

# Struggles Over Symbols and Endogenous Group Norms

**Eric S. Dickson**

**Associate Professor of Politics &  
Faculty Affiliate, Department of Psychology  
New York University  
19 West 4<sup>th</sup> Street, Room 306  
New York, NY 10012  
[eric.dickson@nyu.edu](mailto:eric.dickson@nyu.edu)**

EARLY DRAFT \* COMMENTS WELCOME

## **Abstract**

*Symbols associated with social group membership are frequent objects of contestation in the political sphere. What role do such symbols play in intergroup relations? What accounts for variation in the intensity of symbolic displays across space and time? What motivates some governments to attempt to regulate or ban the use of social identity symbols? This paper explores these questions within the context of a novel, behavioral game-theoretic model of social identity. This model grounds individual utility functions in social identity theory; individuals suffer “identity losses” when they fail to follow “behavioral prescriptions” associated with social group membership. The model considers the dynamics of a campaign of assimilation, within the context of an endogenous model of social group identification. In such a campaign, a government compels some social group members to violate a group norm regarding symbolic behavior, leading individuals to suffer identity losses. This tactic potentially weakens some individuals’ social identity attachments, but can also backfire from the government’s perspective by strengthening the social identity attachments of others.*

## 1. Introduction

Symbols and rituals associated with social group membership are frequent objects of contestation in the political sphere. In the context of contemporary global affairs, attempts to regulate use of the headscarf, veil, or other forms of Islamic dress in some countries – and attempts to mandate their use in others – have become emblematic both of divisions within Muslim countries and of conflicts between Muslim and non-Muslim populations in Western ones. The Turkish ban on the use of the headscarf in universities and government buildings has repeatedly proved a flash point in tensions between secular and religiously observant populations in Turkey (White 2002; McGoldrick 2006). A more recent ban on the use of the headscarf in French secondary schools is symptomatic of the unease with which many “native” French view large and growing Muslim populations within that country (McGoldrick 2006; Scott 2007). In Iran, use of the *chador* was banned during the 1930s, became a symbol of Islamist opposition to the regime of Mohammad Reza Shah in the 1970s, and became compulsory by law under Ayatollah Khomeini in 1982 (McGoldrick 2006). Of course, attempts to regulate or ban social identity rites and symbols extend far beyond this family of examples in the historical record; among many other examples, the U.S. Bureau of Indian Affairs banned or regulated Native American rites and cultural practices during the 19<sup>th</sup> Century (White 1991), while Jews and Muslims alike have been compelled to violate the ritual taboo on the consumption of pork in conflicts from Spain through the Levant to Cambodia and China over the past two millennia (Kiernan 2002; Gladney 2004).

The prevalence and intensity of struggles in the political sphere over such symbols and symbolic behavior raise several key questions. What role do social identity symbols play in intergroup relations? What accounts for variation in the intensity of symbolic displays across space and time? What motivates some governments to attempt to regulate or ban the use of social identity symbols? This paper explores these questions through the development of a novel behavioral game-theoretic framework.

In this framework, actors' emotional investments in their social identities are explicitly modeled and play a central role. The model of identity employed in the paper is based on core principles of social identity theory (Tajfel and Turner 1979, 1986; Tajfel 1981; Turner 1984). Within social identity theory, individuals are taken to possess a sense of self or ego that is defined on both an individual and collective basis. The self is constructed via a process of identification in which one associates oneself with other individuals in one's social categories, and differentiates oneself from nonmembers. An individual's efforts to maintain and enhance self-esteem are understood to be key motivations for behavior; when individuals' social identities are salient, self-esteem is greatly influenced and maintained by individuals' social contexts and the categories or roles they fill in their environment.

Among the roles played by social groups is a coordination of expectations among group members as to which behaviors are appropriate or mandatory, and which behaviors are inappropriate or forbidden. Individual self-esteem and perceived well-being are bound up in social identities; and a central part of maintaining social self-esteem is the internalization and obedience of such group norms, or "behavioral prescriptions." Such prescriptions help to define a group's identity, and an individual's observance of them helps maintain her sense of self. Behavioral prescriptions can require positive action, or they can stipulate prohibitions.

Behavioral prescriptions are generally taken to influence individual behavior through an essentially affective causal mechanism; because group members internalize group norms, individuals suffer emotional anxiety and a reduced level of self-esteem when they violate behavioral prescriptions that specify how group members "should" behave. Internalized group norms of this kind can serve a role in promoting group cohesion and coordinated action.

The framework presented in this paper explores one particular mechanism through which social identity symbols<sup>1</sup> may come to be objects of contestation in the political sphere. Within this framework, a government or other agent can choose to regulate or tax use of a social identity symbol, in an attempt to “deactivate” or weaken the social identities of out-group members, or even to conclude a wholesale campaign of assimilation. In order to explore the potential effectiveness of such efforts, the paper first develops a framework in which individual levels of identification with a social group are endogenous, reflecting both individual characteristics as well as structural features of the political economy in which the social group exists. Attempts to regulate or tax the use of social identity symbols can then be understood as attempts to change those structural features of a social group’s environment, leading in turn to outcomes of identity change.

The mechanism through which such change can potentially take place relies on the internalization of social group norms by group members. By manipulating the costs of honoring a behavioral prescription for group membership, a government can increase the likelihood that group members will violate such prescriptions. In the limiting case, governments<sup>2</sup> can even compel such violations using physical force. Such acts of compulsion lead to negative emotional consequences (that is, “identity losses”) for group members, because the duty to follow behavioral prescriptions is internalized. In a setting where identification is endogenous, such acts may therefore insert a psychological wedge between an individual and his or her group identification; one potential way in which the resulting anxiety can be assuaged is through a weakening of one’s sense of social identity, or ultimately a process of identity change.

The structure of the remainder of the paper is as follows. Section 2 offers motivation for the paper’s overall objectives in the context of illustrative historical cases.

---

<sup>1</sup> Except where it is otherwise clear from context, the term “symbols” will be used in a general sense that is also meant to subsume publicly-observable participation in rites or honoring of taboos.

<sup>2</sup> The generic term “government” is to be understood as referring to any sufficiently powerful agent or organization with the ability to regulate behavior of an out-group’s membership. Such agents will often be, but need not always be, actual governments.

Section 3 then presents a framework for incorporating social identities into game-theoretic models, based on the work of Akerlof and Kranton (2000). Section 4 puts this framework to use in developing a simple model in which the extent of individual identification with a social group is endogenously determined, given fixed individual characteristics and structural features of the environment. Section 5 then explores the effects that a government's attempts to regulate or tax social identity symbols may have on the extent of individual identification with the social group. Section 6 offers a brief conclusion.

## 2. Motivation and Historical Examples

This paper presents a simple model of a campaign of *assimilation*, through which a government or other authority may attempt either to reduce the intensity of out-group members' social identity commitments or even to provoke a process through which individuals ultimately change their identity self-categorizations. Such an account resonates in several key respects with both the French and Turkish headscarf bans. Traditional conceptions of citizenship in French political thought emphasize an indivisible national identity, deeply distrustful of other kinds of group commitments (Scott 2007). Similarly, the headscarf ban in Turkey is historically resonant with Ataturk's campaign of advancing a new, secular Turkish national identity (McGoldrick 2006; White 2002) to the exclusion of other ethnic and religious social identities in Turkey.<sup>3</sup>

These attempts to deter individuals from fulfilling their internalized "behavioral prescriptions" can be set beside many other historical examples, in which members of one group either *force* out-group members to carry out an action that is *contrary* to the out-group's norms, or *prevent* out-group members from carrying out actions that are seen as *central* to the out-group's identity. While in particular instances such tactics may

---

<sup>3</sup> The substantive emphases on bans in school settings in both cases is particularly suggestive, given the likely greater degree of malleability in individual senses of identity during youth.

spring from multiple distinct motivations, there can be little doubt that denigration of the social identities of out-group members *per se* – with an intention of provoking changes in the strength and nature of opponents’ group attachments – is a common thread running through a broad range of historical cases.

One excellent example comes in the conquest of the American West and associated attempts to assimilate Native Americans into the prevailing national culture. A key part of the Bureau of Indian Affairs’ assimilation strategy was the systematic repression of essential Native American cultural and religious practices, “by persuasion when possible and by force when necessary.” (White 1991, p. 109) In 1884, for example, the government banned the Sun Dance among the Lakotas, while in 1888, it enforced a prohibition against “bundles,” in which locks of hair from a deceased family member were ritually preserved for a year, binding the deceased’s spirit to the community before ultimately being released in a ceremony. (White 1991, p. 112) Being forbidden from carrying out such crucial cultural obligations took a predictable emotional toll on Native Americans. In the words of the Lakota religious leader Short Bull, “The white people made war on the Lakotas to keep them from practicing their religion... The white people wish to make us cause the spirits of our dead to be ashamed.” (White 1991, p. 113) Such a sense of shame doubtless extended to the living as well. An unquestionable intention of such tactics was to weaken Native Americans’ social and cultural identities, with an ambition that ultimately “Indians would be Christian farmers living in nuclear families on their own land.” (White 1991, p. 111) Preventing Native Americans from carrying out ceremonies and other behavioral prescriptions central to their group identities was a key strategy in this attempt at assimilation.

In other settings, a comparable objective has been pursued by forcing individuals to *take* actions explicitly *forbidden* by their social groups’ behavioral prescriptions, rather than by preventing them from taking actions considered obligatory. A prominent set of examples involves the forced consumption of foodstuffs that are considered ritually unclean and out of bounds for adherents of particular faiths. The consumption of pork, of course, is a taboo both among Jews who observe Kashrut law (Leviticus 11:7-8;

Deuteronomy 14:8) and among Muslims who honor halal laws (in the Qur'an, at al-Baqara 2:173; al-Ma'ida 5:3; al-An'am 6:145; an-Nahl 16:115). A recurring method of persecution that has been used against adherents of both of these faiths through the ages has been to compel them to eat pork, under penalties as severe as death if they failed to do so. Among the oldest historical accounts of this tactic comes from Biblical times; the scribe Eleazar is described as choosing to be executed instead of following his captors' orders to consume pork during a period of anti-Jewish persecution (2 Maccabees 6:18-31). Similar stories can be found sprinkled throughout the historical record, notably during the anti-Jewish pogroms of the Spanish Inquisition. Muslims have also been subject to comparable social identity attacks, numerous times. One violent and disturbingly recent such instance took place in Cambodia, where the Khmer Rouge forced Cham Muslims to eat pork, often under threat of death, as part of their program to eradicate the Muslim identity in Cambodia during the 1970s (Kiernan 2002). Similar tactics have been employed during periods of ethnic conflict in China, where Muslim minorities have been compelled to consume pork by Han Chinese (Gladney 2004) as part of campaigns to weaken the minorities' religious and cultural identities. Such tactics are sometimes – though by no means always – effective in achieving goals of forced assimilation; Gladney reports a family history in which, after a period of persecution many years before, “we [the family] began to eat pork and became assimilated to the Han.” (p. 115)

These historical cases provide but a sampling of the social identity conflicts in which members of one group have attempted to weaken the social identities of out-group members by compelling them to violate out-group norms, in either a positive or in a negative sense. While many of these examples are particularly extreme, in the sweeping extent of their assimilationist motivations as well as in the degree to which many of these campaigns relied on extreme physical brutality, similar tactics can be and have been employed in milder forms as well, and towards more modest objectives. For example, in some settings social identity conversions were encouraged with economic incentives rather than compelled by force; a typical example involves fifteenth- and sixteenth-century tax exemptions offered by Muslim Ottomans to any Christians in Limnos who

chose to convert to Islam (Lowry 1986). Even under these less brutalistic circumstances, however, social life was nonetheless structured so as to impose higher relative costs on individuals who maintained the behavioral norms of their groups' memberships, in the hopes that these incentives would change the behavior of at least some people, and eventually their social self-identifications as well. Clearly, whatever the form it takes and whatever the intensity with which the strategy is pursued, attempts to induce identity change by compelling individuals to violate social identity behavioral prescriptions are a recurring theme across a broad universe of religious and ethnic conflicts.

By what mechanism can such campaigns of compelled assimilation sometimes be effective? The very objective of such campaigns – identity change – suggest the utility of a behavioral game-theoretic model, in which identities are not only modeled explicitly but also allowed endogenously to adjust in the context of the strategic environment. A “classical” game-theoretic setting seems deficient for the exploration of such questions. To the extent that social identities could be modeled at all in such a context, they would most naturally enter through the ways in which they affect individual preferences; however, in classical game-theoretic models, individual preferences are taken to be exogenous and fixed. Through employing a behavioral game-theoretic method, it is possible to explore processes of identity (and preference) change, and explicitly to model the key causal role that emotions may play in processes underlying campaigns of compelled identity change. At the same time, the incorporation of emotions into a framework of strategic interaction offers the potential for novel insights into the underlying dynamics of such campaigns. Such an approach not only offers the promise of more psychologically realistic formal models of politics, but also provides an analytic framework that makes it easier to ask certain kinds of questions that psychological models may find it difficult to answer on their own, for example, under what circumstances such social identity attacks may be more or less likely to be effective.

The central causal mechanism advanced in this section suggests that while the establishment of such group norms may deepen and extend a given social group's capacity for coordinated social or political action, it can also expose an Achilles' heel that

can sometimes be exploited by out-group adversaries who may wish to attack group members' social identity affiliations.

Group members may obey behavioral prescriptions in order to avoid the negative emotional consequences of violating group norms; but if these same group members are *compelled* to violate these behavioral prescriptions, then the negative emotional consequences can become actual rather than hypothetical as they are experienced on the equilibrium path. Such an act can potentially insert a psychological wedge between an individual and his or her group affiliations, which may ultimately lead to a weakened sense of social identity or even identity change.

Consider a situation in which an individual has been compelled by force to violate a behavioral prescription in the manner described above. How is the individual to respond to the resulting identity losses? The model here focuses on two salient alternatives. In the first alternative, the individual simply deals with and suffers this emotional cost and reduced sense of self. In this scenario, an individual's social identity remains intact, despite the fact that it is now associated with emotional losses.

In the second alternative, the individual attempts to avoid or minimize this emotional distress by *reformulating* his social identity attachments. The potential plausibility of such a mechanism of identity change is built, in a sense, into the fabric of social identity theory itself. Individuals develop social identities in part to create and maintain self-esteem in a social context. If a particular social identity becomes sufficiently associated with a sense of emotional anxiety, the imperative to develop a positive sense of self may sometimes drive the individual to redefine himself in his social context. Consider the first motivating example in the previous section. A Native American with a Lakota social identity would, in part, derive her sense of self-esteem from carrying out the rituals considered obligatory in her society. If prevented from carrying out these rituals by an outside force, she would feel a sense of shame and anxiety, and her social identity would be devalued. One way of dealing with this situation would be to de-emphasize her Lakota identity in her self-definition – just as the

Bureau of Indian Affairs would have wanted. Similarly, the forced consumption of pork against an individual's religious dietary restrictions could potentially invoke a sense of anxiety and shame, within the context of the individual's system of religious beliefs. One method of minimizing this sense of uncleanness or shame would be to redefine her religious values, possibly putting her at greater remove from group religious norms, or even to lapse in her religious observation or to convert to another religion under which the act committed is not considered shameful. Naturally, the likelihood that acts attempting to induce forced conversions will be successful varies depending the context.

As such, forcing members of a group to violate a behavioral prescription of that group can insert an emotional wedge between group members and their social identities, which could plausibly lead in some individuals and in some circumstances to a process of identity change. From one perspective, this causal mechanism is somewhat ironic; the anxiety generated by violating group norms, one of the key bits of social glue that can help make groups cohesive, is in this account liable to be exploited by out-group members who wish to *weaken* the group in question.

### **3. The Akerlof-Kranton Model of Social Identity**

This paper conceives of individuals as strategic actors whose utility functions incorporate social identity considerations.<sup>4</sup> Individuals' social identities offer psychological benefits, conceptualized as emotional in nature, in a way that depends among other things on the actions an individual takes and whether or not these actions comport with expected behavior of group members. Incorporating social identity into utility functions in this way makes it possible to consider traditional strategic incentives and emotional ones within a unified framework.

The paper follows Akerlof and Kranton (2000) by adopting a utility function of the following general form:

---

<sup>4</sup> The discussion of the Akerlof-Kranton model and related material in the paper are adapted from Dickson and Scheve (2006).

$$U_i = U_i(\mathbf{a}_i, \mathbf{a}_{-i}, I_i)$$

where individual  $i$ 's utility depends on her actions,  $\mathbf{a}_i$ ; on the actions of other individuals,  $\mathbf{a}_{-i}$ ; but also, unlike in standard game-theoretic models, on  $i$ 's identity or self-image,  $I_i$ . The Akerlof-Kranton model of identity is based on the assignment of social categories. Individuals place themselves and others in society in some finite set of categories,  $\mathbf{C}$ . Let  $\mathbf{c}_i$  be a mapping for individual  $i$  assigning the set of all individuals,  $\mathbf{F}$ , to categories in  $\mathbf{C}$  ( $\mathbf{c}_i: \mathbf{F} \rightarrow \mathbf{C}$ ). Crucially, social categories may be associated with *behavioral prescriptions*  $\mathbf{P}$ , which are sets of actions (or characteristics) deemed appropriate for individuals in given social categories. Finally, individuals are endowed with basic characteristics,  $\varepsilon_i$ , that are not *a priori* assumed to be correlated with social categories. Identity payoffs are then represented as:

$$I_i = I_i(\mathbf{a}_i, \mathbf{a}_{-i}; \mathbf{c}_i, \varepsilon_i, \mathbf{P})$$

In the Akerlof-Kranton framework, a person's identity depends on his or her social categories assigned by  $\mathbf{c}_i$ , which may be exogenous and fixed or endogenously chosen. Identity is also allowed to be a function of the extent to which an individual's own characteristics,  $\varepsilon_i$ , match any ideal characteristics, defined by  $\mathbf{P}$ , associated with the social categories to which he or she is assigned. Most relevant here, identity payoffs may also depend on the extent to which an individual's own actions,  $\mathbf{a}_i$ , and the actions of other actors,  $\mathbf{a}_{-i}$ , correspond to the behavioral prescriptions for social categories, also defined by  $\mathbf{P}$ . The violation of prescriptions associated with social categories is thought to generate emotional anxiety, because individuals internalize relevant group norms, and thus identity losses.

#### 4. A Model of Endogenous Social Identity Payoffs

This section presents a model in which an individual's identity-related payoffs ( $I_i$ ) and social self-categorization (a component of  $\mathbf{c}_i$ ) are endogenously determined. The

model takes into account three aspects of group life, each of which influences the equilibrium level of  $I_i$  in this framework: the production of group public goods, the use of social identity symbols signifying group membership, and the potential for commitments to a group to involve unforeseen entanglements.

#### **4.1. Social Identities, Collective Action, and Group Public Goods**

A large literature in political science has focused on problems of collective action, including models of political participation. Collective action problems are central to political life; individuals can gain access to many of the goods available through political processes only in the context of coordinated action by a group of which the individuals are members. This section applies insights from the Akerlof-Kranton model in the previous section and considers the role that may be played by behavioral prescriptions in facilitating collective action. Indeed, behavioral prescriptions of this kind are quite common. In the context of present-day tensions between social identity groups, the idea of *jihad* within Islam is among the most prominent that fits within this rubric; while, in the history of Islamic thought, this idea has connotations of internal struggle as well as of social and collective action, a key dimension of *jihad* clearly involves a duty, presumed to be binding on believers, to participate in campaigns of collective action under appropriate circumstances. More generally, social groups face a variety of collective action problems, many of them involving the provision of group public goods such as security or political participation.<sup>5</sup> In many contexts, individual participation in public goods provision can be thought of a group norm that is supported by a behavioral prescription for group membership.

In traditional analyses of public goods provision, the key difficulty in encouraging individual contributions is the lack of a direct incentive

---

<sup>5</sup> Although generally not posed in social identity terms, the idea of “duty” as motivating political participation of course has a long lineage in the literature. In the formal literature, this lineage extends back at least as far as the “D term” of Riker and Ordeshook (1968).

to contribute; that is, any return that an individual personally receives from her own contribution decision is insufficient to meet the cost of contributing. The approach taken here is to note that social identities have the potential to motivate even the most myopic actors to do their part in providing public goods despite the costs of doing so.

Specifically, the model here considers a binary public goods contribution decision. It is assumed that a(n exogenous) behavioral prescription directs group members to contribute to the public good. A decision to contribute has direct cost  $c_i(\varepsilon_i) > 0$  for individual  $i$ ; at the same time, individual  $i$  enjoys benefit  $r_i(\varepsilon_i) > 0$  as a direct result of her own contribution.<sup>6</sup> In keeping with the standard public goods framework,  $c_i(\varepsilon_i) > r_i(\varepsilon_i)$ . If instead individual  $i$  chooses not to contribute, she bears the internalized psychological cost  $I_i$  for violating the behavioral prescription to play her part in providing the public good.

Throughout, the paper models individual group members as myopic actors whose behavior is proximally affected by identity-related factors. With respect to public goods provision, this perspective can be interpreted to mean that individuals are motivated to contribute in accordance with an internalized norm, rather than because of other factors of the larger environment, such as an individual's strategic potential to influence the contributions of others through contributing, or a fear that a decision not to contribute may ultimately be discovered and punished by others. Consistent with this perspective, an individual here chooses to provide the public good if and only if the direct utility from doing so exceeds the direct utility from failing to do so<sup>7</sup>, which can be expressed as follows:

---

<sup>6</sup> Here and from here on, the interpretation of  $(\varepsilon_i)$  is broadened from that of Akerlof and Kranton, to include individual characteristics that may be relevant for economic production, etc., in addition to representing the closeness of fit between an individual and her group's "ideal type."

<sup>7</sup> Throughout, it is assumed for convenience that individuals who are indifferent between courses of action (e.g., whether or not to follow a behavioral prescription, whether or not to maintain self-categorization in their "current" group) choose the action more in accordance with group expectations (e.g., to follow the behavioral prescription, to maintain self-categorization in their "current" group).

$$I_i(\varepsilon_i) \geq c_i(\varepsilon_i) - r_i(\varepsilon_i)$$

While these direct costs and benefits determine the behavior of myopic group members, in the larger picture these individuals may of course also benefit from the public goods contributions of others. Specifically, we denote the potential benefit to individual  $i$  of contributions by  $N$  other group members as  $B_i(\varepsilon_i; N)$ . Of course, in many settings, the access of group members to the benefits of group membership may be made contingent on individual behavior, through the *social* enforcement of group norms (as opposed to the way in which internalized norms may internally influence behavior, as described above). This feature of group life is captured by an assumption that, in expectation, an individual loses fraction  $\gamma_{PG}$  of the benefit  $B_i(\varepsilon_i; N)$  if she fails to contribute to the public good, where  $0 \leq \gamma_{PG} \leq 1$ . If  $\gamma_{PG} = 0$ , the group public good is non-excludable with respect to the individual's contribution decision; if  $\gamma_{PG} = 1$ , it is fully excludable with respect to this decision, and if  $0 < \gamma_{PG} < 1$ , it is partially excludable with respect to this decision. As described below, the group public good may also be (partially) excludable with respect to individuals' potential non-use of symbols of group membership. The possibility that access to group benefits may be excludable through the social enforcement of group norms is taken here to be an exogenous fact of life; the decisions underlying such enforcement are not explicitly modeled.

## 4.2. Symbols Signifying Group Membership

As noted in the motivating examples, many social groups are associated with particular symbols or rituals signifying group membership. The model here considers a simple binary decision over whether or not to participate in some form of symbolic behavior, such as the use of identifying religious garb. As in the case of public goods provision, it is assumed that a(n exogenous) behavioral prescription directs group members to participate in the symbolic behavior.

A decision to participate in the symbolic behavior has net direct cost  $k_i(\varepsilon_i)$ , which is interpreted as including both the practical costs of the symbolic behavior as well as any interpersonal costs or interpersonal benefits that directly result from the symbolic behavior.<sup>8</sup> As such,  $k_i(\varepsilon_i)$  may be either positive or negative, with a negative value indicating an overall net direct benefit to engaging in the symbolic behavior. These practical and interpersonal costs and benefits are understood as being measured relative to the practical and interpersonal costs and benefits of non-participation in symbolic behavior, which are normalized to zero. However, these practical and interpersonal factors do not include the psychological costs of failing to conform to group norms; individuals who choose not to participate in the symbolic behavior suffer identity-related losses  $I_i(\varepsilon_i)$  for failing to fulfill their group's behavioral prescription to participate in symbolic behavior.

Given these definitions, individual group members who are myopic in the sense described above will participate in the prescribed symbolic behavior if and only if:

$$I_i(\varepsilon_i) \geq k_i(\varepsilon_i)$$

Finally, it is supposed that non-participation in symbolic behavior may be associated with partial exclusion from the benefits of the public good described earlier, just as partial exclusion was possible for decisions not to contribute to the public good. Specifically, failure to participate in symbolic behavior is associated in expectation with the loss of fraction  $\gamma_S$  of the benefits of the public good, where  $0 \leq \gamma_S \leq 1$ . It is assumed that the two forms of exclusion described by  $\gamma_{PG}$  and  $\gamma_S$  are additively related, so that an

---

<sup>8</sup> As an example of what is meant by "directly result," decreased access to employment opportunities by virtue of engaging in the symbolic behavior would count as a direct interpersonal cost, while increased access to employment opportunities by virtue of engaging in the symbolic behavior would count as a direct interpersonal benefit. Such interpersonal costs and benefits are features of the social and economic environment in which individual group members find themselves; these are taken to be exogenous to the model. Note that potential exclusion from the group public good based on failure to engage in symbolic behavior is not counted as a "direct" cost, and is considered separately below.

individual who fails both to contribute to the public good and to participate in symbolic behavior receives fraction  $1 - \gamma_{PG} - \gamma_S$  of the public good's potential benefits. This of course is true only so long as  $1 - \gamma_{PG} - \gamma_S \geq 0$ ; if  $1 - \gamma_{PG} - \gamma_S < 0$ , the individual simply receives zero benefits from others' contributions to the public good.

### **4.3. The Potential for Commitments to a Group to Involve Unforeseen Entanglements**

Throughout, it has been assumed that individual group members are myopic actors who can be motivated to engage in pro-social behaviors through their social identities; an individual  $i$  suffers identity losses  $I_i(\epsilon_i)$  when she fails to conform to a behavioral prescription for group membership. While it is easy to imagine that myopic individuals may sometimes be well-served by internalizing group norms in this way, such an attachment to one's social identity group also has a potential down side. Group elites with the ability to shape behavioral prescriptions may use this power in a way inconsistent with individual interests, as when leaders in pursuit of power or their own private interests whip up ethnic fervor that leads to counterproductive intergroup conflict. In a complicated, changing world, a form of agency in which individuals suffer identity losses when failing to follow group norms exposes individuals to the risk that the group norms that seem salient tomorrow may entail unforeseen costs.

This potential down side to a more intense commitment to a group – that is to say, to a higher value of  $I_i(\epsilon_i)$  – is captured in the model in the following way. Suppose that a behavioral prescription exists indicating that group members should undertake an action of cost  $z$ , where  $z$  is a random variable uniformly distributed between 0 and  $Z$ , and the value  $Z$  is assumed to be higher than the maximum value of  $I_i$  that will ever be observed in practice. Individuals who undertake the action receive no benefit. Given this, a myopic individual for whom  $I_i(\epsilon_i) < z$  will choose not to undertake the action, and suffer identity losses  $I_i(\epsilon_i)$ , while a myopic individual for whom  $I_i(\epsilon_i) \geq z$  will instead undertake the action, paying cost  $z$ . Integrating over the distribution of costs  $z$ , in expectation an individual pays a cost of “unexpected entanglements” equal to  $I_i(\epsilon_i) (Z - \frac{1}{2}I_i(\epsilon_i))$ . Because

$Z > I_i(\varepsilon_i)$ , this expression is increasing in  $I_i(\varepsilon_i)$ ; straightforwardly, individuals who have lower values of  $I_i(\varepsilon_i)$  undertake the costly action for a smaller range of costs  $z$  (and suffer lower identity losses when failing to undertake the action for larger  $z$ ). Thus, individuals with lower values of  $I_i(\varepsilon_i)$  suffer lower expected costs from the unforeseen entanglements that stem from commitments to a group.

#### **4.4. The Endogenous Generation of Identity-Related Payoffs**

The above expressions for individual behavior include a number of parameters – for example,  $c_i(\varepsilon_i)$ ,  $r_i(\varepsilon_i)$ ,  $\gamma_{PG}$ ,  $\gamma_S$ , and  $k_i(\varepsilon_i)$  – that reflect the structure of economic and social interactions within a particular group. In contrast,  $I_i(\varepsilon_i)$  represents a fundamentally psychological quantity, one that helps to shape individual decision making within this economic and social context.

This paper pursues an intuition that the nature and extent of individual identification with social groups can, in turn, be shaped by structural factors. In particular, the approach adopted here endogenizes the intensity with which individuals internalize group norms, captured above by the “identity losses”  $I_i(\varepsilon_i)$  suffered by individuals who violate these norms.

The logic of the approach is as follows. Because  $I_i$  proximally influences decision making, different values of  $I_i$  have different implications for individual outcomes within the larger environment. The model supposes that, in an evolutionary sense, an individual's  $I_i$  will eventually drift towards the value that would optimize her decision making within the larger context in which she finds herself. As such, the equilibrium values of  $I_i(\varepsilon_i)$  derived here are to be thought of as the long-run outcomes of some less-than-fully-conscious adjustment process, rather than as the objects of conscious and instantaneous choice by group members. Indeed, individual decision making in the present, as described above, is influenced by individual identities in a binding way. However, over the long run, the nature of these identities adjust in a way that makes them well-suited to individual social circumstances. This feature of the model – that identities

are binding or “sticky” in the short run, but adjust over the long run – should be kept carefully in mind when considering the analysis to follow.

The previous section described three ways in which identities affect outcomes and decision making; the central contribution of this section will be to show how the balance of considerations thus described induces equilibrium values of  $I_i(\epsilon_i)$  that reflect underlying structural realities. On the one hand, identities can shape pro-social behavior in a way that leaves myopic individuals (and the groups of which these individuals are members) ultimately better off – in this case, by providing a proximal motivation for individuals to undertake costly contributions to a public good, a task that would not be carried out under myopic self interest as depicted above. At the same time, increasingly intense identity attachments (higher and higher values of  $I_i$ ) increase the downside risk that individual commitments to carrying out group norms may lead to costly losses through unforeseen entanglements.

As suggested by the inequalities in the previous section, the nature of individual behavior – and therefore the nature of individual payoffs – varies across a few key thresholds as  $I_i$  increases. The details of this dependence on  $I_i$  vary according to the structural parameters of the model. Figure 1 depicts this dependence for a particular set of parameter values. In particular, in the setting relevant to Figure 1,  $c_i(\epsilon_i) - r_i(\epsilon_i) > k_i(\epsilon_i) > 0$  – participating in symbolic behavior has a positive net direct cost, but has lower net direct cost than public goods provision – and the benefits of the public good are partially excludable based both on decisions regarding public goods contributions and on symbolic behavior.

\* Figure 1 About Here \*

In such a setting, when  $I_i = 0$ , individuals engage neither in symbolic behavior nor in public goods contributions. Individual payoffs then simply reflect the non-excludable portion of the group public good. As the value of  $I_i$  increases, individual payoffs initially fall. This is true for two reasons. First, higher values of  $I_i$  make individuals susceptible

to more costly unforeseen entanglements without offering any compensating benefits. And second, higher values of  $I_i$  increase the identity losses suffered because the individual has failed to carry out the behavioral prescriptions associated with group membership.

Individual utility discontinuously increases, however, as  $I_i$  increases past  $I_i = k_i$ . For values of  $I_i$  larger than  $k_i$ , a group member now participates in the prescribed symbolic behavior. This has two important implications. First, the individual is no longer partially excluded from the benefits of the public good on the grounds of failing to engage in the relevant symbolic behavior; and second, as  $I_i$  continues to increase beyond  $k_i$ , a failure to employ the group symbol no longer places increasing downward pressure on individual utility. Nonetheless, in the region where  $I_i$  falls between  $k_i$  and  $c_i - r_i$ , individual utility continues to decrease, both because of the rising expected cost of unforeseen entanglements and because of increasing identity losses suffered from failure to carry out the behavioral prescription mandating the contribution to public goods.

The group member's utility again discontinuously increases at  $c_i - r_i$ , for parallel reasons to those behind the increase at  $k_i$ ; for  $I_i \geq c_i - r_i$ , the individual contributes to the public good, removing the final source of partial exclusion from its benefits. And, individual utility once again resumes a downward trajectory past  $I_i = c_i - r_i$ , because of the ever-increasing expected cost of unforeseen entanglements.

Similar figures can easily be derived for any values of the model's parameters. With the intuitions from Figure 1 in mind, the remainder of this section instead presents several general results concerning the intensity of identity attachments (values of  $I_i(\varepsilon_i)$ ) in equilibrium and the relationship these intensities and the structural parameters of the model.<sup>9</sup> As noted above, at any given time, individual behavior is determined by the current value of  $I_i$ . However, in the long run, the intensity of group identification has the ability to adjust – as does, potentially, an individual's self-categorization. If an individual continues to self-categorize as a member of her original group,  $I_i$  adjusts to a value that

---

<sup>9</sup> The proofs corresponding to these results are omitted from this version of the paper.

maximizes individual utility<sup>10</sup>; this equilibrium value is denoted  $I_i^*(\varepsilon_i)$ . If an individual would experience higher utility by ceasing to self-categorize as a member of the group at all, the individual accordingly changes her self-categorization; this is the case only if an individual's "outside value"  $Q_i(\varepsilon_i)$  – an exogenous utility that black boxes the relative attractiveness of abandoning self-categorization in the original social group – exceeds the maximum utility the individual could receive by retaining her original self-categorization.<sup>11</sup>

**Result 1.** *In equilibrium, four outcomes are possible for individual  $i$ :*

- (1)  *$i$  may self-categorize as a member of the group, with  $I_i^*(\varepsilon_i) = c_i - r_i$ ;*
- (2)  *$i$  may self-categorize as a member of the group, with  $I_i^*(\varepsilon_i) = k_i$ ;*
- (3)  *$i$  may self-categorize as a member of the group, with  $I_i^*(\varepsilon_i) = 0$ ;*
- (4)  *$i$  may not self-categorize as a member of the group at all.*

Figure 1 communicates the most basic intuition behind this result. In the Figure, utility is a decreasing function of  $I_i$  everywhere aside from the discontinuous jumps at  $I_i = c_i - r_i$  and  $I_i = k_i$ . This leaves a utility function with three local maxima, corresponding to outcomes (1), (2), and (3). If  $i$  continues to self-categorize as a member of her initial group, she must therefore have a value of  $I_i^*(\varepsilon_i)$  corresponding to one of these local maxima. Of course, in Figure 1, one of these local maxima is also a global maximum, so in this case the equilibrium value of  $I_i^*(\varepsilon_i)$  is unique. However, for different values of the model's parameters, it may be the case that any of the three local maxima corresponding

---

<sup>10</sup> It is well worth noting that the relevant individual utility functions that are maximized over time include the identity-related payoffs experienced by individuals, consistent with the logic of the mechanism described in Section 2. As such, the maximand is distinct from a biological fitness, which would make reference only to material parameters. Instead, the idea here is that individuals over time adjust their identities in a way that maximizes experienced utility, which of course includes the identity-related payoffs themselves.

<sup>11</sup> For example, an individual whose characteristics allow her to "pass" as a member of a different social group may have a higher value of  $Q_i(\varepsilon_i)$  than another individual whose characteristics do not allow her to "pass." Intuitively, the first individual may be more able to enjoy benefits from self-categorization in another group, thereby giving her a higher "outside value."

to these different values of  $I_i$  is the largest, so that in general, there remain three possible outcomes where self-categorization as a group member is retained. Alternatively, if the “outside value” of leaving the group is sufficiently high – that is, if the utility she would experience by doing so exceeds the utility she could experience as a group member, for any value of  $I_i$  – she may cease to self-categorize as a group member at all.

The Result indicates that the intuition garnered from Figure 1 carries over to all possible sets of parameter values allowed by the model, in the sense that only these four outcomes are possible regardless of the parameter values. However, it is not the case that *all four* outcomes are possible for *any given* set of parameter values. One important result along these lines considers the nature of the group public good:

**Result 2.** When the group public good is fully non-excludable ( $\gamma_{PG} = \gamma_S = 0$ ), only two outcomes are possible for individual  $i$ :

- (1)  $i$  may self-categorize as a member of the group, with  $I_i^*(\epsilon_i) = 0$ ;
- (2)  $i$  may not self-categorize as a member of the group at all.

The intuition behind this result is that when the public good is not at all excludable, there is no discontinuous jump in individual utility as  $I_i$  increases. Instead, utility is simply a decreasing function of  $I_i$ ; whether or not an individual retains self-categorization as a group member simply depends on the individual’s “outside value.” As such, in equilibrium, an individual will never internalize group norms (in the sense of having positive  $I_i^*(\epsilon_i)$ ) unless the public good is at least partially excludable. If partial excludability is interpreted as resulting from some process of external enforcement, this result implies that the possibility of external enforcement is a necessary condition *for the internalization of group norms*.

## 5. Contesting Identities in Intergroup Conflict

The previous section developed a framework in which the intensity of an individual’s identification with a social group emerged endogenously as a function of

various structural parameters of a group's economic and social context (and of the individual's personal characteristics). The present section employs this framework to explore the intuitions that were latent in the historical examples and made explicit in the hypothesized causal mechanism of Section 2. Crucially, when a government targets a social group by regulating or placing a tax on the use of a social identity symbol, that government is transforming the structural parameters of the environment in which the group and its members must function. When a government attempts to affect a program of assimilation or identity attenuation using such a policy tool, should we expect it to achieve its ends, or may the government's tactics sometimes backfire, causing an *increase* in individuals' intensities of commitment to carrying out group norms?

In order to address this question, consider a pre-existing social group with the exogenous behavioral prescriptions given in the previous section, supposing that each individual  $i$  who self-identifies as a member of the social group has the equilibrium intensity of identification  $I_i^*(\varepsilon_i)$  that is appropriate given her personal characteristics and the operative values of the structural parameters. Then, suppose that a government transforms the situation by placing a tax  $T$  on use of the social identity symbol, so that individual  $i$  now pays net direct cost  $k_i(\varepsilon_i) + T$  (rather than  $k_i(\varepsilon_i)$ ) when choosing to follow her group's behavioral prescription by engaging in the symbolic behavior. Once this has taken place, play proceeds as before; individual group members make decisions about public goods contributions and symbolic behavior, and do so according to their *pre-existing* intensity of identification  $I_i^*(\varepsilon_i)$ . That this value  $I_i^*(\varepsilon_i)$  is still the relevant one reflects the framework's intuition that social identities adjust over the long run but are binding in terms of their influence on short-run decision making. Finally, in the long run, identities once again adjust according to the logic of a less-than-fully-conscious evolutionary process as was previously described. Once this process is complete, a given individual  $i$  who retains self-categorization as a group member achieves some new equilibrium value  $I_i^{**}(\varepsilon_i)$ , which in principle need not be the same as  $I_i^*(\varepsilon_i)$ . Alternatively, individual  $i$  may engage in a process by which she drops self-categorization as a group member altogether. This sequence of events, and its timing, are well-matched with the causal mechanism described in Section 2, through which

individuals who are compelled to violate group norms by failing to carry out the requisite symbolic behavior may undergo a process of identity adjustment in light of the violation they have committed.

What effect may the government's tax have on individual behavior and individual identification? Interestingly, there are a number of possibilities.

**Result 3.** Consider an individual  $i$  who self-categorized as a member of the group with  $I_i^* = k_i$  before the government imposed the tax. At the end of the long-run readjustment process following imposition of the tax, four outcomes are possible:

- (1)  $i$  may self-categorize as a member of the group, with  $I_i^{**}(\varepsilon_i) = c_i - r_i$ ;
- (2)  $i$  may self-categorize as a member of the group, with  $I_i^{**}(\varepsilon_i) = k_i + T$ ;
- (3)  $i$  may self-categorize as a member of the group, with  $I_i^{**}(\varepsilon_i) = 0$ ;
- (4)  $i$  may not self-categorize as a member of the group at all.

As such, the imposition of the tax may cause the individual to increase her intensity of internalization of group norms ( $I_i^{**}(\varepsilon_i) > I_i^*(\varepsilon_i)$ ), to decrease her intensity of internalization ( $I_i^*(\varepsilon_i) > I_i^{**}(\varepsilon_i)$ ), or to change her self-categorization so as to leave the group altogether.

Interestingly, a given individual who had  $I_i^*(\varepsilon_i) = k_i(\varepsilon_i)$  prior to the imposition of the tax may be successfully assimilated; may retain her self-categorization but with a reduced intensity of attachment; or may emerge with an *increased* intensity of attachment in a way that would be counterproductive given the government's intentions. The first thing to note is that such an individual would have engaged in the requisite symbolic behavior before the tax was imposed ( $I_i^*(\varepsilon_i) = k_i(\varepsilon_i)$ ), but does *not* do so after the tax is imposed (since  $I_i^*(\varepsilon_i) = k_i(\varepsilon_i) < k_i(\varepsilon_i) + T$ ). As a result, the individual suffers identity losses, which potentially can be assuaged through a re-evaluation of the individual's social identity. The new local maximum at  $I_i^{**}(\varepsilon_i) = k_i(\varepsilon_i) + T$  is relatively less attractive than the old local maximum at  $I_i^*(\varepsilon_i) = k_i(\varepsilon_i)$ , because an individual for whom  $I_i^{**}(\varepsilon_i) = k_i(\varepsilon_i) + T$  engages in the same behaviors and receives the same social benefits but at

higher costs. It may nonetheless be the case that  $I_i^{**}(\epsilon_i) = k_i(\epsilon_i) + T$  is a better outcome for an individual than the three other feasible alternatives; however, the relatively lower level of attractiveness of the new local maximum compared to the corresponding old one means that, depending on the values of the model's parameters, any of the outcomes can occur in equilibrium.

Of course, the aggregate effect of imposing a tax on an entire population depends on the distribution of individual characteristics within the population; however, the logic behind Result 3 indicates that a government's attempted campaign of assimilation may either have the desired effect or be counterproductive in the aggregate.

## 5. Discussion and Conclusions

This paper has presented a simple behavioral game-theoretic model of social identity symbols in the political sphere. In the model, the intensities with which individual group members internalize group norms is endogenous, and depend on the structural features of the social and economic context of the group. In an extension to the model, a government that attempts a "campaign of assimilation" by taxing use of a group's social identity symbol may under some circumstances achieve its objectives while causing a counterproductive backlash in which social identity attachments are actually *strengthened* under other circumstances.

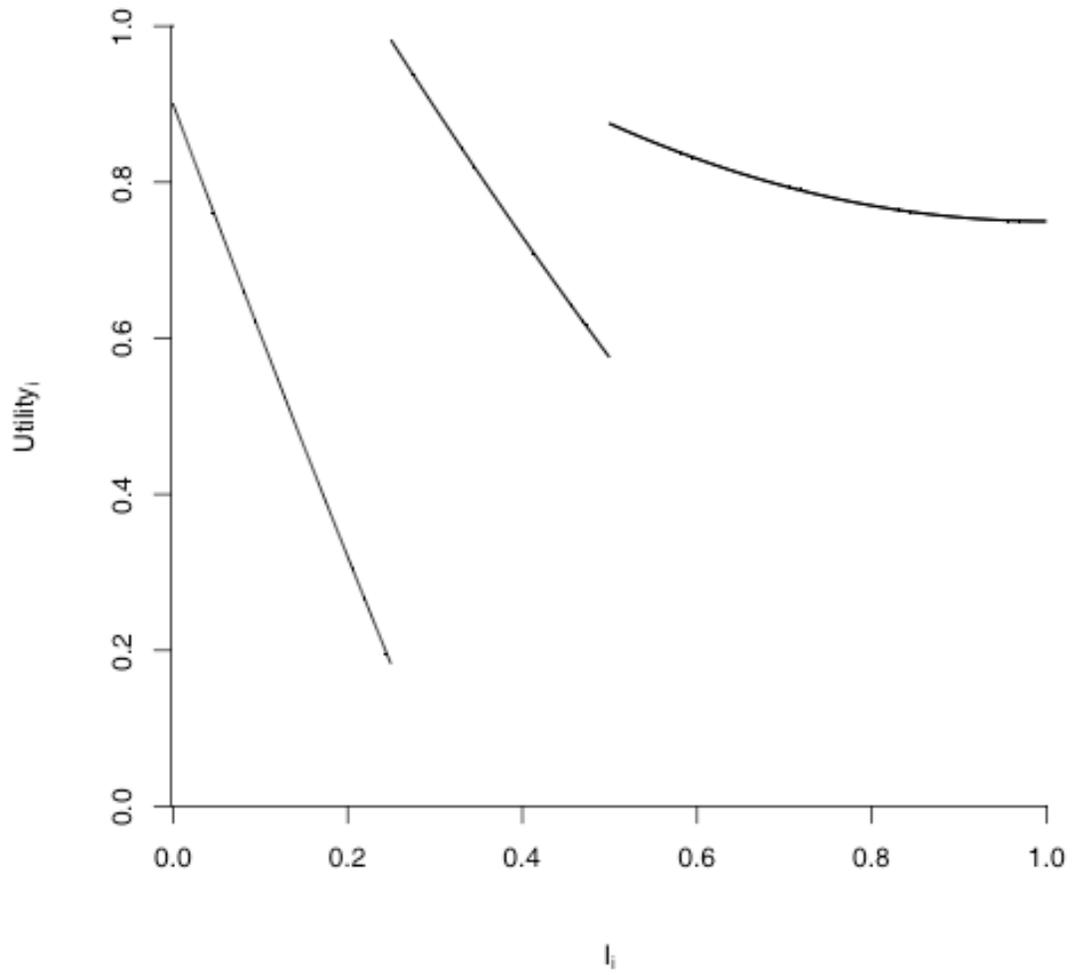
Of course, the models explicating these basic mechanisms are extremely simple in themselves, and also leave unmodeled a variety of factors which can affect individuals' incentives to use social identity symbols (or group leaders' incentives to advocate their use). Other scholars (e.g., Patel 2006) have explored, for example, the impact that social pressures may have on decisions to use, or not to use, identity symbols of varying degrees of intensity. In some settings, such pressures may be more or less orthogonal to the dynamics of intergroup political competition that are the focus here, but in other settings, such alternative accounts may be intimately bound up in these dynamics. Such interactions remain a subject for future research. Nonetheless, from a game-theoretic

perspective, this paper contributes to a small but growing literature on the endogenous formation of social identities (e.g., Shayo 2005, Penn 2008).

## References

- Akerlof, George and Rachel Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics* 115: 715-753.
- Bendor, Jonathan, Daniel Diermeier, and Michael Ting. 2003. "A Behavioral Model of Turnout." *American Political Science Review* 97(2): 261-280.
- Calvert, Randall L. 1992. "Leadership and Its Basis in Problems of Social Coordination." *International Political Science Review* 13(1): 7-24.
- Calvert, Randall L. 1995. "The Rational Choice of Social Institutions: Cooperation, Coordination, and Communication," in *Modern Political Economy*, ed. by J.S. Banks and E.A. Hanushek. Cambridge UK: Cambridge University Press.
- Chandra, Kanchan.
- Chwe, Michael Suk-Young. 2001. *Rational Ritual: Culture, Coordination, and Common Knowledge*. Princeton: Princeton University Press.
- Cook, Michael. 2003. *Forbidding Wrong in Islam*. Cambridge: Cambridge University Press.
- Dickson, Eric S. 2008. "Leadership, Followership, and Beliefs About the World: Theory and Experiment." New York University Working Paper.
- Dickson, Eric S. and Kenneth Scheve. 2006. "Social Identity, Political Speech, and Electoral Competition." *Journal of Theoretical Politics* 18(1): 5-39.
- Gladney, Dru C. 2004. *Dislocating China: Reflections on Muslims, Minorities, and Other Subaltern Subjects*. Chicago: University of Chicago Press.
- Kiernan, Ben. 2002. *The Pol Pot Regime: Race, Power, and Genocide in Cambodia Under the Khmer Rouge, 1975-79 (2<sup>nd</sup> ed.)*. New Haven: Yale University Press.
- Lupia, Arthur and Jesse Menning. 2006. "When Can Politicians Scare Citizens Into Supporting Bad Policies? A Theory of Incentives with Fear-Based Content." University of Michigan Working Paper.
- McGoldrick, Dominic. 2006. *Human Rights and Religion: The Islamic Headscarf Debate in Europe*. Oxford: Hart Publishing.
- Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge: Harvard University Press.

- Patel, David Siddhartha. 2006. "Concealing to Reveal: The Changing Informational Role of Islamic Dress." Stanford University mimeo.
- Penn, Elizabeth Maggie. 2008. "Citizenship versus Ethnicity: The Role of Institutions in Shaping Identity Choice." *Journal of Politics* 70(4).
- Riker, William and Peter Ordeshook. 1968. "A Theory of the Calculus of Voting." *American Political Science Review* 62(1): 25-42.
- Scott, Joan Wallach. 2007. *The Politics of the Veil*. Princeton: Princeton University Press.
- Sen, Amartya. 2006. *Identity and Violence: The Illusion of Destiny*. New York: W.W. Norton and Company.
- Shayo, Moses. 2005. "A Model of Social Identity with an Application to Redistribution." Working Paper.
- Tajfel, Henri. 1981. *Human Groups and Social Categories*. Cambridge, UK: Cambridge University Press.
- Tajfel, Henri, and John Turner. 1979. 'An Integrative Theory of Intergroup Conflict' in *The Social Psychology of Intergroup Relations*, William G. Austin and Stephen Worchel (eds). Monterey, CA: Wadsworth.
- Tajfel, Henri, and John Turner. 1986. 'The Social Identity Theory of Intergroup Behavior' in *Psychology of Intergroup Relations*, Stephen Worchel and William G. Austin (eds). Chicago: Nelson-Hall.
- Turner, John. 1984. 'Social Identification and Psychological Group Formation' in *The Social Dimension: European Developments in Social Psychology Vol. 2*, Henri Tajfel (ed). Cambridge, UK: Cambridge University Press.
- White, Richard. 1993. *"It's Your Misfortune and None of My Own": A New History of the American West*. University of Oklahoma Press.
- White, Jenny B. 2002. *Islamist Mobilization in Turkey*. Seattle: University of Washington Press.
- Wilson, Rick K. and Carl M. Rhodes. 1997. "Leadership and Credibility in N-Person Coordination Games." *Journal of Conflict Resolution* 41(6): 767-791.



**Figure 1.** Experienced utility as a function of  $I_i$ , per the discussion in the paper. For  $k_i(\varepsilon_i) = 0.25$ ,  $c_i(\varepsilon_i) = 0.6$ ,  $r_i(\varepsilon_i) = 0.1$ ,  $\gamma_{PG} = 0.15$ ,  $\gamma_S = 0.40$ ,  $Z = 1$ , and  $B_i(\varepsilon_i; N) = 2$ .