Social identity modeling: past work and relevant issues for socio-cultural modeling

Jonathon Kopecky
Nathan Bos
Ariel Greenberg
Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Rd  Laurel, MD 20723
{jonathon.kopecky, nathan.bos, ariel.greenberg}@jhuapl.edu

ABSTRACT: Many of today’s political conflicts are based on social identity differences, and sides are drawn up along ethnic, religious, ideological lines. Socio-cultural modeling efforts need to be able to incorporate realistic social identity dynamics that are based in academic literature and build on prior work. This paper reviews four modeling efforts in this area: Aptima’s SCIPR, Salzarulo’s Metacontrast model, Lustik’s PS-I model, and Johns Hopkins APL’s SILAS. Each is analyzed as to its mix of descent (permanent, inherited) and flexible identities, how each handles changing salience using Turner’s Accessibility x Fit model, and how each uses data for grounding and validation.

1. Modeling Social Identity

“It is increasingly apparent how many of the dangerous conflicts around the world are defined in terms of some variant of ‘identity politics’” (Lustick, 2002). Tutsi versus Hutu violence in Rwanda, Sunni versus Shia violence in Baghdad and Serb versus Bosniak violence in Bosnia and Herzegovina are a few recent examples of conflicts in which social identity (in addition to the usual political and economic factors) were critical causes and of conflict. There is a current emphasis on modeling in the human social, cultural, and behavioral area, (HSCB) and the dynamics of social identity should be a prominent part of these models. However, social identity is not the most easily tractable topic area for modeling, with components that are complex, highly contextual, and have important individual differences.

Identity refers to a person’s collective identity. All individuals have a sense of belonging to multiple identity groups. Since the 1950’s psychologists have used the simple “Twenty statements test” to gauge self-concept (Kuhn & McPartland, 1954) where participants make 20 statements in the form of “I am __.” Responses tend to fall into five groups, one of which is social categorization, or social identities. Social identity responses might be “I am Christian”, “I am American”, or “I am a Teamster.” Individual may have many social identities along dimensions of ethnicity, religion, politics, economics, and ideology, among others.

Knowing an individual’s identity affiliations can be the key to understanding attitudes and opinions, as individuals tend to adopt opinions compatible with their salient identity groups (Haslam & Turner, 1992; Haslam & Turner, 1995; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). Identity can help explain the day-to-day behavior of individuals when rituals, mores, practices, or more subtle behavior patterns are associated with identity groups. (Abdelal, Herrera, Johnston, & McDermott, December 2006). Understanding the pattern of identities in a population is a key to understanding conflicts, predicting both where conflicts are most likely to occur, and predicting how groups are likely to align in a conflict situation.

Modeling identity is more complex than simply modeling demographic differences, however, which means that modelers have to do more than simply recreate populations with known ethnic, religious, and political statistics. There is an extensive literature in the behavioral sciences dealing with the definitions, implications, and malleability of social identities. It is not the purpose of this paper to comprehensively review this literature. Nor is it probably feasible (or necessary) for socio-cultural models to incorporate every social and psychological nuance of identity. Some issues are more critical than others for modeling, which will be the focus of this paper.
Although identities themselves can change over time (e.g. the ‘Catholic’ identity may become more secular or more religious over a generation), we limit our discussion to models of time scales in which the properties of the actual identity as constant. This paper will focus a subset of issues that are important for models in the HSCB domain about how individuals select a particular identity, which can be used to model political trends, conflict, and related social issues. We will address two overarching modeling issues:

1. Identity permanence. How do modelers differentiate between identities that can change easily and those that cannot?
2. Identity salience. How are individual or group identities expected to change in importance based on the situation?

We will also discuss, in the context of prior examples, three related issues:

3. Identity and influence. How do individuals in a social network affect each others’ identity affiliations?
4. Ingroups and outgroups. How do identity groups define themselves in comparison to each other, and what are the resulting dynamics?
5. Relationships between identity groups. How do identity groups show affinity or rivalry with each other, and how does this affect alignments in conflict situations?

1.1 Identity permanence
There is an important distinction between descent identities such as ethnicity, which are relatively fixed, and flexible identities such as political party affiliation. Across cultures, an identity may vary in being assessed as descent or flexible.

Descent identities are identities that individuals are born with and that are difficult to impossible to change, especially in the short-term (cf. ‘stickiness’ in the political science literature, e.g. Chandra, 2006). Obvious examples of descent identities are ethnicity and race. Individuals who identify as African-American are going to have some connection to the African-American identity group their entire lives. While this identity may be nuanced or augmented, it cannot be changed to a completely different group (e.g. Asian). Religion can also be treated as a descent identity. Although, technically individuals can convert from one religion to another, this is very difficult if not impossible in many parts of the world and usually carries a high cost, such that most models should regard this variable as permanent. Descent identities should not necessarily be considered exclusive, however. Conversion or intermarriage may tie a person to more than one identity group. A Caucasian woman with an African-American husband and children may adopt a strong affiliation to that identity group, even though it was not hers by birth. Descent identities are augmented, but not replaced. Even in the case of conversions or intermarriage, an individual’s original religious or ethnic identity still affects behavior. People carry multiple descent identities, although they often differ in salience, as will be discussed.

Flexible identities are those identities which individuals can change fairly easily with relatively low cost. The most commonly modeled flexible identities are political-party affiliations and occupation. It is usually possible to switch political parties or occupations, and usually the barriers are much lower than those related to changing religions. Ideologies that blend the social and political are a third common example of flexible identities: ‘environmental activist’; ‘evangelical conservative’; and ‘moderate Islamist’ might be examples. Not every belief constitutes an ‘identity’, (e.g. ‘Ford truck advocate’ probably does not need to be modeled as an identity group in most sociocultural models); but beliefs that connect people to larger groups with established norms and that affect a variety of behaviors may need to be modeled as such.

Some identities, such as social class, may need to be treated as descent identities in some settings and flexible in others. In regions of the world known to have strong class distinctions and low economic mobility, social class and even occupation may be a descent identity; but these should be treated as flexible in most parts of the developed world.

1.2 Determining salience: Accessibility x Fit
While every individual can hold multiple identities of multiple types, the importance of these identities can change radically from one circumstance to another.
Understanding when particular identities are salient is a critical capability. We will use the concepts of Accessibility and Fit, which are aspects of Turner’s Social Categorization Theory (SCT; Turner et al., 1987, Bruner, 1957, Blanz, 1999) as a way of thinking about differences in salience. Salience is the product of a relatively permanent ‘accessibility’ parameter and a contextual ‘fit’ for a particular identity (Salience = Accessibility * Fit).

Individuals have self-identities that are more or less salient. For one individual, their religion may be the most important component of their identity, while for another, an economic identity (e.g. ‘successful businessman’) may be most salient. In our research group’s work modeling Nigeria, ethnic loyalties were thought to be particularly important. For this research, we benefitted from a data source which asked questions directly about salience. The data source is Afrobarometer (www.afrobarometer.org), which is a repeated survey of a number of West and South African countries. In addition to collecting demographic information for each respondent, Afrobarometer asked each respondent: “Besides being (your country’s nationality), which specific group do you feel you belong to first and foremost?” The answer to this question would be the non-national identity most salient at that moment. This data showed how salience varied across individuals, and also how it varied systematically across different segments of the Nigerian population. For example, religious identity was most salient for Muslim Hausas in Nigeria, while ethnic identity was most salient for Christian Igbos. There was also considerable individual variation—each group included some individuals with strongly salient ethnic, religious, political, and economic identities.

Accessibility is the ease of retrieving a given identity to mind, similar to the ‘availability heuristic’ from cognitive psychology (Tversky & Kahneman, 1973). Identities that are more familiar or carry more emotional valence are more accessible. For example, it is relatively easy for Americans to retrieve the identities ‘Christian’ or ‘Muslim,’ and generally harder to retrieve some other religious identities (e.g. ‘Rastafarian’, ‘Sunni’, ‘Shintoist’). The harder it is to retrieve a particular identity, the less likely a person is to categorize either themselves or another into that category.

Some identity categories are also more accessible than others. Most individuals have ethnic, religious, and occupational identities, but they are not equally accessible. Research has shown that ethnic, religious, and political identities tend to be more accessible than occupational or relational identities (such as ‘husband’ or ‘son’, Deaux, Reid, Mizrahi, & Ethier, 1995).

One example where accessibility affects perception is in the American perception of 9-11 hijackers. Although 15 of the 19 were from Saudi Arabia, ‘Saudi Arabian’ has low accessibility for most Americans, so very few Americans noticed or remembered that the hijackers were Saudi Arabian. However, both ‘Muslim’ and ‘Iraqi’ had much higher accessibility, thus the identity of the hijackers was more easily perceived to be Muslim (which was accurate) or Iraqi (which was not). (This may help explain why over 40% of Americans felt that Iraq played a direct role in the 9-11 attacks, Wolf, 2007).

Fit is the degree to which a particular context activates particular identities. While accessibility is considered to be a relatively fixed feature of an identity for any given individual, contextual fit can vary widely. Current events can strongly interact with particular identities. We saw that in America after 9/11, people’s American identity was more salient than their political identity, because of the ‘fit’ between the events and national identities. Lewis (2007) showed that identity affiliation in Nigeria changed markedly between 2001, 2003, and 2005, with ethnic identities significantly higher in the first and last. His explanation: elections were being held near the 2001 and 2005 data collection events, and Nigerian elections have often been seen as contests between ethnic groups. Using SCT terminology, the context of Nigerian national elections had a high degree of ‘fit’ with ethnic identities.

Fit can also be affected by a particular social context. Nigerian expatriates living in the U.S. may become particularly conscious of their Nigerian identity, especially in the company of other Nigerians. When we review Salzarulo’s model, the use of metacontrast
rations to quantify comparisons between identity groups will be relevant to this kind of fit.

Framing or re-framing of an event can attempt to change the ‘fit’ of events, and thus change which identities becomes salient. This is one of the techniques used by Al Qaeda to try to elicit sympathy for themselves, by portraying Al Qaeda actions done against specific American or European targets as part of a conflict between Muslims in general and Western powers in general. In the language of SCT, Al Qaeda tries to create a fit between specific events and identities that are highly accessible to their audience: Islam and the West.

## 2. Four models of social identity

We will review four modeling efforts where social identity plays a large role. We will describe each model’s unique strengths, and compare how they handle identity permanence (descent versus flexible identities) and identity salience (with components of accessibility and fit). Each model also brings in additional theoretical issues, which will be described in the context of each model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Types of identities</th>
<th>Data sources</th>
<th>Key features</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCIPR</td>
<td>Flexible: Political opinions</td>
<td>Grounded and validated with IRA attack data and voting results from Northern Ireland</td>
<td>Models influence using a bounded confidence model. Includes multiple overlapping identities and uses a simple social network for influence.</td>
</tr>
<tr>
<td>Salzarulo’s MetaContrast model</td>
<td>Flexible: Belief-based social categories</td>
<td>Synthetic</td>
<td>Illustrates how polarization and extremism can occur due to combination of attraction to ingroups and repulsion from outgroups</td>
</tr>
<tr>
<td>PS-I</td>
<td>Flexible and Descent: Political/cultural identity groups</td>
<td>Author’s regional expertise</td>
<td>Models geographic clusters, or ‘polities’, and spread of identities through a population</td>
</tr>
<tr>
<td>SILAS</td>
<td>Flexible and Descent: Ethnic, Religious, and political identity affiliations</td>
<td>2001 Afrobarometer survey of Nigeria used for grounding and validation</td>
<td>Models how internal conflicts between identities may be resolved; models ‘common enemy’ dynamic</td>
</tr>
</tbody>
</table>

Table 1.1 Overview of social identity models

**Aptima’s SCIPR** (Simulate Cultural Identities for Predicting Reactions to Events) is an agent-based model of opinion dynamics (Grier, Skarin, Lubyansky, & Wolpert, 2008). A collection of agents maintains a set of possible identities, where each identity is defined by a set of beliefs. Each agent also has a synthetic social network of associates, largely determined by geographic proximity. As the model runs forward in time, agents influence each other to try to draw others closer to their beliefs, and thus influence political party affiliation.

Central to the SCIPR model is a model of ‘bounded confidence’. Agents hold beliefs and also have a degree of confidence associated with those beliefs. This confidence strongly constrains how easily they can be influenced by other agents. When the model is running agents try to influence the other agents in their social network, but can only influenced by them if 1) the two agents are demographically similar, and 2) the influence message being sent is close enough to the receiving agent’s current beliefs agent to fall within that agent’s confidence parameters. Agents with less confidence are both more likely to listen to agents whose starting position is dissimilar to their own and more easily persuaded by new messages. Agents with very strong confidence are very resistant to changing opinions, although in the absence of reinforcing messages from similar agents, confidence does decay over time.

For the 2006 paper cited here, the SCIPR model was used to try to reproduce broad changes in opinion dynamics of Northern Ireland residents during ‘the troubles’ by comparing outputs with election results.
Aptima’s model cites Salzarulo’s work and its use of bounded confidence is similar.

Salzarulo’s Metacontrast model also focuses on opinion dynamics (Salzarulo, 2006). Agents have positions on a single issue, with a continuous number representing their opinion. Similar to the SCIPR model, each agent has a bounded confidence which affects who the agent will listen to and how much they may be swayed by an alternate position. Persuasion in this model is equivalent to an agent moves along the continuum of opinions toward a different position held by another agent. A unique feature of Salzuro is that the model includes both attraction and repulsion forces; agents move toward ‘identities’, or opinion positions that they want to join, and also try to move away from opinion positions that they define themselves against.

Salzarulo uses the principle of meta-contrast from social categorization theory (Haslam & Turner, 1992; Turner et al., 1987) to judge similarity and cohesion of identity groups. Groups (Salzarulo calls them categories) form when a cluster of agents perceive that the differences between them are small, and the distance between them as a group and other individuals in a group is large. More precisely, agents calculate the mean pairwise difference between all individuals in the model and compare it to the mean different pairwise differences to a subset of agents that form a candidate group. Groups form from clusters with a low ratio of group differences to context differences.

Once these groups form, agents act to reinforce group membership. Groups observe which individuals are most central, or prototypical of the group, and move to reduce differences between themselves and their group prototype. At the same time agents seek to maximize the difference between themselves and agents outside of the group. This is consistent with prior psychological studies of identity dynamics (Tajfel & Turner, 1986); ingroups often consolidate their identity by trying to clearly differentiate themselves from other groups, referred to as ‘outgroups’.

The Salzuro model produces three interesting effects that may be particularly useful for modelers. First, it produces polarization of opinions between groups. Because Salzarulo’s agents actively change opinions to move away from outgroups and toward the center of ingroups, they can result in groups clustered at the extreme ends of an opinion continuum, although this does not always happen. Polarization clearly happens in the real world, but often fails to happen in other influence models where over time agents become homogenized; Salzarulo provides a plausible mechanism for polarization to occur.

Second, Salzuro’s model produces an effect where agents whose opinions are prototypical of their identity group have very high confidence in their opinions. Because other agents in the group are moving toward them as central figures, and no force is pulling them away from their own center, the confidence of prototypical agents increases. This again corresponds to the real-world observation: group leaders tend to be very certain of their opinions. Salzarulo does not use the term ‘leader’; his model speaks only of more- or less-prototypical members; but it would be a natural extension to use his mechanisms to name these prototypical members as group leaders, and use these mechanisms to explain (at least partially) observed high levels of leader confidence.

Third, Salzarulo introduces a mechanism for context to affect identity. Salzuro’s explorations show that the formation and differentiation of groups in the metacontrast model are strongly influenced by the profile of agents in the initial model, i.e. the social context. Salzarulo’s explorations do not take the next step of varying the context within model runs, but one can easily imagine changing context within a larger model and observing the resulting effects on identity. This could model the strengthening of identity in a context where that identity is the minority; e.g. the previous example of Nigerian expatriates in the US context feeling a strengthening national identity.

Salzarulo’s work is a pure modeling effort, so has not (to our knowledge) been grounded or validated against real-world datasets.

Lustick’s PS-I model, is also focused on political opinions and persuasion, and particularly focused on regionally coherent ‘polities’, or identity groups. PS-I is intended as an open, general framework and has
been applied by the author (a Middle Eastern expert) in several settings. The example used in Lustick (2002) is a fictional country called ‘Middle Eastern Polity’ (MEP). MEP is represented by a rectangular grid populated with 2260 agents. Each agent represents a population aggregate, but behaves similarly to individuals in other models. There are 19 ‘identities’ present in the model, that vie for influence within and between agents. These are mostly political identities, but also include elements of religious, ethnic, and economic identities. Three examples are ‘Fundamentalist Islam’ (religious/political), ethnic Kurd (ethnic), and modernized Islam (religious/political). Each agent has a repertoire of 2-6 ‘identities’ that they hold. Only one identity is ‘active’ at a time, but the others may maintain lower levels of activation that are important to the model. A geographic cluster of agents with the same activated identity is referred to as a polity.

The pattern and initial activation levels of identities are how PS-I handles accessibility. The model also includes fit of contextual events. Model runs include disruptive, short-term events originating outside the model; e.g. a terrorist incident in a nearby country. The effects of these external events are determined by the existing pattern of activations moderated by tables of ‘bias’ specific to event types. These bias tables are what implement fit in PS-I. When the model runs, agents influence their neighbors and polities spread, shrink, or disappear across the landscape of the country. As in the other political influence models, similarity between agents determines influence. Lustick’s model also includes varied agent ‘personalities’ which are important to the influence dynamics but will not be described here.

PS-I has been used to study the volatility and common patterns of identities through simulated countries such as ‘Middle Eastern Polity’, and has also achieved some success validating against historical data. A focal point of study has been predicting regime instability. Other noteworthy strengths of this model are its ability to combine across identity types; and the ability to model of larger-scale geographic trends.

**SILAS** (Social Identity Look-Ahead Simulation), is in development by the authors at Johns Hopkins University Applied Physics Laboratory. SILAS focuses on identity-based conflicts. It attempts to predict how individuals with multiple identity affiliations will align in a conflict that may activate more than one of their identity groups.

The model includes two layers: individual agents (people), and abstracted identity groups. Each individual agent is modeled on single respondent to the Afrobarometer 2001 survey. Each agent was also given a synthetic social network of other agents based on assumptions about levels of cross-ethnic and cross-religious affiliations in Nigerian society (no data was available for this).

**Figure 1.1.** One individual (far right) and the network of identities joined by affinity links that the individual is affiliated with

Identities are modeled as objects that are separate from, but connected to individual agents by ‘affiliation’ links. Identities are arranged in separate hierarchies for three types: ethnic, religious, and political identities. Groups have affinity relationships with each other, both within and between hierarchies which are set by comembership data derived from Afrobarometer 2001 data. So, for example, the ‘affinity’ between the Hausa ethnic group identity...
and the Muslim religion was set to correspond to the percent of Hausa Afrobarometer respondents were Muslim. (This was an asymmetric network; the affinity from Muslim to Hausa corresponds to the percent of Muslims who are Hausa.)

Each individual in this model was affiliated with multiple identity groups, usually there was one ethnic, one religious, and one political affiliation. As a default, the weight (accessibility) of each affiliation link was set to 1 to indicate membership with a group. We used the Afrobarometer data on most-favored identities to increase this weight to 2 when such a response was given (recall that each individual in the model is based on an actual Afrobarometer respondent). These permanent affiliation weights are SILAS’s representation of identity accessibility.

Running the SILAS model begins with a conflict event between any two identity groups. (e.g. Muslim versus Christian; Igbo versus Ijaw; or Igbo versus Muslim). The groups do not have to be of the same type. The two groups in the conflict spread positive sentiment about themselves and negative sentiment about their opponent in the conflict. These sentiments spread through the abstracted identity model along affinity links. The strength of the affinity links was used as a multiplier of the strength of the sentiment. Sentiment, both positive and negative, spreads between identities and down to individuals. Spreading activation is limited to minimize feedback loops among identities. When the model is finished running, many individuals will have received positive and negative sentiment about the identities involved in the conflict. Some will have received both through separate channels, and will weigh the level of each to determine where they stand on the conflict. Some individuals will have received no sentiment messages, or equal positive and negative sentiment, and so will remain neutral. The temporary sentiment messages with varying levels of activation are how SILAS represents situational fit.

There are two notable features of SILAS. First, SILAS can predict the opinion of conflicted individuals; i.e. those that have identity links (direct or indirect) to multiple parties in a conflict, as described. A formative evaluation study used SILAS to predict political party affiliation based on an individual’s identity links. The model was constructed using known co-memberships, and then run on the same dataset with political affiliation links removed. (We chose to train on the entire set rather than reserve part for validation, because of the small cell size of some affinities). The conflict event was a simulated election between the three major political parties in Nigeria at the time of the 2001 Afrobarometer survey. The SILAS model correctly predicted 72% of known party affiliations. We compared this with a more conventional regression analysis, which predicted 76% correct. We were disappointed that SILAS did not outperform conventional regression, but pleased to be close. We hope to be able to improve the model with more highly localized data.

SILAS’s second notable feature is reproducing the ‘uniting against a common enemy’ effect. The hierarchical arrangement of identities allowed inference beyond stated groups, e.g. the model knew that an individual who self-identified as a Baptist was a Christian. In a conflict between a Catholic group and a Muslim group, the Baptist will receive stronger sentiment messages from the Catholic than the Muslim identity groups because of the shared Christian identity. The dynamic of uniting against a common enemy is well documented in the real world, but previous models did not necessarily reproduce it, or produce it only as a byproduct of other kinds of similarity.

We are seeking, but have not yet found a dataset that could be used to test the validity of the ‘common enemy’ effects. We are also seeking to extend the SILAS model to reproduce the corollary ‘sibling rivalry’ effect. Sibling rivalry is an effect where, in the absence of a common enemy, peer groups in a hierarchy may be particularly prone to conflict. This may be useful in modeling ‘horizontal inequality’ conflicts between peer groups, which Stewart (2000) argues are one of the most common types of conflicts in third-world countries. We have experimented with adding negative affinity, or ‘rivalry’ links between peers in the identity hierarchy, but initial runs with these in the model yielded unsatisfying results.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accessibility?</th>
<th>Fit?</th>
</tr>
</thead>
</table>
Modeling social identity is an important capability for valid socio-cultural models. The need to synthesize a broad and diverse literature and the need for new modeling techniques make this a difficult but also very interesting problem.

The models reviewed in this paper were all narrowly focused on a few identity-related issues. Future models will also need to integrate with a broader set of behavioral, economic, and political dynamics, which should be a focus of current research.

References


