability to identify members of out-groups, as well as their in-group. It follows that individuals who identify strongly with their own ethnic group should be able to identify their own group well and also be better at identifying the ethnicity of other individuals. Thus, although identifiability might be difficult on average, we explore the possibility that identifier characteristics, such as a strong identity, facilitate the identification of others. We measured the strength of ethnic identity for identifiers and estimated its effect on successful identification of others' identities.

We conducted a randomized experiment in a large township in the Eastern Cape of South Africa to evaluate our hypotheses about the impact of identity strength. Although our primary focus is on the identity strength of the individual attempting to identify others, we also examine information about the person being identified. We consider whether language, name, geographical heritage, and ethnic symbols have an impact on an identifier's ability to identify others correctly. These different types of information – shown in photographs and short video clips – are important because each one could communicate important features of a person's ethnic identity. They could also be used as cues for someone to differentiate among members of ethnic groups during times of ethnic conflict or other ethnic interactions.

Some of the results of our experiment are consistent with past research, but we also offer novel insights. First, we find that individuals are not able to identify each other very easily. Correct identification rates are, on average, even lower than those found in Habyarimana et al (2007; 2009). Given the low baseline, we then explore what causes variation in identifiability. Consistent with social identity theory, individuals with a stronger attachment to their own group appear better at identifying members of other groups in a number of experimental conditions, which is consistent with our hypothesis and with past research on Jews in which only photographs were used (e.g., Allport and Kramer 1946). Finally, we find that individuals with a strong identity are more easily deceived by other individuals who lie about their identity, leaving open the possibility that a very strong attachment to one's group does not make ethnic targeting foolproof.

The experiment applies particularly to homogenous areas within countries, as the setting of this study consisted primarily of Xhosa identifiers in the Eastern Cape. South African townships are generally quite homogenous, making this study applicable to South Africa. Caution should be exercised in generalizing the results to more diverse areas and future research could examine identifiability in such contexts, similar to Habyarimana et al (2009). More generally, the results do suggest possible questions about one of the core assumptions in the ethnic conflict literature. If ethnicity is not easily identifiable, then theoretical and empirical analyses should account better for the uncertainty in individual decision-making during episodes of violent conflict. Existing literature should also account for variation on characteristics such as the strength of ethnic identity that condition individual behavior and lead to different outcomes.<sup>1</sup>

In what follows, we first present literature on ethnic identifiability, and ethnic strength, following which we propose theoretical connections between the strength of ethnic identity and ethnic identifiability. We then outline the experimental model and discuss the results showing that

ethnic strength plays an important, but not always consistent, role in correct identification. We then conclude by discussing the implications that these findings have for the literature on ethnicity.

## **2 Literature Review**

Ethnicity could have an impact in many political, social, and economic settings; moreover, it is a principal component of civil violence worldwide. And yet, attempts to understand whether and how ethnicity matters have proven complicated. For one, the concept of ethnicity is complex and heavily debated across several disciplines, leading to significant confusion about what ethnicity is as well as about how it should matter. Even where there is a consensus that ethnicity matters, the characteristics that are important in a given context are not well understood.

One of the most accepted conceptualizations of ethnicity is that it is based on a “myth of a collective ancestry”. By this understanding, ethnicity includes many individual and group attributes, such as skin color and culture, as well as group-based categories such as tribes and castes (Horowitz 1985: 52-53). Most of the ethnic conflict literature accepts Horowitz’s general conceptualization without being more specific about how the definition applies under different circumstances.

Because ethnicity is not usually defined precisely (e.g., Alesina et al 2003), measures have been equally vague and unclear. Quantitative studies of civil violence employ various measures capturing different elements of ethnicity: ethnic diversity (e.g., Fearon and Laitin 2003), ethnic group distinctness (Kirschner 2007), geographic concentration of ethnic groups (Toft 2005), and ethnic polarization (Alesina et al 2003). These measures capture only isolated aspects of ethnicity (Cederman and Girardin 2007; Chandra and Wilkinson 2008; Posner 2004).

Recently, some scholars have called for greater caution in conceptualizing and measuring ethnicity (Posner 2004; Chandra 2006b; Chandra and Wilkinson 2008). Offering a refinement of the concept of ethnicity, Chandra (2006b) argues that the defining characteristic of ethnicity is that it captures attributes associated, or believed to be associated, with *descent*, such as skin color, hair type, physical features, name, language, birth place, and surname. While this allows for a variety of attributes to characterize ethnicity, it offers a more precise umbrella to categorize ethnicity and separates it from other more abstract concepts such as culture.

Chandra's conceptualization is insightful, and begs the critical question: is ethnicity identifiable? Ethnic similarities or differences might not be relevant if they are not readily identifiable. We concur with Chandra (2006b) that ethnicity represents descent-based attributes, but we focus on exploring the conditions under which descent-based attributes are actually identifiable, as some have argued (Horowitz 1985; Chandra 2004).

Very few studies directly question the assumption that ethnicity is identifiable.<sup>2</sup> Horowitz (1985) contends that individuals can identify each other with relative ease, which potentially explains the mostly uncritical acceptance of the identifiability assumption. In discussing caste distinctions in India, he notes that "caste origins could easily be detected without a visible 'rank sign'" and continues: "This, in fact, is generally true" (45). Chandra (2004) indicates that more costless information should be available about ethnic identity, such as name, physical features, speech, and dress, than about nonethnic identities.

Clearly, Horowitz (1985) and Chandra (2004; 2006a) do not believe that ethnic identification is always easy. Although they cite cases in which identification occurred fairly easily, he discusses others in which it was more difficult (47-49). Horowitz (2001: 129) also suggests that identification

of prospective victims can be easy, but notes that in many cases, “it takes some effort to distinguish potential victims from members of other groups...” He suggests that potential perpetrators of violence are especially likely to attempt to distinguish among groups, if the costs of making a mistake are very high. Chandra (2004: 42) notes that “The multiple sources of costless data about an individual’s ethnic memberships mean that an observer can typically guess an individual’s ethnic identity on the basis of a relatively superficial interaction, even though such a guess may turn out to be erroneous.”

The prominent literature on ethnic security dilemmas (Posen 1993; Saideman et al 2002), furthermore, contends that in the presence of “anarchy”, individuals seek security within ethnic groupings, which in turn leads to ethnic targeting. Perhaps because the ability to identify others is a critical assumption, Sambanis and Schulhofer-Wohl (2008) call for a closer look at the identifiability assumption. Theories and research in the areas of in-group sanctioning (Fearon and Laitin 1996), assimilation (Laitin 1995), and pork spending (Fearon 1999), furthermore, all depend on the assumption that identification is possible. These theories assume that identification occurs so that individuals can reward or punish others appropriately or maintain their group boundaries.

Systematic research on identifiability is not completely absent from scholarly debate, although it has been confined primarily to the psychology literature. Following World War II, scholars studied the identifiability of Jews using photographs for identification (Allport and Kramer 1946; Savitz and Thomasson 1959; see Rice and Mullen 2003 for a review) and demonstrated the difficulties of accurate ethnic identification. More recently, a study conducted in Uganda examined whether individuals can identify each other given different information, such as name, language, heritage, and cultural information (Habyarimana et al, 2007). They found that individuals were not highly successful at identifying each other and that there was variation across individuals and groups.

We build on and extend earlier literature by examining how the *strength of one's ethnic identity* affects her ability to correctly identify others. In general, scholars have argued that identity strength is a key dimension of interethnic relations (Gibson and Gouws 2003) and there is some evidence that it affects ethnic identifiability (Allport and Kramer 1946; Dorfman, Keeve, & Saslow 1971). We now consider this possibility more closely.

### 3 Theory

Social identity theory posits that individuals identify more strongly with their own ethnic group when they have a more salient connection to it as well as feelings of superiority towards another group (Blascovich et al 1997; Tajfel and Turner 1979, 1986; Tajfel 1981; Horowitz 1985, 144).<sup>3</sup> A person who has a stronger ethnic identity highlights the positive in her own group relative to the negative in other groups (Tajfel and Turner 1986). By differentiating, she emphasizes her group's uniqueness, which leads to out-group prejudices and possibly hatred (Blascovich et al 1997).<sup>4</sup>

Put differently, strong identification with one's own group is frequently associated with pride in one's ethnicity and prejudice towards others (Blascovich et al 1997, 1364; de Figueiredo and Elkins 2003). Tajfel (1981) argues that the stronger one associates herself with a group, the more out-group antipathy she will hold. Therefore, a person with a stronger ethnic identity has greater pride in her ethnic group, because it provides a feeling of separateness. This, in turn, contributes to more antipathy towards other ethnic groups.

We contend that in-group pride and out-group prejudice motivate a stronger ethnic<sup>5</sup> to know more about the characteristics of ethnic groups because she will want to differentiate between her group and other groups. Ties to a group not only make people more aware of identifiable characteristics of other groups as a whole, but also of characteristics of *individual members* of other



groups. It is one thing, for example, for a Hutu to believe that, on average, Tutsis are taller, lighter skinned, and have narrower noses than Hutus, but it is quite another for a Hutu to meet a Tutsi and correctly identify whether he or she is Hutu or Tutsi. Identification is complicated due to high intermarriage rates and because ethnic attributes do not reduce to dichotomous categories – skin color, height, and other descent-based attributes are a matter of degree.

A stronger ethnic has greater knowledge of individuals belonging to various ethnic groups because she takes the time to learn the differences between members of her own ethnic group and others. Someone with greater pride in her ethnic group does not want to confuse outsiders as insiders because that would “defile” what she perceives as the uniqueness of her ethnic group (Blascovich et al 1997). Pride in one’s ethnic group comes from seeing it as distinct and superior to other groups, so it would be defiling to include inferior individuals into the superior ethnicity. In short, in-group pride can motivate the stronger ethnic to identify those they meet correctly because of the feelings of superiority that accompany their ethnic pride.

An individual with a weaker ethnic identity has a more difficult time differentiating between individuals within various ethnic groups because she does not have the same depth of knowledge. The weaker ethnic will not be as concerned with differentiating between her own group and others because she does not feel the same level of pride in her group and, therefore, is not as concerned about “defiling” it by including those that do not belong (Blascovich et al 1997). Thus, we hypothesize:

***Hypothesis 1:*** *A stronger ethnic is likely to be more successful than a weaker ethnic at correctly identifying others.*<sup>6</sup>

This hypothesis offers a somewhat generic expectation about identifiability that follows social identity theory closely. Ideally, we would like to understand the determinants of successful

identifiability in more specific situations. In reality, individuals have different types and amounts of information that they use to identify each other. Individuals might use available information about names, language, and accents to identify other people. Moreover, individuals might make a judgment about ethnic identity based on stated information about an identifyee's ethnicity, parents' ethnicity, modes of dress, and traditional foods. Individuals need to use available information to place others in an appropriate ethnic category.

Processing available information is challenging. Language can be a deceptive sign, because a person could speak her own language *or* the language she expects the other person to speak. Although one could change languages, accent and proficiency in a language also communicate important information. A stronger ethnic is likely able to discern whether someone is speaking her language as a second language, but may not be able to know which language is the person's first language. Other important pieces of information are an individual's first name and surname. Because the stronger ethnic has a greater knowledge of what differentiates ethnic groups, she will be better able to differentiate which name belongs to which language and group.

Individuals might also make a decision about ethnic identity based on more information. An individual might explicitly identify her own ethnicity, for example, and offer supporting evidence that confirms the stated ethnicity. When a person tells a stronger ethnic what her ethnic identity is, the stronger ethnic can look for cues that demonstrate the person is telling the truth. Because the stronger ethnic is aware of the differentiating qualities between ethnic groups, she can use those cues as evidence, whereas the weaker ethnic might not be able to match the information to relevant ethnic categories. Accordingly, we hypothesize:

*Hypothesis 2: A stronger ethnic is likely to be more successful than a weaker ethnic at correctly identifying others given more true information about a person's ethnic identity.*

In some cases, individuals might deliberately provide misinformation to the identifier in order to pass as a member of another ethnic group. During the Rwandan genocide, for example, some Tutsis attempted to pass as Hutus to avoid being killed. Similar incidents occur in many episodes of civil violence. The same types of information discussed above might be offered, but in reference to a different group. A stronger ethnic can look for the differentiating cues to decide whether a person belongs to the group to which she purports to belong. A person trying to pass as a member of another ethnicity needs to send the “correct” information in order to deceive the identifier. Because a stronger ethnic has a greater amount of ethnic knowledge relative to a weaker ethnic, one might expect that a stronger ethnic is less likely to be deceived.

On the other hand, strong ethnics are more likely to know the stereotypical cues to rely on as general rules of thumb; people that do not know the stereotypes should not be able to use any of the information provided them in a reliable way. If an individual sends the stereotypical cues to a stronger ethnic (who knows the stereotypes), the stronger ethnic will likely know just enough to recognize that the cues belong to a certain group, but not enough to understand whether the person is lying with the cues. Thus, a stronger ethnic that sees or hears the stereotypical cues will likely be fooled more easily. In this instance, the relative ethnic strength that the stronger ethnic has actually works against him/her when attempting to identify others. A Tswana individual, for example, might more easily fool a strong ethnic Xhosa into believing she is actually Sotho if the Tswana individual provides enough information about Sothos and the strong ethnic Xhosa identifies those cues with Sothos. A weaker ethnic Xhosa will likely not know one way or the other whether the individual is

Sotho, because she does not know the stereotypical cues, and thus be more likely to guess as if at random. Thus, we hypothesize:

***Hypothesis 3:** A stronger ethnic is likely to be less successful than a weaker ethnic at identifying those that lie about their ethnic identity.*

We test these hypotheses using an experiment that we detail in the next section.

## **4 Experimental Design**

We conducted a randomized experiment in a large township that we call Mayibuya in the Eastern Cape Province of South Africa.<sup>7</sup> Mayibuya is an exclusively black area that was created under the apartheid regime. The township suffers from poverty and unemployment: just over half of the population lives on an annual income of less than \$1,400 and 62% are unemployed (South Africa Statistics 2001).

The experiment consisted of presenting a photo followed by a randomly selected video of multiple people (hereafter ethnic “representers”) from the nine major black ethnic groups to individuals recruited in Mayibuya (hereafter “subjects”). The subjects attempted to identify the ethnicity of the representers. For each photo and video, the subject made a guess about the representer’s ethnicity based on the 2001 census’s list of the nine major black African groups.<sup>8</sup>

### *4.1 Ethnic Representers*

The *representers* were recruited based on their ethnic identity, age, and gender to ensure variation and proportionality. From May to June 2008, we recruited representers from throughout South Africa. The recruitment efforts yielded a sample of 76 representers that was proportional to the *national* demography of the black ethnic groups. That is, the percent Xhosa in the sample closely

matched the percent Xhosa in the population, the percent Zulu in the sample matched the percent Zulu in the population, and so forth.<sup>9</sup> The sample was less representative of the geographic regions of South Africa: of the 76 ethnic representers, 46 live in the Eastern Cape Province (60%), 22 in Gauteng (29%), 5 in Limpopo (7%), 2 in Northwest (3%), and 1 in KwaZulu Natal (1%). The ethnic representers, therefore, came from five of the nine provinces, but a few were born and/or raised in Mpumalanga and Free State Provinces; therefore, the representer sample has some representation from seven of the nine provinces.

We obtained background and demographic information from the ethnic representers and then took a bust picture and recorded 10 *separate* videos designed to capture various components of ethnicity. In the videos, the ethnic representer does only the following: (1) states first given name,<sup>10</sup> (2) states surname,<sup>11</sup> (3) greets in primary language,<sup>12</sup> (4) greets in English, (5) states and argues for true identity,<sup>13</sup> (6) states and argues for true identity with supportive ethnic symbol in the background, (7) states and argues for true identity with contradictory symbol,<sup>14</sup> (8) states and argues for false identity, (9) states and argues for false identity with supportive symbol, and (10) states and argues for false identity with contradictory symbol.<sup>15,16</sup> Each of these conditions captures various degrees of what Chandra (2006b) calls “descent-based attributes”. Some of these characteristics are easier to manipulate and some are more visible than others; and all of them offer varying amounts of information about ethnicity. Giving one’s first name or greeting in English provide far less information, for example, than stating one’s surname.

We categorize photographs and videos into “signs” and “signals” (Horowitz 1985; Habyarimana et al’s 2007). The photograph and videos 1-4 tested the impact of *signs* on ethnic identifiability: the “manifestation[s] of group membership that [are] beyond an individual’s control, at least in the short run” (6) such as physical characteristics, accents, and language. Videos 5-10 test

*signals*: “action[s] taken by an individual in order to communicate membership of an identity” (6). We chose these videos to capture the types of information that might be communicated about ethnicity in real-world settings.

Ethnic representers received 20 Rand (about US \$2.47) for participation and 2 Rand (~\$0.25) for each time they successfully passed as a member of a different ethnic group. The flat participation payment compensated representers for videos 1-7, and additional compensation was given for videos 8-10 as an incentive for representers to lie well. We capped winnings at 20 Rand per person (~\$2.47) for a total possible payment of 40 Rand.

#### 4.2 *Subjects*

The *subjects* were recruited exclusively from Mayibuya through poster advertisements covering all parts of the township. The subjects’ participation included a pre-study questionnaire, the viewing of photographs and videos in which subjects attempted to guess the representer’s identity, and a post-study questionnaire. The pre-study questionnaire measured the strength of each subject’s ethnic identity using Phinney’s (1992) Multigroup Ethnic Identity Measure (MEIM). Appendix A explains the measure. After completing the questionnaire, each subject viewed the photo and one of the videos (randomly selected) for all 76 ethnic representers.

Each subject attempted to guess the ethnic identity after each photo and video resulting in a total of 152 guesses per subject. With 62 subjects, the experiment yielded a total of 9,424 observations. Each subject received 20 Rand (about US \$2.47) for participation, and 50 South African cents (~\$0.07) for each correct guess to incentivize serious participation. We capped winnings at 30 Rand per person (~\$4.29) for a total possible payment of 50 Rand (~\$6.76).

#### 4.3 *Validity*

The experiment was carried out in a laboratory setting using pre-recorded photos and videos, which raises questions about validity. In what follows, we discuss key choices that we made to maximize our ability to make inferences beyond the laboratory. We also consider the limitations of the study's design and interpret our results accordingly. Our discussion highlights several key choices: the ethnic composition of the subjects, the information provided in each experimental condition, and the generalizability of the study beyond the Eastern Cape of South Africa

All subjects were Xhosas from Mayibuya Township, which is about 90% Xhosa. It is possible that individuals in homogeneous areas might not interact with others enough to identify them with significant consistency.<sup>17</sup> Theoretically, however, the most homogeneous groups might be those of greatest interest, because of their involvement in ethnic violence. In a study of how geography and settlement patterns affect violence, Toft (2005, 34) finds that “a group's concentration in a region of a state serves as practically a necessary condition for violence, whereas urbanism and dispersion are practically sufficient conditions for nonviolent political activity.” Saideman and Ayres (2000), furthermore, find that concentrated ethnic groups are more likely to fight secessionist wars. Varshney (2001) and Anderson and Paskeviciute (2006) argue that heterogeneity, if contributing to a vibrant civil society, might be more peaceful than relatively more homogeneous regions. Taken together, a great deal of evidence suggests that a consideration of areas with homogenous identity structures is important. Future research would nonetheless benefit from comparing homogeneous and heterogeneous areas more explicitly.<sup>18</sup>

We carefully constructed each experimental condition to capture a certain aspect of ethnic identity, especially categories posited by others as important in categorization: name, features, speech, and dress (Chandra 2004). We also present information that one might be able to gather when two people meet for the first time: clearly people evaluate each other based on facial cues or

clothing; both of these are reflected in the photos and, along with additional information, in the videos. Furthermore, it is easy to obtain an individual's first name or surname, which provide information about ethnic identity. Those seeking to identify others also rely on language cues, such as accents or expressions. The first five experimental conditions capture all of these factors. Representers provided still other information in the signal videos.

The ethnic representers have a vast repertoire of ethnic characteristics with which to communicate information about an ethnic identity in the signal videos, we provided specific scripts to prevent variation outside of the bounds of the experiment. Across the videos, ethnic representers addressed four topics that more closely tied the information to descent-based attributes (Chandra 2006b) of their identity: their stated ethnicity, their parents' ethnicities (to capture one element of social networks), the region their family comes from (to capture settlement and migration patterns), and up to three articles of traditional clothing (to capture cultural factors).

Although this does limit the information that is presented and used to identify, we feel that it is important to isolate how these four specific ethnic characteristics have an impact on identifiability because they are easily observed and the information can be fairly easy to obtain in real-world settings. For example, if someone wanted to know if their new neighbor was Zulu, she could easily ask the person (or another person in the neighborhood) where the family moved from. When it is revealed that the family comes from Kwa-Zulu Natal, then the person may feel that she needs no more information in order to judge what ethnic group the family belongs to. If she wanted further evidence, she might look to see what type of food the family eats or the traditional clothes the family wears when going to funerals and other occasions. Although these attributes are only presented verbally and in a laboratory setting rather than in action and real life, they capture important aspects



of ethnicity in a realistic yet controlled way in the laboratory that allows us to effectively evaluate the impact of these types of signals as well as for the sign videos.

The information provided in the experimental conditions does not directly capture other potentially important factors such as official documents or statements/denunciations by other individuals. We attempted to capture these possibilities indirectly by including substantiating or countervailing evidence during the signal videos. To do this, we placed a symbol associated with a group in the background of some of the signal videos.<sup>19</sup> In some of the signal videos we placed ethnic symbols on the wall behind the ethnic representer. At times these symbols supported what the individual was saying about herself, and at other times it contradicted what she was saying. These symbols gave the subjects some information beyond the representer's word to inform the subject's assessment. Thus, when a Xhosa person gave evidence of her Xhosa heritage with the picture of a Zulu warrior behind her, the subject may deduce that there is something amiss and would have to decide which to believe: the word of an individual or the context surrounding his interaction with her.

In short, we have attempted to construct experimental conditions that reflect a wide variety of realistic types of information that people might actually obtain about another person's ethnic identity.<sup>20</sup> Beyond the study's ability to reflect aspects of real-world circumstances, it is important to understand the applicability of this study beyond Mayibuya and the Eastern Cape of South Africa to other areas of South Africa and other developing countries. This study easily represents and applies to Mayibuye and the rest of the Eastern Cape Province because it is quite representative of the ethnic and socio-economic environment of the rest of the province (Statistics South Africa 2001).

Mayibuye is socio-economically representative of South Africa, but not ethnically representative to the same extent because South Africa itself is, in some areas, much more diverse. The Gauteng Province is extremely ethnically diverse, while Kwa-Zulu Natal is more Zulu than the Eastern Cape is Xhosa. Thus, we would assume, based on the ethnic diversity make-up of the Eastern Cape, that the results could be applied to other ethnically homogenous areas within South Africa. The results might also apply to individuals who spend a lot of time with co-ethnics even if they live in ethnically diverse areas. South Africa is not necessarily representative of the rest of Africa; it is far more developed and enjoys a relatively more stable political and economic environment. However, it is, like much of Africa, ethnically diverse and like much of the rest of Africa has certain areas within it that are more homogenous than others. For example, in Uganda, Kampala is dominated by Baganda but is much more ethnically diverse than any other area in the country; all other areas of the county are home to various majorities that dominate linguistic, cultural, political, economic, and social life. It would be safe to say that this study would apply to ethnically homogenous regions within ethnically heterogeneous countries.

South Africa's ethnic conflict history, apartheid, to the extent it was implemented in South Africa, is quite unique. But even under apartheid, a substantial amount of violence occurred between black ethnic groups rather than between whites and blacks. The Eastern Cape experienced particularly high levels of violence making it a suitable location to consider ethnicity in conflict or post-conflict zones. Thus, South Africa with its national ethnic diversity, ethnically homogeneous regions, and punctuations of violence, presents a representative and suitable location to conduct the research.

#### *4.4 Data Analysis*

Our primary goal is to evaluate whether the strength of ethnic identity affects individuals' abilities to guess the ethnic identity of others correctly. The dependent variable in each regression is *guesses* (correct guess=1, incorrect guess=0), which we regressed on the *ethnic strength* score and control variables all measured prior to the beginning of the experiment. Because the *guesses* variable is binary, we estimated a series of logit regressions, clustering on the ethnic representer. Descriptive statistics for all of the variables appear in Table 1.

We measure *ethnic strength* according to the Multigroup Ethnic Identity Measure (MEIM) (see Appendix A). *Subject age* and *ethnic representer age* control for whether experience makes the subject better able to identify people or the representer better able to represent one's group. Presumably the older a person is, the more aware they are of other groups such that they can identify and represent better. *Subject gender* and *ethnic representer gender* control for any gender-specific effects, though we have no a priori expectation about whether men or women are better guessers.

We included three education measures. First, *schooling* is the highest grade that the subject has completed. Second, *schooling father* is the highest grade that the subject's father completed, and third, *schooling mother* is the highest grade the subject's mother completed. We included all three measures to capture general knowledge about other groups and to capture socio-economic status without directly asking about income. We expect that those who have more educated backgrounds will know more about various ethnicities and guess more accurately.

The *Years in Mayibuya* variable measures how long the subject has lived in the township. Because Mayibuya is predominantly Xhosa, it is possible that those who have lived outside of the area have a better knowledge of other ethnic groups.<sup>21</sup> The final control variable, *religious activity*, measures how often a subject attends church or other religious ceremonies or gatherings and ranges

from never to more than twice a week. This variable controls for the religiosity of the subject rather than for religion itself.

Because subjects viewed and guessed on 152 photographs and videos, there are at least three potential biases due to time that we are concerned about. First, the subject might get better at guessing as she familiarizes herself with the experiment and the types of videos. Second, she might get tired and sloppy over time not fully considering the information presented in each photo or video. Third, she might learn quickly and become better in the middle of the experiment, but then get tired and less accurate over time.<sup>22</sup> We attempt to understand time effects by including an *experimental order* variable as well as the order variable squared to capture nonlinear dynamics. The *experimental order* variable is ordinal, ranging from 1 to 152 for each subject, and identifies when the subject viewed each photo or video. If subjects learn and get better, we would expect a positive relationship; if they get tired and sloppy, we expect a negative relationship; and if there is a nonlinear effect, the squared term should capture it. Table 1 reports the summary statistics for all variables.

**[TABLE 1 ABOUT HERE]**

## **5 Results**

The results show that the probability of successfully identifying the ethnicity of others is, on average, fairly low. Xhosas guess the identity of members of their own group as well as other groups correctly less than 50% of the time. Xhosas correctly guessed other Xhosas in 45% of the overall cases, which represents the highest probability. Table 2 reports the ratio of times that a subject guessed a certain ethnicity (rows) to the times that the subject *correctly* guessed the presenter's identity (column). For example, row two is Pedi, column five is Tsonga, and the number in the corresponding cell is 0.06. In this case, when Xhosas saw Tsongas, they mistakenly guessed Pedi 6%

of the time. Because the columns indicate the representer's stated identity, the numbers within the columns sum to 1 (before rounding). The principal diagonal (lightly shaded cells) represents the probability of correct identification for each given group. The off-diagonal represents the probability of incorrect identification where the darkly shaded cells represent the highest probability of incorrect identification. For example, the cell in the eighth row and third column reports that Xhosas confused Sothos for Xhosas 23% of the time, which is the highest probability of incorrect identification of the Sotho group.

**[TABLE 2 ABOUT HERE]**

Xhosas never correctly identified any ethnic group a majority of the time; the highest percentage they achieve is 45% when identifying their co-ethnics. Apparently, Xhosas consider some others as very similar: they incorrectly identify Pedi, Sotho, Tswana, Venda, and Zulu most often as being Xhosa (darkly shaded cells). These “errors of inclusion” illustrate the percentage of time subjects erroneously include members of other groups in their own (Habyarimana et al 2009). The Xhosas in our sample had error of inclusion rates, ranging from 13-23%. This gives an initial look at people's ability to identify ethnicity correctly, but we hypothesized earlier that strength of ethnic identity might be an important mitigating factor. It is important to note that although rates of successful identification are low generally, we can still examine the impact of ethnic strength because our interest is in *relative*, not *absolute*, identification. Thus Table 2 only presents an absolute picture of low identification, the remainder of the paper will investigate relative success based on ethnic strength. Yes, Xhosas may not be able to easily identify others' ethnicity, but are stronger ethnic Xhosas relatively better than weaker ethnic Xhosas? We now turn to a consideration of the ethnic strength variable and disaggregate the conditions (photos and videos) to evaluate the success rate given different types of information.

Table 3 reports a series of models estimating the effect of a subject’s ethnic identity strength. We first consider all conditions pooled together (Model 1) and then each of the five “sign” conditions separately (the picture and five videos). Model 1 shows that without differentiating among the conditions, Xhosas with stronger ethnic identities are more successful at identifying others’ ethnicities (at the .01 significance level).

**[TABLE 3 ABOUT HERE]**

To aid substantive interpretation, we calculated predicted probabilities for the identity strength measure for each regression. Table 4 reports the change in one’s ability to identify others correctly given a change from the lowest observed strength score (2.5) to the highest strength score (4). These calculations are based on a baseline probability of 0.27, and the conditions under which the strength score was a statistically significant indicator are shaded in the table. For Model 1, we found a 7.7% increase in the predicted probability of guessing the representer’s identity correctly. Pooling the conditions potentially masks disaggregated dynamics, however.

**[TABLE 4 ABOUT HERE]**

Models 2 through 6 show that a “stronger ethnic” is not necessarily more successful given more information. One important finding is that a stronger ethnic is significantly better at identifying another when only seeing the person (Model 2: Photograph), this is interesting because it suggests that a stronger ethnic is relatively better at identifying physical characteristics than a weaker ethnic. When a Xhosa only hears a person’s first name or only receives a greeting (whether in the person’s own language or English), the strength of ethnic identity does not have a significant impact on her ability to correctly identify others. The coefficients are in the expected positive direction but do not meet the 0.1 level of statistical significance. Some of these results are unsurprising. We might

expect that an individual's first name would convey little information about ethnicity; this appears to be the case even for a stronger ethnic. An individual's surname and greeting in the native language should, however, convey more information to distinguish ethnic identities.

An individual's surname in South Africa conveys significantly more information about group identity. Surnames are associated with language and tradition more than first names and surnames are passed down to children. We had a strong expectation that stronger ethnics could connect surnames to particular identities and the results show that this is the case. The finding is statistically significant at the 0.05 level and Table 4 provides a prediction of the finding's substantive significance. Although we expected a greeting in the representer's native language to aid correct identification, the results do not confirm that expectation. A likely explanation is that different ethnic groups often use similar words in their greetings, even though accents might be different. Depending on which part of the country they come from, for example, Pedis, Tswanas, and Sothos all use the same greeting regardless of their language. Also, most ethnic groups in South Africa have more than one greeting, which complicates the information the greeting provides.

Moving from Table 3 to Table 5, we now examine the effects of the signal videos. Models 7 through 9 consider correct guessing in the context of the videos in which the person is telling the truth; Models 10 through 12 consider the deceptive videos. In all of them, the person being recorded first states the ethnicity she belongs to or is trying to pass as, and then gives evidence for why she is a member of the stated group. The results of Model 7 show that subjects with a stronger ethnic identity are significantly (at the .01 level) better at correctly identifying people when representers state their ethnicity, offer additional information beyond a name or greeting, *and* tell the truth. This result seems intuitive, but recall that the representers tell the truth and lie in an equal

number of the videos. All subjects know that there is an equal probability of viewing an honest versus a dishonest video and, thus, must be cautious about believing an individual's stated identity.

In Models 8 and 9 (Table 5), representers again truthfully reveal their identity, but now cultural symbols belonging to the stated group and to a different group appear in the background respectively. The results are not much different when including the cultural symbols, even with the contradictory symbol. In both cases, the cultural symbols increased marginally the predicted probability of guessing the representer's identity correctly. It is important to note that the correct symbol (Model 8) in the truthful videos increased correct identification for the strong ethnic more than the incorrect symbol (Model 9). Thus, the ethnic context of an encounter appears to matter.

**[TABLE 5 ABOUT HERE]**

Coupled with the results of the deceptive videos, these results reveal an interesting relationship: having a stronger ethnic identity makes one more likely to identify those telling the truth, but significantly *less* likely to identify the ethnicity of others when they are lying. Models 10, 11, and 12 have a statistically significant negative effect between the strength of ethnic identity and one's ability to guess correctly, though Models 10 and 12 offer stronger support.

Social identity theory (Blascovich et al 1997) suggests that stronger ethnics know what differentiates other-ethnics from co-ethnics; weaker ethnics are less clear about what makes them different. Accordingly, a stronger ethnic might be more likely to believe any individual that can send the stereotypical cues that allow the stronger ethnic to differentiate between groups. When representers give information about their true ethnicity, strong ethnics know enough about other ethnicities to know whether they, in fact, are cues relevant to that group. The weaker ethnic may or may not be aware of these cues and, on average, will be more uncertain about whether the cues



communicate a meaningful characteristic of the group. The result is that because a stronger ethnic is more aware of cues, she is likely to believe the person if given the stereotypical cues.

Only a few of the control variables were related to guessing others' identity correctly. In Models 4, 6, 8, and 9, the Number of Years in Mayibuye increased the likelihood of correct identification, which suggests that the homogeneity of the population might reinforce ethnic boundaries, a result that is consistent with the argument that violence may originate in homogeneous areas. In Models 11 and 12 (deception videos), the more years the subject had been in Mayibuya, the less likely they were to guess correctly. These results offer some indirect support for the possibility that the homogeneity of the subjects did not consistently bias the results away from correct identification.

In four of the models, religious activity made successful identification of others less likely. The subject's gender had a positive impact in four of the models (1, 2, 7, 8, and 9). These results appear to indicate that, on average, men are better than women at guessing others' identities correctly. This result may be due to higher levels of migrancy by males in South Africa. Curiously, the subject's level of education was not significantly related to correct identification (nor was the father's or mother's level of education).

The experimental order variable never had a strong substantive impact on successful identification and it was only statistically significant in Model 6. This is encouraging and suggests that subjects did not learn to guess better as they progressed further into the experiment. It also suggests that subjects did not necessarily tire of guessing, but continued to try to identify the ethnic representers. It is possible that some subjects experienced one or the other, but that the two effects

cancelled each other out. We paid subjects for successful guessing to avoid these problems and we believe that this incentive served its purpose in keeping subjects motivated.

Finally, we considered several additional models to understand better how the homogeneity of Mayibuya might matter. We first considered a control for whether the ethnic representer currently lives in the Eastern Cape. This control sought to check whether the ethnic representers actually characterized their ethnic groups as opposed to someone who has been among Xhosas and has “acclimatized” to the Eastern Cape. The results of all the regression models are qualitatively the same. We then measured the number of languages a subject speaks as a proxy for how much exposure the subject has to people outside of the Eastern Cape (a factor that might also be captured by gender, years in Mayibuya, and the strength score). The results are again very similar to those reported in the paper, except in condition 2 the result is no longer significant and in condition 12 the result is only weakly significant ( $p=0.057$ ). In a final set of tests, we collapsed the data so that all non-Xhosa groups were in a single non-Xhosa category to test whether Xhosas correctly identified other Xhosas or any non-Xhosa. In these models, conditions 1, 3-4, and 6-8 were qualitatively the same as the results reported in the paper, model 5 becomes weakly significant ( $p=0.075$ ), and models 2 and 9-12 are no longer significant, suggesting that, when considered dichotomously, the effect of duping strong ethnics is not as robust as in previous models.

In sum, the empirical results support Hypothesis 1 because a stronger ethnic is generally more successful at identifying the ethnicity of others. We do find that the stronger ethnic is more successful when given certain types of information (surname and cues about correct identity), especially information that is more relevant to ethnic identity (Hypothesis 2). Hypothesis 3 is also supported because a stronger ethnic is less able to identify deceptive representers, apparently

believing the stereotypical cues provided by others, though these results are not as robust to dichotomizing ethnicity into Xhosa vs. non-Xhosa categories.

## 6 Conclusions

We have found evidence that a stronger personal ethnic identity increases one's ability to identify the ethnicity of others correctly; this impact is not consistent or straight forward, however. Pooling all conditions, the stronger ethnic is better at identifying others than is the weaker ethnic. But the effect varies depending on the type of information: when a stronger ethnic is given only limited information (except for the case of surname), she is not significantly better at identifying others. But when the stronger ethnic is told the identity of another (whether right or wrong), she is more prone to believe that person – correctly in some cases, but incorrectly in other cases. In short, a stronger ethnic is more prone to identify or believe others if they can give her sufficient correct information.

These findings offer insights to the ethnic conflict literature, especially prominent work such as the ethnic security dilemma (Posen 1993). The main premise of the ethnic security dilemma is that an ethnic group will start to mobilize for confrontation or conflict if it sees or perceives its rival ethnic group is mobilizing for such an event. However, this mobilization for mobilization's sake may not occur if ethnicity is not readily identifiable or is easily faked. For example, if an individual or group cannot readily identify the ethnicity of the perceived aggressors (whether on a group or individual level), then the individual or group will not respond in the theorized way. Similarly, if someone can lie about her identity and pass due to a strong ethnic's dupability, then that passing individual could either disguise which group is mobilizing or possibly present a different ethnic group (namely, the one she passed as) as the one that is mobilizing. Thus, the security dilemma

could be far more complex than originally theorized. Duping a stronger ethnic has important implications for how exactly the ethnic security dilemma could play out.

In the context of ethnic cleansing and genocide, having a stronger ethnic identity is a two-edged sword: a stronger identity might facilitate easier identification of others but the stronger ethnic might be fooled easier as well. Potential perpetrators of violence with stronger identities know the correct signs and symbols and if potential victims can give these signs and symbols, the potential perpetrator might indeed be convinced. Our results indicate that if a person is facing sanctions based on ethnicity, the best response is to fake a different ethnicity if one can produce the stereotypical information about another group. Of course, this is not always possible, because some descent-based attributes are not easily changed (Chandra 2006b).

Clearly much remains to be done in order to understand the conditions under which individuals can identify each other in order to target them. Extending the research into other geographic areas will be an important test of the generalizability of the results. Some, but not all, of our results are similar to those found in a quite different context (Habyarimana et al 2007); carrying out the research in other contexts could help sort out demographic or cultural differences, such as the heterogeneity or homogeneity of ethnicity, in the ability to identify others. Furthermore, our results show that while groups generally cannot correctly identify others reliably, there is cause to focus on stronger ethnics in future studies of ethnic identifiability. It is possible, for example, that identity strength is also picking up some other underlying concepts. Theories of social dominance, for example, suggest that hierarchy and power relations have an important impact on intergroup relations (Sidanius and Pratto 1999). The results of this study suggest that ethnic identifiability is complex and misunderstood and continued research on this topic could help elucidate the broader interest in ethnicity in the political science literature.

## References

- Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic Growth* 8 (2): 155-194.
- Allport, G., and B. Kramer. 1946. "Some Roots of Prejudice." *Journal of Psychology* 22: 9—39.
- Anderson, Christopher, and Aida Paskeviciute. 2006. "How Ethnic and Linguistic Heterogeneity Influence the Prospects for Civil Society: A Comparative Study of Citizenship Behavior." *Journal of Politics* 68(4): 783-802.
- Blascovich, Jim, Natalie A. Wyer, Laura A. Swart, and Jeffery L. Kibler. 1997. "Racism and Racial Categorization." *Journal of Personal and Social Psychology* 72 (6): 1364-1372.
- Brubaker, Rogers, Mara Loveman, and Peter Stamatov. 2004. "Ethnicity as Cognition." *Theory and Society* 33(1): 31-64.
- Cederman, Lars-Erik and Luc Girardin. 2007. "Beyond Fractionalization: Mapping Ethnicity onto Nationalist Insurgencies." *American Political Science Review* 101 (Feb): 173-185.
- Chandra, Kanchan. 2004. *Why Ethnic Parties Succeed: Patronage and Ethnic Head Counts in India*. Cambridge: Cambridge University Press.
- Chandra, Kanchan. 2006a. "Counting Heads: A Theory of Voter and Elite Behaviour in Patronage Democracies" In Herbert Kitschelt and Steven Wilkinson (Eds.) *Patrons, Clients or Politics: Patterns of Political Accountability and Competition*. Cambridge: Cambridge University Press.
- Chandra, Kanchan. 2006b. "What is Ethnic Identity and Does it Matter?" *Annual Review of Political Science* 9: 397-424.
- Chandra, Kanchan and Steven Wilkinson. 2008. "Measuring the Effect of 'Ethnicity'." *Comparative Political Studies* 41 (4-5): 515-563.
- de Figueiredo, R. and Z. Elkins. 2003. "Are Patriots Bigots? An Inquiry into the Vices of In-group Pride." *American Journal of Political Science* 47 (Jan): 171-188.
- Dorfman, Donald, Susan Keeve, and Carl Saslow. 1971. "Ethnic Identification: A Signal Detection Analysis." *Journal of Personality and Social Psychology* 18(3): 373-379.
- Fearon, James D. 1999. "Why Ethnic Politics and 'Pork' Tend to go Together." Presented at MacArthur Foundation-sponsored conference on Ethnic Politics and Democratic Stability. Wilder House, IL, University of Chicago, May 21-23.
- Fearon, James D. and David D. Laitin. 1996. "Explaining Interethnic Cooperation." *American Political Science Review* 90 (4): 715-735.

- Fearon, James D. and David D. Laitin. 2003. "Ethnicity, Insurgency and Civil War." *American Political Science Review* 97 (Feb): 75-90.
- Ferree, Karen. 2004. "The Microfoundations of Ethnic Voting: Evidence from South Africa." *Afrobarometer Working Paper No. 40*. Cape Town, South Africa: Institute for Democracy in South Africa.
- Gibson, James L. 2004. *Overcoming Apartheid: Can Truth Reconcile a Divided Nation?* New York: Russell Sage Foundation.
- Gibson, James, and Amanda Gouws. 2003. *Overcoming Intolerance in South Africa: Experiments in Democratic Persuasion*. Cambridge: Cambridge University Press.
- Gourevitch, Philip. 1998. *We Wish to Inform you that Tomorrow We will be Killed with our Families*, New York: Farrar, Straus and Giroux.
- Habyarimana, James, Macartan Humphreys, Daniel Posner, and Jeremy Weinstein. 2007. "Placing and Passing: Evidence from Uganda on Ethnic Identification and Ethnic Deception." Presented at the American Political Science Association, Chicago, IL.
- Habyarimana, James, Macartan Humphreys, Daniel Posner, and Jeremy Weinstein. 2009. *Coethnicity: Diversity and the Dilemmas of Collective Action*. Russell Sage Foundation Press.
- Hale, Henry. 2008. *The Foundations of Ethnic Politics: Separatism of States and Nations in Eurasia and the World*. Cambridge: Cambridge University Press.
- Horowitz, Donald L. 1985. *Ethnic Groups in Conflict*. Berkeley, CA: University of California Press.
- Horowitz, Donald L. 2001. *The Deadly Ethnic Riot*. Berkeley, CA: University of California Press.
- Kirschner, Shanna, A. 2007. "Fear and Ethnic Civil War Duration." Presented at the Journey's workshop. Available at: <<http://www.saramitchell.org/kirschner.pdf>>.
- Laitin, David D. 1995. "Marginality: A Microperspective." *Rationality and Society* 7 (January): 31-57.
- Lindzey, G., and S. Rogolsky. 1950. "Prejudice and Identification of Minority Group Membership." *Journal of Abnormal and Social Psychology* 45: 37—53.
- Phinney, Jean S. 1992. "The Multigroup Ethnic Identity Measure: A New Scale for use with Diverse Groups." *Journal of Adolescent Research* 7 (April): 156-176.
- Posner, Barry. 1993. "The Security Dilemma and Ethnic Conflict." *International Security* 35(1): 27-47.
- Posner, Daniel. 2004. "Measuring Ethnic Fractionalization in Africa." *American Journal of Political Science* 48 (4): 849-863.

- Posner, Daniel. 2005. *Institutions and Ethnic Politics in Africa*. Cambridge: Cambridge University Press.
- Posner, Daniel. 2007. "Regime Change and Ethnic Cleavages in Africa" *Comparative Political Studies* 40(11): 1302-1327.
- Quanty, Michael, John Keats, and Stephen Harkins. 1975. "Prejudice and Criteria for Identification of Ethnic Photographs." *Journal of Personality and Social Psychology* 32(3): 449-454.
- Rice, Diana, and Brian Mullen. 2003. "Isaac, Ishmael, and Janus: Past and Future Lessons Regarding the Ethnic Categorization of Faces." *Applied Cognitive Psychology* 17(9): 1129-1147.
- Saideman, Stephen, and William Ayres. 2000. "Determining the Causes of Irredentism: Logit Analyses of Minorities at Risk Data from the 1980s and 1990s." *Journal of Politics* 62(4): 1126-1144.
- Saideman, Stephen, David Lanoue, Michael Campenni, Samuel Stanton. 2002. "Democratization, Political Institutions, and Ethnic Conflict: A Pooled Time-Series Analysis, 1985-1998." *Comparative Political Studies* 35(1): 103-129.
- Sambanis, Nicholas and Jonah Schulhofer-Wohl. 2008. "What's in line? Partition is not the solution to civil war." Unpublished manuscript.
- Savitz, Leonard, and Richard Thomasson. 1959. "The Identifiability of Jews." *American Journal of Sociology* 64 (5): 468-475.
- Sidanius, Jim, and Felicia Pratto. 1999. *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression*. Cambridge: Cambridge University Press.
- Statistics South Africa. 2001. Census 2001. At: <[http://www.statssa.gov.za/census01/html/RSA Primary.pdf](http://www.statssa.gov.za/census01/html/RSA_Primary.pdf)>. 23 April 2008.
- Tajfel, Henri. 1981. *Human Groups and Social Categories: Studies in Social Psychology*. New York: Cambridge University Press.
- Tajfel, H and J. C. Turner. 1979. "An Integrative Theory of Intergroup Conflict." In *The Social Psychology of Intergroup Relations*, ed. W. G. Austin and S. Worchel, 33-47. Monterey, CA: Brooks/Cole.
- Tajfel, H and J. C. Turner. 1986. "The Social Identity Theory of Intergroup Behavior." In *Psychology of Intergroup Relations*, ed. S Worchel and W. G. Austin, 7-24. Chicago: Nelson-Hall.
- Toft, Monica. 2005. *The Geography of Ethnic Violence: Identity, Interests, and the Indivisibility of Territory*. Princeton, NJ: Princeton University Press.
- Varshney, Ashutosh. 2001. "Ethnic Conflict and Civil Society: India and Beyond." *World Politics* 53 (3): 362-398.

**Table 1: Summary Statistics**

Variable	Mean	Standard Deviation	Minimum	Maximum
Guess	0.27	0.45	0	1
Ethnic Strength	3.42	0.34	2.5	4
Subject Age	30.06	11.32	16	59
Subject Gender	0.53	0.5	0	1
Schooling	5.44 <sup>+</sup>	1.17	4	7
Schooling Father	4.47	2.07	1	8
Schooling Mother	4.5	1.51	2	8
Years in Mayibuya	20.24	11.77	1	59
Religiosity	2.61	1.6	1	7
Ethnic Representer Age	29.74	11.89	16	72
Ethnic Representer Gender	0.54	0.5	0	1

<sup>+</sup>Converting 5.44 to a qualitative interpretation shows that on average, the subjects have at least completed high school because 5 represents “completed high school” and a 6 represents “completed some university or college”.



**Table 2: Xhosas' Probability of Correct/Incorrect Classification**

Guess	Correct Identity								
	Ndebele	Pedi	Sotho	Swati	Tsonga	Tswana	Venda	Xhosa	Zulu
Ndebele	0.22	0.07	0.05	0.05	0.07	0.06	0.08	0.06	0.08
Pedi	0.06	0.17	0.06	0.08	0.06	0.08	0.04	0.05	0.07
Sotho	0.15	0.15	0.22	0.08	0.08	0.14	0.08	0.08	0.09
Swati	0.07	0.07	0.07	0.14	0.08	0.08	0.12	0.05	0.07
Tsonga	0.06	0.08	0.06	0.06	0.18	0.07	0.08	0.06	0.06
Tswana	0.02	0.1	0.11	0.15	0.12	0.18	0.06	0.06	0.08
Venda	0.17	0.1	0.07	0.1	0.17	0.09	0.18	0.07	0.08
Xhosa	0.13	0.15	0.23	0.15	0.16	0.18	0.22	0.45	0.18
Zulu	0.12	0.13	0.12	0.19	0.09	0.13	0.14	0.13	0.28

Note: reports the ratio of times in which a subject guessed a certain ethnicity (rows) to the times in which the subject correctly guessed the representer's identity (column). The principal diagonal (lightly shaded cells) represents the probability of correct identification. The off-diagonal represents the probability of incorrect identification where the darkly shaded cells represent the highest probability of incorrect identification.

**Table 3: Ethnic Strength and Identification In Different Conditions**

Dependent Variable: Correct Guesses

VARIABLES	(1) All Conditions	(2) Photograph	(3) First Name	(4) Surname	(5) Greet Own Language	(6) Greet English
Subject: Strength Score	0.266*** (0.072)	0.145* (0.098)	0.334 (0.286)	0.585** (0.298)	0.115 (0.334)	0.342 (0.465)
Subject: Age	0.000 (0.002)	0.000 (0.004)	-0.017 (0.012)	-0.028 (0.011)	-0.001 (0.010)	-0.020 (0.017)
Subject: Gender	0.143*** (0.057)	0.175** (0.087)	-0.306 (0.228)	-0.052 (0.269)	-0.162 (0.204)	-0.538 (0.306)
Subject: Education	0.025 (0.024)	-0.019 (0.035)	0.179* (0.117)	-0.019 (0.115)	0.211** (0.097)	-0.073 (0.133)
Subject: Father's Education	-0.005 (0.011)	-0.004 (0.021)	0.028 (0.058)	-0.078 (0.060)	0.071* (0.051)	0.005 (0.060)
Subject: Mother's Education	-0.015 (0.017)	-0.066** (0.027)	0.049 (0.065)	0.019 (0.068)	0.074 (0.084)	0.010 (0.105)
Subject: Years in Mayibuya	0.003 (0.002)	-0.001 (0.003)	0.008 (0.010)	0.018** (0.010)	0.007 (0.009)	0.032** (0.019)
Subject: Religious Activity	-0.031** (0.017)	-0.018 (0.026)	0.075 (0.064)	-0.060 (0.071)	-0.009 (0.062)	0.074 (0.092)
Representer Age	0.013** (0.007)	0.018** (0.009)	0.007 (0.014)	0.012 (0.013)	0.015 (0.019)	0.040*** (0.015)
Representer Gender	0.056 (0.134)	0.031 (0.196)	0.060 (0.349)	-0.091 (0.297)	0.103 (0.325)	0.467* (0.326)
Experimental Order	-0.002 (0.002)	-0.004 (0.004)	-0.002 (0.009)	-0.009 (0.009)	-0.004 (0.009)	-0.020** (0.012)
Order Squared	0.000* (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000** (0.000)
Constant	-2.372*** (0.359)	-2.093*** (0.534)	-3.304*** (1.263)	-2.046** (1.145)	-2.752** (1.325)	-3.578** (1.990)
Observations	9402	4712	492	483	476	424

S. E. in parentheses; \* =  $p < .1$ , \*\* =  $p < .05$ , \*\*\* =  $p < .01$  (significance tests are one-tailed). Ethnic representers are those in the pictures/videos and subjects are those that are viewing the videos. Standard errors clustered on ethnic representer.

**Table 4: Substantive Impact of Identity Strength on Correct Guesses**

Condition	Strength Score $\Delta$	Probability $\Delta$	Condition	Strength Score $\Delta$	Probability $\Delta$
All Conditions	(2.5 to 4)	7.7	Signal True Identity	(2.5 to 4)	31.3
Photograph	(2.5 to 4)	3.1	Signal Own Identity with Own Symbol	(2.5 to 4)	33.4
First Name	(2.5 to 4)	9.7	Signal Own Identity with Other Symbol	(2.5 to 4)	34.4
Surname	(2.5 to 4)	16.6	Signal Other Identity	(2.5 to 4)	- 11.0
Greet: Own Language	(2.5 to 4)	4.3	Signal Other Identity with Other Symbol	(2.5 to 4)	- 7.4
Greet: English	(2.5 to 4)	6.1	Signal Other Identity With Own Symbol	(2.5 to 4)	- 13.4

**Table 5: Ethnic Strength and Identification In Different Conditions**

Dependent Variable: Correct Guesses

VARIABLES	(7) Signal Own (True) ID	(8) Signal Own w/Own Symbol	(9) Signal Own w/Other Symbol	(10) Signal Other (Fake) ID	(11) Signal Other ID w/Other Symbol	(12) Signal Other ID w/Own Symbol
Subject: Strength Score	1.018*** (0.332)	1.269*** (0.412)	1.064*** (0.357)	-1.371** (0.812)	-0.713* (0.505)	-0.919** (0.498)
Subject: Age	0.017* (0.012)	0.026** (0.014)	0.021** (0.011)	-0.017 (0.039)	0.010 (0.019)	0.009 (0.025)
Subject: Gender	0.796*** (0.281)	0.409* (0.280)	0.770*** (0.280)	0.038 (0.571)	0.584 (0.361)	0.205 (0.311)
Subject: Education	0.064 (0.121)	-0.140 (0.129)	-0.022 (0.104)	0.131 (0.296)	0.078 (0.153)	0.103 (0.123)
Subject: Father's Education	-0.112 (0.045)	0.013 (0.060)	0.016 (0.049)	-0.024 (0.099)	0.101 (0.077)	-0.010 (0.080)
Subject: Mother's Education	-0.003 (0.084)	0.053 (0.098)	0.066 (0.086)	-0.046 (0.194)	-0.062 (0.115)	0.177 (0.129)
Subject: Years in Mayibuya	0.014 (0.011)	0.017* (0.013)	0.023** (0.013)	-0.011 (0.023)	-0.047*** (0.017)	-0.025** (0.013)
Subject: Religious Activity	-0.192*** (0.071)	-0.152** (0.087)	-0.194** (0.077)	0.061 (0.181)	0.057 (0.113)	0.221** (0.114)
Representer Age	0.013* (0.009)	0.001 (0.010)	0.005 (0.008)	-0.019 (0.027)	0.005 (0.013)	0.020 (0.014)
Representer Gender	-0.097 (0.222)	0.610** (0.273)	0.296 (0.220)	0.141 (0.617)	-0.179 (0.336)	0.413 (0.315)
Experimental Order	-0.004 (0.010)	0.001 (0.012)	0.016 (0.008)	-0.002 (0.027)	0.002 (0.015)	0.007 (0.015)
Order Squared	-0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)
Constant	-3.046*** (1.192)	-3.819*** (1.550)	-5.012*** (1.488)	2.518 (2.480)	-0.965 (1.855)	-2.057 (1.717)
Observations	487	461	470	462	493	442

S. E. in parentheses; \* = $p < .1$ , \*\*= $p < .05$ , \*\*\*= $p < .01$  (significance tests are one-tailed). Ethnic representers are those in the pictures/videos and subjects are those that are viewing the videos. Standard errors clustered on ethnic representer.

## Endnotes

---

<sup>1</sup> While we are motivated by the literature on ethnic conflict and violence, we note that the results of this study might have implications for other research domains on ethnicity. Identifiability is likely important in studies of parties, voting, clientelism, and patronage (for examples, see Chandra 2004, 2006a; Posner 2005; Posner 2007; Ferree 2004).

<sup>2</sup> Notable exceptions include Allport and Kramer (1946), Lindzey and Rogolsky (1950), Dorfman, Keeve, & Saslow (1971), Quanty, Keats, and Harkins (1975), Habyarimana, et al (2009).

<sup>3</sup> While social identity theory could apply to any type of group, ethnicity is one important category within which people sort themselves.

<sup>4</sup> Social identity theory has had a strong impact on the field of ethnic relations, but its foundations have been examined extensively and often found to be fragile (e.g., Brubaker, Loveman, and Stamatov 2004; Hale 2008; Yamagishi and Kiyonari 2000). The basic argument that minimal group categorization can produce group-oriented behavior finds support in nearly all work on social identity (Hale 2008).

<sup>5</sup> We use the term “stronger ethnic” to refer to someone who has a stronger tie to his/her ethnic group relative to other members of the group.

<sup>6</sup> Note that this hypothesis makes no claims about the baseline levels of successful identification. We are primarily interested in what causes the variation around the baseline, regardless of how high or low it is. Our theoretical argument suggests that, because many individuals are not strong ethnics, the baseline should be relatively low. We present results about both the baseline levels and the variation around those levels in the results section.

<sup>7</sup> The name Mayibuya is substituted for the actual township name to preserve anonymity.

<sup>8</sup> The identities from the census represent what Habyarimana et al (2009) call “benchmark” identities. It is possible that they differ from what individuals would subjectively report if asked. We used the benchmark identities because, in the South African context, they capture the identities that individuals typically use. In Gibson’s (2004) study, black South Africans almost always identified themselves either as South African or as one of the ethnic identities on the census. That is, individuals identified themselves with one of the major categories, rather than another sub-category.

<sup>9</sup> According to the 2001 Census this distribution is as follows: Zulu 23.8%, Xhosa 17.6%, Pedi 9.4%, Tswana 8.2%, Sotho 7.9%, Tsonga 4.4%, Swati 2.7%, Venda 2.3%, and Ndebele 1.6% (Statistics South Africa 2001).

<sup>10</sup> The ethnic representer was asked to give the name with which she would introduce herself in a conversation. In English, she says “My name is [insert name]”

<sup>11</sup> In English, she says, “My surname is [insert surname]”

<sup>12</sup> For videos 3 and 4, the ethnic representer has the freedom to choose if she wanted to greet casually or formally. She also chose if she wanted to greet one person or many people. This was done so that each subject could greet in the way he/she felt most comfortable.

<sup>13</sup> In videos 5– 10, the individual states the appropriate ethnic identity and then proceeds to argue why she belongs to the purported ethnic group by using the following script: “I am a [insert ethnic group]. My father is a [insert ethnic group] and my mother is a [insert ethnic group]. My family comes from [insert province or town of origin]. And traditionally, we wear [insert up to three articles of traditional clothing].”

---

<sup>14</sup> For video 7, the “contradictory symbol” is the symbol of the assigned false ethnic identity; e.g. If a Xhosa is going to be assigned to convince people that she is Zulu, then in video 7, she is saying that she is Xhosa with the Zulu symbol behind her.

<sup>15</sup> For video 10, the “contradictory symbol” is the symbol of the individual’s true ethnic identity; e.g. If a Xhosa is assigned to convince people that she is Zulu, then in video 10, the person is saying she is Zulu with the Xhosa symbol behind her.

<sup>16</sup> The “false” identity was assigned prior to meeting with each ethnic representer. We first assigned at least one person in each ethnic group to pretend to be a member of one of the other eight ethnic groups so that we covered all groups, and then randomly assigned the remaining ethnic representers. For example, we recorded 23 Zulus, so we first assigned one man and one woman to pretend to be each of the other eight ethnic groups, then we randomly assigned the remaining seven individuals making sure that none of these seven pretended to be members of the same group. When there were less than eight ethnic representers from a given ethnic group, we randomly assigned the false ethnic group.

<sup>17</sup> A homogenous group also may not be attentive to differences in ethnicity, unless there is a compelling reason to be concerned. In the case of Xhosas in South Africa, their dominance was threatened by Jacob Zuma’s ascent to the presidency of the ANC on the 18<sup>th</sup> of December 2007. Given the historical rivalry between Xhosas and Zulus, it could be important for Xhosas to differentiate between Xhosas and Zulus, if not Xhosas and outsiders more generally.

<sup>18</sup> We address the homogeneity issue in several other ways. First, the rationale for having subjects from only one group was to explore variation on ethnic strength *within* the group as opposed to across groups, which we discuss more below. Given resource constraints, we chose to emphasize this dimension in order to understand strength of identity as clearly as possible. Second, although Mayibuya is relatively homogenous, we learned from interacting with Mayibuya residents that many do indeed interact with people from other ethnic groups, though not as much as someone that lives in the more heterogeneous areas of Johannesburg. In addition to interpersonal interactions, all news in the area is national and most people follow the news regularly. Third, in our analysis we control for the number of years the subject has lived in Mayibuya to understand how insulated they have been.

<sup>19</sup> We chose obvious symbols for each group so that the symbols would be archetypical symbols and have approximately equal effects. As an example of how we introduced symbols, for a Xhosa representer, we recorded three separate videos. In them, a Xhosa representer tried to convince the subject that she is Xhosa: (1) without any symbol, (2) with the correct picture of a Xhosa in the background, and (3) with the incorrect Zulu picture. The same representer would then be recorded three more times trying to convince the subject that she is Zulu: (1) without any picture in the background, (2) with the accompanying Zulu symbol, and (3) with the Xhosa symbol.

<sup>20</sup> The focus of the experiment is on individuals. States may also be involved in defining, identifying, and verifying ethnicity, but they are not the only actors that do so. Furthermore, the involvement of the state may fluctuate over time. In many countries, such as South Africa, the state has been far more involved in the past than it is now. Ultimately, individuals and groups of individuals might care about ethnicity, because it can affect their personal welfare on a daily basis. Thus, the individual-level does not capture the full range of interest on identifiability, but it is an important step.

<sup>21</sup> This measure is imperfect, because it picks up age to some extent. Those individuals who have lived in the township for short periods of time might have just moved in or they might just be younger subjects.

<sup>22</sup> It is also possible that individuals would begin successfully, get worse, and then get better, though this option seems unlikely.