JOURNAL OF PUBLIC ECONOMIC THEORY
Department of Economics 415 Calhoun Hall Vanderbilt University Nashville, TN 37235 USA (615) 343-8495 (fax)

Submission Number: PED11-11-00017

Decline of Class: A Group-Theoretic Approach

Y. stephen Chiu
University of Hong Kong

Weifeng Zhong
Northwestern University

Abstract

We present a framework of group cooperation and competition in which agents are concerned not only about their material payoffs but also about their psychological payoffs, derived from working with others per se. We show a material foundation to such psychology --- the stronger a group's psychological preferences are, the greater the group's bargaining power will be in determining its terms of cooperation with other groups. We also generate implications that are consistent with two contemporary phenomena --- the decline of class in the politics of industrial economies and the salience of race in the third world.

Submitted: March 12, 2011.

Decline of Class: A Group-Theoretic Approach*

Y. Stephen Chiu† and Weifeng Zhong‡

February 4, 2011

Abstract

We present a framework of group cooperation and competition in which agents are concerned not only about their material payoffs but also about their psychological payoffs, derived from working with others per se. We show a material foundation to such psychology — the stronger a group's psychological preferences are, the greater the group's bargaining power will be in determining its terms of cooperation with other groups. We also generate implications that are consistent with two contemporary phenomena — the decline of class in the politics of industrial economies and the salience of race in the third world.

Keywords: class; race; intergroup relationship; group identity; group loyalty

JEL: D74, H00, O10

1 Introduction

This paper is motivated by two related phenomena. The first is the declining importance of class in contemporary politics in developed economies. It has been a concern that under universal suffrage, because the median voter has a lower income than the mean voter's, the poor will succeed in achieving a redistribution of income from the rich by choosing a 100% income tax. But this has not been the case and one proposed reason for this is that voters are not only concerned about economic issues, but are also concerned about non-economic issues

*We would like to thank Jimmy Chan, Bin Chen, Steve Ching, Eric Chou, Simon Fan, Vai-Lam Mui, Wing Suen and seminar participants at HKU and HKUST for their comments. All errors are our own. Financial support (HKU742709H) from the Research Grants Council of Hong Kong is gratefully acknowledged.

†Chiu (corresponding author): School of Economics and Finance, 918 K. K. Leung Building, University of Hong Kong, Pokfulam Road, Hong Kong. Phone: +852-2859-1056. Fax: +852-2548-1152. Email: schiu@econ.hku.hk.

‡Zhong: Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston IL 60208, USA. Email: w-zhong@kellogg.northwestern.edu.

like religion, race, abortion, anti-drug legislation, etc.[1] In a two-dimensional preferences model, Roemer (1998) shows that an equilibrium refinement will select an equilibrium in which the tax rate is substantially lower than unity, which is nonetheless predicted by the model in which voters are concerned only about income. What remains to be explained is how the multi-dimensionality of preferences is determined at the outset, and to what extent it is "manufactured" by the rich.

The second phenomenon is the salience of racial conflict over class conflict in third-world countries despite high income inequality.[2] The pork theory, as proposed by Fearon (1999) and Caselli and Coleman II (2010), starts with the observation that unlike other social dimensions race is the most easily recognizable and the least likely to change. Thus race-based coalitions provide the strongest warranty for agents to share the "pork" *ex post* and hence the strongest incentive for them to grip it *ex ante*. A recent theory by Esteban and Ray (2008) argues that because capital and labor are complementary in technology, social conflict takes place between a coalition of capitalists and workers of one race and those of another race.

In this paper, we provide an alternative theory to explain both phenomena. Unlike the Roemer theory, we assume neither specific voting institutions nor fixed multidimensional preferences. Unlike the pork theory, we do not assume that race imposes a more recognizable, less changeable social marker than class does; nor do we assume any complementarity between capital and labor as in the complementarity theory. What we do assume here are that capitalists constitute only a minority in the population and that agents derive utility from interacting with others *per se*. We argue that these simple properties could lead to interesting implications consistent with the two aforementioned phenomena.

Our model, which will be presented in more detail in Section 2, is as follows. Imagine a community populated by agents with different characteristics (both economic and non-economic). They are partitioned into exhaustive and mutually exclusive groups so that those in the same group share the same characteristics (including preferences). Entry into and exit out of a group is forbidden so that no asymmetry between class and non-class dimensions,

[1] Poole and Rosenthal (1991) show that roll call votes in the US congress are best explained by a two-dimensional model. Laver and Hunt (1992) present evidence that democratic politics are multi-dimensional in a sample of over twenty countries. Other reasons that prevent the poor from exploiting the rich include: (1) a tax rate close to 100% would induce so much disincentive so that the rich would stop creating value; (2) that upward mobility makes a low tax rate much more acceptable to the poor (see Putterman 1997 for a delineation of these different considerations). For country studies, see, e.g., Carmines and Stimson (1989) and Huckfeldt and Kohfeld (1989) on the US and Kitschelt (1994) on Europe.

[2] Although the definitions of race, nationality, and ethnicity are different, we use race uniformly when the distinction is unimportant.

or race and non-race dimensions, is presupposed. Agents interact among themselves to generate material production, obtaining material payoffs; they also derive (positive or negative) utility from the interactions *per se*, obtaining what we call psychological payoffs. Because agents in the same group are of the same race and religion, for instance, they enjoy greater psychological payoffs when working with each other than with any outgroup member.

We use the term *group identity*[3] to describe the shared psychological preferences of a group. It consists of a vector of coefficients characterizing the group's *intragroup* and *intergroup amicability* (with hostility interpreted as negative amicability). On the one hand, agents want to work with outgroup agents to increase their material payoffs. On the other hand, they are unwilling to do so unless compensated enough to make up the losses of psychological payoffs due to dilution of intragroup interactions. Hence, by assuming that groups bargain in a cooperative game (Shapley value bargaining), we are able to determine each group's welfare as a function of its own group identity, as well as the group identities of other groups. This also allows us to study the incentive for a group to shape its psychological preferences or to manufacture/manipulate the psychological preferences of another group.

The way we model the utility function follows Alesina and La Ferrara (2000), who study residents' decisions to contribute to a public good in a racially heterogenous community. In their model, each resident's utility depends on the proportion of residents of the same race as himself or herself in the whole population.[4] Using US survey data on attitude toward redistribution, Luttmer (2001) finds that, controlling for income, individuals increase their support for welfare spending as the share of local recipients from their own racial group rises. Luttmer refers to this as group loyalty. We think that assuming some kind of psychological payoffs when race is concerned is a good short-cut in the modeling.

Some comments about the psychological preferences are in order. First, because we want as little asymmetry as possible between race and any non-race dimension, we assume that psychological payoffs are involved whether or not the dimension is race. Admittedly, class seems to be less sensational than race is, but we think it is better to have it as an implication, rather than as a starting point, of the investigation. Second, no doubt having the same language or religion or culture, etc., facilitates cooperation and exchange by lowering transaction costs. Should this be the main underlying channel that affects

[3] It can alternatively be called group solidarity or group loyalty.

[4] Currarini, Jackson, and Pin (2009) provide a friendship network formation theory and test it using a survey data set among high school students in the US. In the theoretic formulation, the student's utility depends on how many friends he has, as well as the races of his friends.

a group's bargaining power, the psychological preferences referred to in this paper would then be a disguised notion that reflects efficiency gains or losses arising from decreasing or increasing transaction costs. Our study is then an investigation on how decreasing or increasing transaction costs affects group payoffs and the associated incentives to change the transaction costs.

In Section 3, we study how a group's material payoff is related to its psychological preferences. We find that a group's material payoff (i) is increasing in its intragroup amicability, as well as the intergroup amicability towards it; (ii) is decreasing in the intragroup amicability within and intergroup amicability among other groups; and (iii) somewhat surprisingly, may be increasing in its amicability toward some outgroup. The basic idea is that, by working with outgroup members, group members will be diluting their own interactions — as well as the interactions among outgroup members — and their bargaining power will be strengthened or weakened dependent on the various group identities. Results (i) and (ii) are fairly intuitive. Result (iii) suggests intriguing counter-intuitive spillovers between groups. As a whole, the analysis demonstrates a material foundation to psychological preferences.

Sections 4 and 5, the core of the paper, examine a specific model of the general framework laid out in Section 2. There are four groups of agents characterized by two dimensions: class and race. For concreteness and without implications, we call them white capitalists, white workers, black capitalists, and black workers. We examine the incentives of two groups with a common dimension to strengthen their mutual amicability by one unit; to make the exercise non-trivial, we require that the other two groups will automatically strengthen their mutual amicability by one unit as well. A potential alliance between the former two groups is said to exist if, among other conditions, they indeed benefit from such a strengthening of identity. We show that a potential alliance need not exist but, whenever it does, is always unique (except for knife-edge cases).

Two main results are found regarding the presence of potential alliance.

1. There is asymmetry because a potential alliance between (white and black) capitalists and a potential alliance between (white and black) workers. The former never exists under realistic parameters while the latter does.

2. In the case where the capitalists are predominantly of the same race, say, white, a potential alliance between white capitalists and white workers exists if the white population is below a critical proportion, and no potential alliance of any kind exists

otherwise.

The first result is consistent with the observation that the capitalists always preach the universality of values and it is the workers who advocate the importance of class (see Przeworski and Sprague, 1986). The parameter values that enable a potential alliance between white and black capitalists are such that the capitalists-over-workers ratio of a particular race is too high (say, greater than 2.5) to be realistic. Thus the property of capitalists, i.e., the elites in general, being a minority in the population plays a subtle, crucial role in leading to the first result.

The second result suggests exacerbated racial relationship when the (economically) dominant race is a minority group in the population (e.g., the whites in South Africa) but mitigated racial relationship when the dominant race is sufficiently large. Thus it sheds light on the argument that workers' group identity is manipulated and their class reconciliation position manufactured by the rich.[5] Our result is not only consistent with the argument. Because in our notion of potential alliance both groups benefit from its formation, our result also points out an overlooked possibility that those who were allegedly manipulated indeed benefit from the manipulation.

The second result also suggests an asymmetry of the pattern of conflict — racial conflict is more salient than class conflict once we assume that the capitalists are predominantly of the same race. By using some alternative, more relaxed definition of potential alliance, a potential alliance between white and black workers may become feasible. But the conditions under which it exists are still stringent and the general implication of the second result holds true.

We do not address the alliance formation as a non-cooperative game, prior to the cooperative game in which groups negotiate with each other on the division of the worth of the grand coalition. The reason is that equilibrium analysis of this type requires that the decisions of all four groups be checked, making the presentation too cumbersome and demanding on the reader (our preliminary equilibrium analysis, nonetheless, suggests similar results as under the potential alliance analysis).

Section 6 studies extreme identity in contrast with moderate identity studied in earlier sections. When a group's psychological preferences are too strong, the group may not want

[5] According to Roemer (1998), there is a view "in Left circles" that "the Right deliberately creates a certain non-economic issue — or tries to increase the salience of some such issues for voters — as a means of pulling working class voters away from Left parties, thereby driving economic policies to the right" (pp. 417).

to join a sub-coalition or even to form the grand coalition for material production, leading to efficiency loss when material payoffs are concerned. We show that a group with such an extreme identity may benefit from weakening its identity, or from reduced intergroup hostility between two outgroups. We also discuss some possible roles of extreme identity in economic development in this section before we conclude in Section 7.

1.1 Literature Review

The phenomenon that members of a group treat each other differently from the way they treat non-members have long been documented in the social science literature, evidenced in the terminology and theory such as ethnocentrism (Sumner, 1906), homophily (Lazarsfeld and Merton, 1954),[6] and the self-categorization theory.[7] Chen and Li (2009) present experimental evidence on ingroup altruistic behavior.[8] Bernheim (1994) models esteem, defined as the public perceptions of an individual's type, as part of individual preferences and uses it to explain an individual's conformity to social norms.

In the economic and psychology literature, some works have focused on belief manipulation. Mui (1999) models witch hunt as a game in which citizens respond to rumors about some fringe group of the society. Building on Romer (1995), Glaeser (2005) studies an election model in which, prior to the election, political entrepreneurs can spread rumors that the minority outgroup is harmful to the majority ingroup. This can be seen as the "supply side" of manipulation. Relatedly, Bénabou and Tirole (2006) study a model in which agents have insufficient will power, suggesting that people may be better off from, thus receptive to, the fostering of a belief that the world is just. This can be seen as the "demand side" of manipulation. Our approach is in line with this literature in that we adopt the concept of potential alliance (in terms of manipulating identities) of which potential allies have mutual incentives to form.

Akerlof and Kranton bring identity, defined as an individual's sense of self, into formal economic analysis (2000), applying it to education (2002) and organization (2005) in par-

[6] Ethnocentrism, coined by Sunmer a century ago, refers to the tendency to believe that one's own race or ethnic group is the most important and that some or all aspects of its culture are superior to those of other groups. Homophily refers to a tendency of various types of individuals to associate with others who are similar to themselves.

[7] Led by the seminal work of Tajfel et al. (1971), Tajfel and Turner (1979), and Turner et al. (1987), the categorization theory treats the formation of groups as a psychological process in which individuals categorize others as well as themselves. Once the process is completed, people become susceptible to biased information processing such that the information that enhances inter-category difference receives more attention.

[8] The evolutionary approach sheds light on group attribute and behavior. See Bowles and Gintis (2004) and Choi and Bowles (2007) on the survival of parochial altruism — preferences with both intragroup altruism and intergroup hostility — under evolutionary pressure.

ticular; Shayo (2009) uses the same approach to study redistribution. Bénabou and Tirole (2010) develop a theory of identity management and study how various psychological notions could be useful to individuals. The consequences of having such ingredients as identity in individual preferences are also experimentally studied (e.g., Eckel and Grossman, 2005 on intragroup activities and Charness et al, 2007 on intergroup interactions).

2 The Model

Consider a community consisting of a set of agents of measure N partitioned into n exhaustive and mutually exclusive groups; we use $1, \cdots, n$ to denote these groups. We use $s_1, \cdots, s_n$ to denote the membership size in each such group. There is a literature that studies how a group, especially a religious group, can be fortified.[9] An important question as it is, in this paper we examine different issues by simply assuming that, members in the same group have already overcome their collective action problem. As a result, we can adopt the following assumption.

Assumption 1 *Each group acts as a single decision maker.*

Agents produce output according to a characteristic function $v(\cdot)$: the value of the output produced by group i when it works alone is $v(i)$; the value of output produced jointly by group i and group j is $v(i \cup j)$, etc. The characteristic function satisfies the following standard property (see, for instance, Shapley 1953).

Assumption 2 *The characteristic function $v(\cdot)$ is strictly superadditive, i.e., $v(R_1 \cup R_2) > v(R_1) + v(R_2)$, where R_1 and R_2 are any two disjoint collections of groups.*

Strict superadditivity corresponds to the scenario where agents are strictly complementary in production and, as a result, formation of and cooperation in the grand coalition is socially optimal. This assumption is consistent with the diversity-in-production approach adopted by Alesina and La Ferrara (2005).[10]

[9] Adopting the club good approach, Iannaccone (1992, for the case of Christian sects), Berman (2000, Jewish sects), and Berman (2005, Islamic sects) provide an explanation to behavior such as prohibitions, self-sacrifices, bizarre behaviors, and even violent activities in some extreme religious sects — efficient sects that provide local public goods to their members overcome the free-ride problem by imposing stringent requirements as a costly signal of "commitment" to the community, thus allowing access to the club good.

[10] They point out that the diversity-in-production approach is consistent with the Dixit-Stiglitz (1977) production function, into which the efforts of the agents of different groups enter as differentiated inputs.

In addition to material products, agents are also concerned about who they work with; more specifically, agents derive psychological utility from working together *per se*. An interesting special case is homophily, i.e., agents prefer working together with other agents in the same group. In the context of ethnicity, for instance, an individual may feel more comfortable dealing with a member of her ethnic group than with a non-member; in the context of religion, an individual may feel more comfortable dealing with someone who shares the same religion than someone who does not.[11]

More generally, we represent group i's psychological preferences by a vector $\mathbf{a}_i \equiv (a_{i1}, \cdots, a_{in})$ of coefficients characterizing the group's intragroup and intergroup amicability. Put differently, a_{ij} measures how much group i members want to work with group j members. By forming a coalition R with some other groups, group i will obtain psychological payoffs

$$\alpha_i(R) \equiv s_i \times \left(\frac{s_i}{s_R} \times a_{ii} + \sum_{j \subset R \setminus i} \frac{s_j}{s_R} \times a_{ij} \right), \tag{1}$$

where $s_R = \sum_{j \subset R} s_j$.

The right-hand side (RHS) of (1) has a natural interpretation. Within a certain period of time, members in the coalition engage in pairwise matching so that each member spends an equal amount of time with every other member. For a member in group i, in particular, she will spend s_i/s_R of her time with group i's members, and s_j/s_R of her time with group j's members, where $j \subset R \setminus i$. This accounts for the term in the parentheses in the RHS of (1). Given s_i members in group i, the RHS represents the total psychological payoffs that group i members will collectively obtain when coalition R is formed. If $R = i$, i.e., if group i chooses to work alone, its psychological payoffs will simply be $\alpha_i(i) = s_i a_{ii}$.

Our approach of putting psychological payoffs into utility can be viewed as an extension of that of Alesina and La Ferrara (2000), who assume that an agent's utility from being in a community depends on the ratio of members to non-members of the agent's group in the community. Furthermore, in a study about the formation of friendship, Currarini, Jackson, and Pin (2009) construct a utility function where an agent's utility depends on the number of ingroup friends as well as the number of outgroup friends the agent has. Roemer puts

[11] Although it is not straightforward to see this type of psychological utility in the context of class in contemporary industrial nations, one can more easily locate it in pre-World War writings (see, for instance, Veblen (1899) in connection with the upper class; see the chapter on Karl Polanyi in Drucker (1998), in connection to the working class). Hobsbawn (1994, pp. 395-408) describes how the identity of the working class in western countries was enhanced through the collective life experience in the interbellum period and how it was weakened through the individualist life style in the post-war period.

non-economic factors directly behind voters' preferences.

We call the vector $\mathbf{a}_i$ group i's *identity* and the n components of the vector as the group's *identity coefficients or amicability coefficients.* Given $\mathbf{a}'_i \equiv (a'_{i1}, \cdots, a'_{in})$ and $\mathbf{a}''_i \equiv (a''_{i1}, \cdots, a''_{in})$ if $a'_{ii} \geq a''_{ii}$ and $a'_{ij} \leq a''_{ij}$ for any $j \neq i$, we say that group i has a stronger identity under $\mathbf{a}'_i$ than under $\mathbf{a}''_i$, or $\mathbf{a}'_i$ is an *(identity) strengthening* of $\mathbf{a}''_i$. (Note that $\mathbf{a}_i = (a_{i1}, \cdots, a_{ii}, \cdots, a_{in})$ with $a_{ii} > a_{ij}, \forall j \neq i$ corresponds to the aforementioned homophily phenomenon.) Each group's identity coefficients are exogenously given, and our main exercise is to perform comparative statics of group welfare with respect to these coefficients. (In later sections, we will study how groups would want to change their identity coefficients if they were able to do so.) Note that, despite similarities, the notion of group identity we adopt is different from the notion of identity that is well known in the literature. The latter is formally introduced to economics in the seminal paper by Akerlof and Kranton (2000) and in there, defined from an agent's perspective, identity means the agent's sense of self.[12]

Next, we assume that a group's total payoffs equal the sum of its material payoffs and psychological payoffs. Thus, given a coalition R, the total utility of its two member groups is given by

$$u(R) \equiv v(R) + \sum_{j \subset R} \alpha_j(R).$$

We call $u(\cdot)$ the total characteristic function.

Assumption 3 *The total characteristic function $u(\cdot)$ is strictly superadditive, i.e., $u(R_1 \cup R_2) > u(R_1) + u(R_2)$, where R_1 and R_2 are any two disjoint unions of groups.*

Assumption 3 states that, because the absolute values of a_{lm} are small enough for all l and m, even when psychological payoffs are also taken into account, the formation of the grand coalition is still efficient. While used in a large portion of the paper, we will relax the assumption in Section 6. We call the two scenarios the moderate identity and extreme identity scenarios.

Given the total utility function, we assume that each group obtains its own Shapley value (Shapley, 1953) taking each group as an individual player. This can be understood as an intergroup bargaining exercise in which the worth of the grand coalition is divided among all

[12] There is a literature explaining the formation of identity. For instance, Bisin and Verdier (2000, 2001) and Bisin et al. (2010) model identity as the result of cultural transmission and socialization, while Darity et al. (2006) interpret identity using an evolutionary game.

member groups. Assumption 1 justifies the treatment that each group enters the bargaining as a single decision maker. Assumption 3 implies that forming the grand coalition is indeed efficient.

More specifically, group i will obtain a total payoff of

$$\phi_i(N) \equiv \sum_{T \not\ni i} \frac{|T|!\,(n-|T|-1)!}{n!}\left[u(T\cup i)-u(T)\right], \tag{2}$$

where $|T|$ is the number of groups in coalition T. There is a natural interpretation here. Imagine that groups arrive at the scene in a random order, then (2) is just the weighted average of group i's marginal contribution to each conceivable coalition that it joins.[13]

We use $\beta_i(N)$ to denote group i's material payoffs, where

$$\beta_i(N) \equiv \phi_i(N) - \alpha_i(N).$$

When our discussion is restricted to a smaller union of groups, $R \subset N$, the total payoffs and material payoffs that group i obtains are denoted by $\phi_i(R)$ and $\beta_i(R)$, respectively, and are calculated in a similar manner. To economize the notation, we define $\beta_i \equiv \beta_i(N)$, $\phi_i \equiv \phi_i(N)$, and $\alpha_i \equiv \alpha_i(N)$.

Here we are most interested in the effect of identity coefficients on the material and total payoffs of groups. To this end, we will study how an infinitesimal change in the former influences the latter. One can augment this approach by taking into account the costs incurred to change identity coefficients. As will become clear, each group's payoff — through Shapley value bargaining — turns out to be a linear function of identity coefficients. If we consider a sufficiently convex cost function, there will then be a small extent to which identity is optimally changed. However, we abstract from this embedded optimization problem by looking only at the marginal benefits of identity change, but not at the marginal costs.

3 A Material Foundation of Identity

In this section, we study group i's welfare as a function of identity coefficients, demonstrating that a stronger identity usually implies a higher material payoff. This thus provides

[13] For noncooperative foundations that provide a natural interpretation of this payoff, see, e.g., Gul (1989), Hart and Mas-Colell (1996), and Dasgupta and Chiu (1998). For related formulation, see Aumann and Myerson (1986) and Maskin (2003). See Segal (2003) for the subtle issue arising when the effective number of bargainers is changed because of various forms of integration.

a material foundation to the psychology we study. (Because the total payoffs can be increased unboundedly by strengthening a group's identity, we do not focus on total payoffs in subsequent analysis.) We leave the determination of identity to subsequent sections. The following proposition is easy to obtain (unless otherwise stated, all proofs are relegated to Appendix B).

Proposition 1 *Under Assumptions 1-3, we have*

[1] *For all* j, k *(not necessarily distinct),* $\sum_{i=1}^{n} \partial\beta_i/\partial a_{jk} = 0$.

[2] *For all* i, j, k *(not necessarily distinct),*

  (*i*) *if the group size of any one of them is zero, then* $\partial\beta_i/\partial a_{jk} = 0$*;*

  (*ii*) *adding (removing) any group* $\ell \neq i, j, k$ *with* $s_\ell = 0$ *into (from) the economy does not change the value of* $\partial\beta_i/\partial a_{jk}$.

[3] $\partial\beta_i/\partial a_{ii} > 0$ *for* $s_i \in (0, N)$*;*

[4] $\partial\beta_i/\partial a_{jj} < 0$ *for any distinct* i, j *and* $s_i > 0$ *and* $s_j > 0$*;*

[5] $\partial\beta_i/\partial a_{ji} > 0$ *for any distinct* i, j *and* $s_i > 0$ *and* $s_j > 0$*;*

[6] $\partial\beta_i/\partial a_{jk} < 0$ *for any distinct* i, j, k *and* $s_i > 0, s_j > 0,$ *and* $s_k > 0$*;*

[7] *for* $j \neq i$ *and* $s_i > 0$ *and* $s_j > 0$,

  (*i*) $\partial\beta_i/\partial a_{ij} > 0$ (< 0) *if* $(s_i + s_j)/N < \frac{2}{n(n-1)+2}$ $\left(> \frac{1}{2}\frac{n-2}{n-1}\right)$*; and*

  (*ii*) *for* $n = 3$, $\partial\beta_i/\partial a_{ij} > 0$ *if and only if* $(s_i + s_j)/N < 1/4$.

Result 1 is rather straightforward. Given that the total material payoff to be divided among groups is equal to $v(N)$ and is constant any change in an identity coefficient leads to gains and losses for individual groups, but the sum of the changes is zero. Result 2 states that any group with zero mass plays no role at all in the psychological game. Any change in its own identity coefficient will not affect its own material payoff, nor that of any other group; any change in an outgroup's intergroup amicability towards this zero-mass group will have no effect on any group's material payoff either; adding or removing any zero-mass group does not change the pre-existing psychology at all. The underlying reason is that, under random arrival, the addition of any zero-mass group to any pre-existing coalition of groups does not dilute the psychological payoffs of each member group in the coalition.

The remaining five results are pertinent to non-zero mass groups. Results 3 to 6 are intuitive. For an individual group, self love is good in terms of material payoff (result 3); the self-love of outgroups is bad (result 4); being loved by outgroups is good (result 5); and the mutual love between two outgroups is bad (result 6). The key is the substitutability between material utility and psychological utility. Take result 3 as an example. The stronger the amicability among group i members, the greater the psychological satisfaction they will lose when cooperating with outgroup members. To persuade them to stay in the grand coalition, therefore, a larger amount of material payoffs have to be allocated to them.

Result 7 is more subtle. Presupposition has it that loving others always hurts as long as material payoff is concerned. The result says that it is true only if the combined size of the ingroup and the outgroup being loved is sufficiently large. To understand the intuition, suppose group i's amicability toward group j is now stronger. On the one hand, having a greater psychological utility by cooperating with j reduces i's material payoff because of substitution between material and psychological payoffs. On the other hand, the participation of groups other than i and j in the grand coalition dilutes the increased psychological utility group i would otherwise obtain through working solely with group j — this increases the compensation by other groups to group i for the latter to stay in the grand coalition. The change in i's material payoffs thus depends on the net effect of these two effects. Note that the larger j is, the larger the first effect becomes, thus the more i has to compensate j. Besides, the larger i and j together is, the smaller the size of all other groups, thus the smaller the dilution effect, and hence the less all other groups have to compensate i. Therefore, the net effect on i's material interests is positive when the combined size of i and j is not too large.

3.1 Identity Strengthening

The material foundation of identity discussed suggests that an individual group does have the incentive to modify and in most cases strengthen its identity. Consider the following two ways of identity strengthening (through media, education, etc.):

- Group i is said to engage in *outward identity strengthening* (*the outward strategy* for short) if it chooses to decrease its amicability (or increase its hostility) toward one particular out-group j, i.e., to decrease a_{ij}, for one $j \neq i$.
- Group i is said to engage in *inward identity strengthening* (*the inward strategy*) if it

chooses to increase its intragroup amicability, i.e., to increase a_{ii}.

The material benefit of the inward strategy is $\partial\beta_i/\partial a_{ii}$ while the material benefit of the outward strategy targeting group j is $-\partial\beta_i/\partial a_{ij}$. Hence, as far as material payoff is concerned, the outward strategy is more profitable if and only if

$$\frac{\partial\beta_i}{\partial a_{ii}} < \max_j \left(-\frac{\partial\beta_i}{\partial a_{ij}}\right), \tag{3}$$

It is straightforward to obtain the following lemma.

Lemma 1 *Under Assumptions 1-3, we have*

$$\frac{\partial\beta_i}{\partial a_{ii}} = \sum_{j\neq i} \left(-\frac{\partial\beta_i}{\partial a_{ij}}\right). \tag{4}$$

Lemma 1 means that, in terms of group i's material payoffs, increasing a_{ii} by one unit while holding a_{ij} constant for all $j \neq i$ is the same as decreasing a_{ij} by one unit for all $j \neq i$ while holding a_{ii} constant. Comparing (3) with (4), one can easily verify that (3) holds only if there exists some $k \neq i$ such that $\partial\beta_i/\partial a_{ik}$ is positive. Put differently, a necessary condition for the outward strategy to be the optimal strategy for group i is that there exists some group k, $k \neq i$, such that group i can benefit from increasing amicability towards it. The conditions for this latter scenario were studied in Proposition 1.7.

In particular, imagine the case where there are only three groups, 1, 2, and 3. Then group 1 finds the outward strategy (targeting toward group 3) better than the inward strategy if and only if $\partial\beta_1/\partial a_{12} > 0$, which occurs when $(s_1 + s_2)/N < 1/4$. Holding s_2 constant, this means that s_1 must be sufficiently small. Therefore, despite a common presumption that a small group is usually a victim of an intergroup conflict, it need not be the case. In our framework, a small group has strong incentives to engage in outward-looking identity strengthening, such as stereotyping or airing grievances against a specific, large outgroup, and a large group has strong incentives to engage in inward-looking identity strengthening, such as glorifying own group's history, etc.. We leave these questions for future studies.[14]

[14] It suggests that when a large group seems to be "bullying" a small group, it may be in fact only be a small segment of the former that is doing so. Galeser (2005) models a small segment of the agents in the majority who spread hatred against a minority and calls these agents political entrepreneurs.

4 Class and Race: Preliminaries

In this section and the next, we investigate which dimension among multiple dimensions is the most prominent one. We first present some auxiliary results.

Lemma 2 *Consider some* $s_1 > 0$,

[1] *if* $n \geq 2$ *and* $s_n = N - s_1 > 0$, *then*

$$\frac{\partial \beta_1}{\partial a_{1,1}} + \frac{\partial \beta_1}{\partial a_{n,n}} = 0;$$

[2] *if* $n \geq 3$ *and* $s_{n-1} + s_n = N - s_1 > 0$, *then*

$$\frac{\partial \beta_1}{\partial a_{1,1}} + \sum_{i,j \in \{n-1,n\}} \frac{\partial \beta_1}{\partial a_{i,j}} \begin{cases} = 0 & \text{if} \quad \text{either } s_{n-1} = 0 \text{ or } s_n = 0, \\ < 0 & \text{if} \quad s_{n-1} > 0 \text{ and } s_n > 0. \end{cases}$$

The first result states that, when the population is concentrated in two non-zero-mass groups, a simultaneous increase in the two groups' intragroup amicability leaves each group's material payoff unchanged. Whereas a unit increase in $a_{1,1}$ increases β_1 by a certain amount, a unit increase in $a_{n,n}$ decreases β_1 by exactly the same amount. The second result studies a similar identity strengthening exercise in which, along with a one-unit increase in $a_{1,1}$, there are also one unit increases in $a_{n-1,n-1}, a_{n-1,n}, a_{n,n-1}$, and $a_{n,n}$. When either s_{n-1} or s_n equals zero, this exercise becomes exactly the same as the one studied in the first result (using Proposition 1.2), and β_1 remains unchanged. However, when s_{n-1} and s_n are both non-zero, β_1 is reduced. Thus group 1 would find the two opponent groups ($n-1$ and n) more intimidating when they are more even in group size.

To see the intuition, consider the simplest case in which there are three groups: 1, 2, and 3. Group 1's material payoff β_1, by definition, is equal to its total payoff ϕ_1 minus its psychological payoff α_1. In the identity strengthening exercise, the distribution between s_2 and s_3 affects β_1 through ϕ_1, but not α_1.[15] Recall that ϕ_1 is a weighted average of its marginal contribution to the worth of coalitions. Groups arrive in sequence in one of six ways. For the two sequences where 1 arrives the earliest and the two sequences where 1 arrives the latest, the distribution between s_2 and s_3 has no effect on 1's marginal contribution. It

[15] Note that $\alpha_1 \equiv s_1 \left(\sum_{i=1,2,3} a_{1i} s_i / N \right)$. The identity strengthening leads to a change in α_1 by s_1^2/N, independent of s_2 and s_3.

has effect, nonetheless, in the other two sequences, where 1 arrives in between the other two groups. When 1 arrives after $i \in \{2, 3\}$ but before $j = \{2, 3\} \backslash i$, 1's marginal contribution is

$$MC(i) = v(1, i) + s_1 \left(\frac{a_{1,1}s_1}{s_1 + s_i} + \frac{a_{1,i}s_i}{s_1 + s_i} \right) + s_i \left(\frac{a_{i,1}s_1}{s_1 + s_i} + \frac{a_{i,i}s_i}{s_1 + s_i} \right) - (v(i) + s_i a_{i,i}).$$

Upon the identity strengthening (a simultaneous increase in $a_{1,1}, a_{i.i}, a_{i,j}, a_{j,i}$, and $a_{j,j}$ by one), the change in $MC(i)$ becomes (with some manipulation) $s_1 - 2s_1 \frac{s_i}{s_1+s_i}$.

Hence, the presence of i imposes a detrimental effect on the increase in $MC(i)$, and this effect is increasing in s_i but at a decreasing rate. Taking into account both $i = 2$ and $i = 3$, as well as the weights assigned, the change in 1's total payoff is thus $\frac{1}{3}s_1 - \frac{1}{3}s_1 \left(\frac{s_2}{s_1+s_2} + \frac{s_3}{s_1+s_3} \right)$, which is maximized when either $s_2 = 0$ or $s_3 = 0$ and is minimized when $s_2 = s_3$. Put differently, the detrimental effect on the change in ϕ_1 induced by s_2 and s_3 is smallest when either $s_2 = 0$ or $s_3 = 0$ but is largest when $s_2 = s_3$. This subtle insight will prove to be very useful for our subsequent analysis.

4.1 The Model

Consider a model with two dimensions: class and a non-class dimension which we call race. There are four groups: the white capitalist group WK, the white worker group WL, the black capitalist group BK, and the black worker group BL; without loss of generalization, we normalize the total population to unity and denote the grand coalition by T. In our analysis, we will assume that s_{BK} is small.

Consider a thought experiment in which, because of increased intra-racial amicability, there is a simultaneous increase in $a_{WK,WK}$, $a_{WK,WL}$, $a_{WL,WK}$, and $a_{WL,WL}$ (among the whites) by one unit, as well as an increase in $a_{BK,BK}$, $a_{BK,BL}$, $a_{BL,BK}$, and $a_{BL,BL}$ (among the blacks) by one unit. In this case, WK's material gain from this exercise is given by

$$B_{WK}^{I}(R) \equiv \sum_{i,j \in \{WK,WL\}} \frac{\partial \beta_{WK}}{\partial a_{ij}} + \sum_{i,j \in \{BK,BL\}} \frac{\partial \beta_{WK}}{\partial a_{ij}}.$$

One interpretation of the exercise is as follows. The increased intraracial amicability among the whites is due to the spreading of a propaganda by the white elites that enhances the "mutual love of white people". However, the propaganda is not omnipotent — the blacks now feel obliged to enhance the mutual love of black people to the same extent (reciprocity is studied in behavioral economics, e.g., Bolton and Ockenfels, 2000). White elites benefit from the manipulation if $B_{WK}^{I}(R) > 0$.

More generally, consider an increase in *intra-d* amicability for all groups by one unit, where $d = C$ (class), R (race). The material gain to group t is given by

$$B_t^I(d) \equiv \sum_{i,j \in \{t, f_t(d)\}} \frac{\partial \beta_t}{\partial a_{i,j}} + \sum_{i,j \in T \setminus \{t, f_t(d)\}} \frac{\partial \beta_t}{\partial a_{i,j}}, \tag{5}$$

where $f_t(d)$ denotes the group that t is in common with in terms of d. As a result, groups t and $f_t(d)$ feel closer to each other; and so do the other two groups. Note that the relationship between t and $f_t(d)$ is strengthened through making their common identity more salient. This is a kind of inward-looking behavioral strategy and this explains the superscript I on the left-hand side (LHS).

Likewise, we can consider a thought experiment with reduced *inter-d* amicability for all groups by one unit, where $d = C$, R. The material gain to group t is given by

$$B_t^O(d) \equiv -\left(\sum_{\substack{i \in \{t, f_t(d)\} \\ j \in T \setminus \{t, f_t(d)\}}} \frac{\partial \beta_t}{\partial a_{ij}} + \sum_{\substack{i \in T \setminus \{t, f_t(d)\} \\ j \in \{t, f_t(d)\}}} \frac{\partial \beta_t}{\partial a_{ij}} \right). \tag{6}$$

Note that in this case the relationship between two groups is strengthened through making their contrast with the other two groups more saliently. This is a kind of outward-looking behavioral change and this explains the superscript O on the LHS.

Our first result is that, as long as WK's material welfare is concerned, the inward- and outward-looking strategies are simply equivalent.

Lemma 3 *Suppose Assumptions 1-3 hold and there are four groups, WK, WL, BK, and BL. For all (t, d), $B_t^I(d) = B_t^O(d)$.*

Because of this, we will ignore the superscript I and O in in the subsequent exposition and proofs.

4.2 The Benefits of Manipulation or of being Manipulated

We obtain the following proposition.

Proposition 2 *Suppose Assumptions 1-3 hold and there are four groups, WK, WL, BK, and BL.*

[1] *For all* (t, d)*, provided that* $s_{f_t(d)} = 0$ *and* $d' \neq d$*, we have*

$$B_t(d) \begin{cases} = 0 & if \quad s_{f_t(d')} \in \{0, 1 - s_t\}, \\ < 0 & if \quad s_{f_t(d')} \in (0, 1 - s_t). \end{cases}$$

[2] *For all* (t, d)*, provided that* $s_{t'} = 0$ *for some* $t' \notin \{t, f_t(d)\}$*, we have*

$$B_t(d) \begin{cases} > 0 & if \quad s_{f_t(d)} \in (0, (1 + s_t)/2), \\ = 0 & if \quad s_{f_t(d)} = \{0, (1 + s_t)/2, 1 - s_t\}, \\ < 0 & if \quad s_{f_t(d)} \in ((1 + s_t)/2, 1 - s_t). \end{cases}$$

For each result in the proposition, one group has zero mass ($f_t(d)$ in the first and t' in the second). In the following discussion, we focus on the most interesting case — the one in which BK is that zero-mass group. The first result, itself a result of Proposition 1.2 and Lemma 2.2, says that $B_{WK}(C)$ is never positive and in fact must be strictly negative unless $s_{WL} = 0$ or $1 - s_{WK}$ (neither case is interesting though). Hence, it almost always hurts for white capitalists if class is made more salient, given that their potential ally, black capitalists, is of a negligible size. Likewise, when $t = BL$ and $d = C$, the result says that it almost always hurts the black workers if race is made more salient, given that their potential ally, black capitalists, are of a negligible size.

To understand result 2, consider $t = WK$, $d = R$, $f_t(d) = WL$, and $t' = BK$ with $s_{BK} = 0$; hence, black capitalists are of a negligible ratio in the population. The result says white capitalists benefit from increased intra-racial amicability (among the whites and also among the blacks) if and only if the population of white workers, while non-negligible, is small enough.

This result provides useful insights into some real-world episodes. For example, in South Africa, since the ruling class (the white capitalists) is of an ethnic minority, the fraction of white workers in the population is necessarily small. From the white capitalists' point of view, therefore, exacerbating the racial difference between the whites and the blacks may well be "profitable", explaining why ethnic segregation was legalized in the post-WWII era. Elsewhere, however, the majority of the working class is typically of the same ethnicity as the ruling elite, suggesting that mitigating racial difference is in the ruling elites' interest. The case of Singapore is worth a mention. The use of Chinese language is actually suppressed in the island city-state despite the fact that the language is the mother tongue of the majority of the population. According to our current model, such language policy may well be consistent with the interest of the ethnic Chinese elites, *even if racial harmony itself did not create additional aggregate material payoff for the whole nation.*[16]

Proposition 2.2 can also be applied to white workers. Now we have $t = WL$, $d = C$, $f_t(d) = WK$, while $t' = BK$ and $s_{BK} = 0$ as before. WL also benefits from the strengthening racial identity, as long as $s_{WK} \in (0, (1 + s_{WL})/2)$, which is satisfied as long as $s_{WK} < s_{WL}$. This result is not only consistent with the story allegedly popular in the Left Circle as pointed out by Roemer (1998) (see footnote 5). It also suggests an overlooked possibility that those who are allegedly manipulated may indeed benefit from the manipulation.

5 Class and Race: Competing for Allies

5.1 Potential Alliance

Suppose black workers could also use propaganda to increase the amicability among all workers by one unit and as a consequence white capitalists would also increase the amicability among capitalists by one unit. Between the two urges — one by white capitalists based on race and another by black workers based on class — which one will white workers respond to? The result is given in the following lemma.

Lemma 4 *Suppose Assumptions 1-3 hold and there are four groups, WK, WL, BK, and BL. For all (t,d), suppose $s_\tau = 0$ where $\tau = f_{f_t(d)}(d')$ and $d' \neq d$. Then, $B_t(d) > B_t(d')$ if and only if $s_{f_t(d)} < s_{f_t(d')}$.*

Thus, suppose $t = WL$ and $\tau = BK$ with $s_{BK} = 0$, and if a choice has to be made, WL will respond to calling of the smaller group between WK and BL. In case WK is smaller, provided that both $B_{WK}(R) > 0$ and $B_{WL}(R) > 0$, it is indeed in the interest of both WK and WL to strengthen their racial identity regardless of the reaction of the black population. To further understand the incentives of each of the four groups to be closer to another group, we make use of the following definition.

[16] In South Africa, 20.8% of the population were white and 67.6% were black in 1951 (first apartheid census); in Singapore, 77% of the population were ethnic Chinese, 14.8% were Malays, and 7% were Indians in 1970 (shortly after gaining independence); the latest estimate in 2009 of the US population is that 65.4% are white and 12.6% are black.

Definition 1 *Two distinct groups i and j sharing a common dimension d are said to have mutual incentives to form an alliance if $B_i(d) \geq 0$, $B_j(d) \geq 0$, $B_i(d) \geq B_i(d')$, and $B_j(d) \geq B_j(d')$, where $d' \neq d$, and at least one of the first two inequalities must be strict. In this case, we say that a* potential alliance (PA hereafter) *(between i and j) exists.*

First, compared with the status quo, the two groups in a PA are not worse off and indeed at least one is strictly better off through increasing their intra-d amicability despite increased intra-d amicability within the other two groups. Second, neither group will strictly benefit from switching to increasing amicability along the other line (d') given that, should this occur, the other two groups along the same dimension will increase their amicability by the same degree. Third, we call it a *potential* alliance, rather than an *equilibrium* alliance, since we have not specified an alliance-formation game. Another way to study the incentive to form an alliance is to formally study a non-cooperative game in which groups propose and respond to each other. However, such an exercise is too complicated for our purpose here (one would have to examine the strategies of four players, to deal with multiple equilibria, etc.).

The following result simplifies the study of potential alliances.

Lemma 5 *Suppose Assumptions 1-3 hold and there are four groups, WK, WL, BW, and BL.*

[1] *There does not exist multiple PAs along the same dimension.*

[2] *Suppose $s_\tau = 0$ for some τ, then (i) no PA exists that involve τ, and (ii) (letting t be the group with no common dimension with τ) a PA between t and $f_t(d)$ and another between t and $f_t(d')$ co-exist if and only if $s_{f_t(d)} = s_{f_t(d')}$.*

Result 1 means that, whenever there is a PA between two groups, there will not be one between the other two groups. This is simply a result of Proposition 1.1, which states that in any identity strengthening/weakening exercise, the net effect on the material payoffs of all groups must be zero. Result 2 suggests that a PA can only exist between WK and WL or between BL and WL, but not both at the same time. Furthermore, except for some knife-edge cases (with $s_{WK} = s_{BL}$), uniqueness is ensured.

We can represent each set of feasible group sizes as a point in a unit simplex (see Figure 1). Lines aa, bb, cc, dd, and ee are depicted to partition the simplex into different regions.

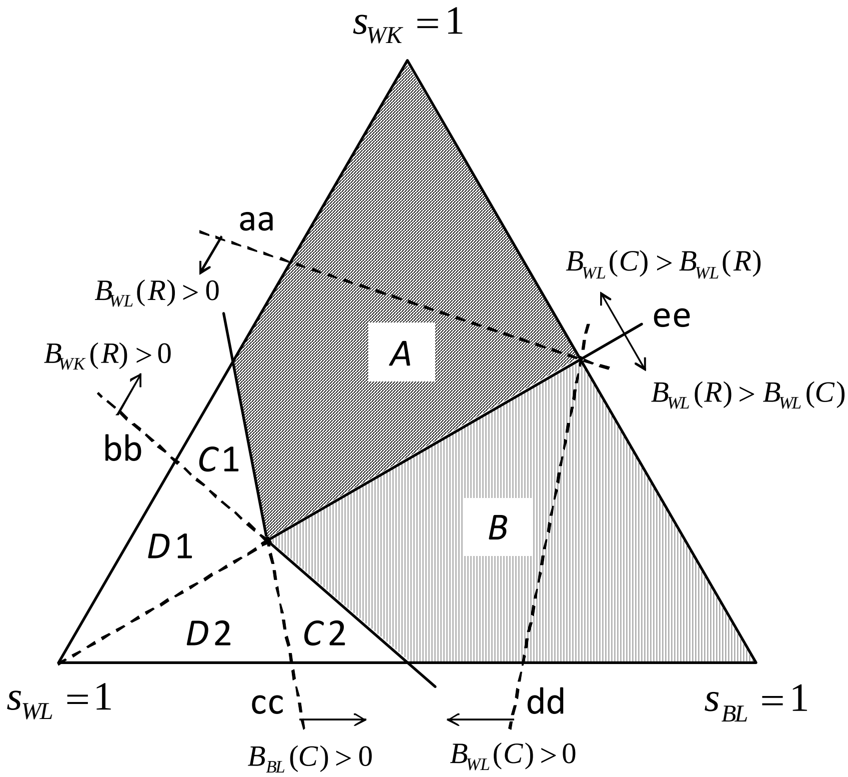

Figure 1: Region A is where a class-based potential alliance between WL and BL exists; region B contains a race-based potential alliance between WK and WL; in regions $C1$, $C2$, $D1$, or $D2$, no potential alliance exists.

Line aa refers to the condition that $B_{WL}(R) = 0$, i.e., $s_{WK} = (1 + s_{WL})/2$; other lines are interpreted similarly. According to Lemma 5.2.ii, more than one PA appears only on line ee. If we ignore this knife-edge case, we obtain the following result.

Proposition 3 *Suppose $s_{BK} = 0$.*

[1] *In region A, there exists only one PA, that between BL and WL;*

[2] *In region B, there exists only one PA, that between WK and WL;*

[3] *In region C1, C2, D1, or D2, no PA exists.*

Note that region A is not a region of interest to us because in that region the population of white capitalists exceeds 20% of the total population and is too large to be realistic.[17] If we restrict our attention to where the population of WK is small ($s_{WK} \leq 20\%$), we only need to consider regions $D1$, $D2$, $C2$, and B. In region B, the race-based PA between WK and WL is present; in $D1$, $D2$, or $C2$, no PA exists. Thus using the notion of PA, a race-based PA is possible, while any class-based PA is not.

Note that in regions $D1$ or $D2$, no PA exists because neither WK nor BL has any incentive to form an alliance, i.e., $B_i(d) < 0$, where $i \in \{WK, BL\}$ and $d \in \{C, R\}$ and thus WL cannot form an alliance because it has no willing partner. Region $C2$ does not contain any PA either, but the reason is different. Because $B_{WK}(R) < 0$, a PA between WK and WL does not exist; because both $B_{BK}(C) > 0$ and $B_{WL}(C) > 0$ but $B_{WL}(R) > B_{WL}(C)$, a PA between BK and WL does not exist either. Definition 1 can be weakened as follows.

Definition 1a *A potential alliance between two distinct groups i and j sharing a common dimension d exists if $B_i(d) \geq 0$, $B_j(d) \geq 0$ (with at least one being strict), and for each $t \in \{i, j\}$, either* (i) *$B_t(d) \geq B_t(d')$ or* (ii) *both $B_t(d) < B_t(d')$ and $B_{f_t(d')}(d') < 0$.*

Definition 1a relaxes Definition 1 in that it allows a group in a PA to prefer an alliance along the other dimension, with an additional restriction that, should this be the case, the potential ally along that other dimension has a strict *dis*-incentive to form it. Then, it is easy to check that Lemma 5 still holds,[18] but now region $C1$ contains a PA between WK and WL (as region B does) while, more importantly, region $C2$ contains a PA between BL and WL (as region A does). However, one can still argue that a race-based alliance is still more plausible than a class-based alliance, as explained in the following.

Given any $s_{WK} \leq 20\%$, we examine the potential alliance as a function of s_{BL}; the exercise is to draw a horizontal line in the unit simplex (such that $s_{WK} \leq 20\%$) to see how the potential alliance varies as the economy moves along the horizontal line. The result is depicted in Figure 2. When s_{BL} is sufficiently low, no PA exists; when it is moderate, a PA between BL and WL exists (not the case if Definition 1 is used); when it is sufficiently

[17] The critical value $s_{WK} = 20\%$ is calculated from the condition that $B_{WK}(R) = 0$ and $B_{BL}(C) = 0$; or the intersection of the bb and cc lines in the unit simplex in Figure 1.

[18] Lemma 5.2.ii under Definition 1a is a little less obvious. Suppose a PA between t and $f_t(d)$ and another between t and $f_t(d')$ co-exist. There are three cases — (1) $B_t(d) = B_t(d')$, (2) $B_t(d) > B_t(d')$, and (3) $B_t(d) < B_t(d')$. In case (2), the PA between t and $f_t(d')$ requires $B_{f_t(d)}(d) < 0$, which contradicts the condition needed for the PA between t and $f_t(d)$ to exist. By symmetry, case (3) leads to a similar contradiction. Hence case (1) is implied, which boils down to $s_{f_t(d)} = s_{f_t(d')}$.

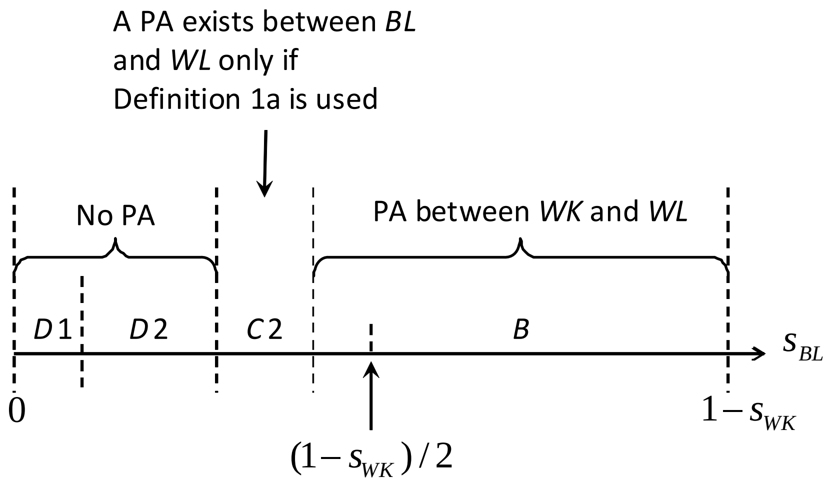

Figure 2: The potential alliance as a function of s_{BL} given that s_{WK} is relatively small (less than 20%). There is thus no PA when s_{BL} is small (regions $D1$ and $D2$); and there is one PA between WK and WL when s_{BL} is large (region B). Region $C2$ contains a PA between BL and WL if Definition 1a is used, but contains no PA if Definition 1 is used instead.

large (but it need not be larger than s_{WL}), a race-based PA exists between WL and WK. Since the second, intermediate range is "smaller" than the third range, a race-based alliance is still more relevant.

5.2 Generalization

Thus far we have studied the problem with the restriction that $s_{BK} = 0$. We argue that so long as its size is small enough, even if it's not zero, the main insight still holds. Recall that the area in Figure 1 of real interest is for which $s_{WK} \leq 20\%$. In what follows, we present simulation results by maintaining this focus and selecting representatively four cases with $s_{BK} = 1\%$, 2.5%, 5%, and 10%. The results are shown in the relevant areas of four normalized simplexes in Figure 3, similar to the unit simplex in Figure 1.

Three observations can be made. First, when the fraction of black capitalists is sufficiently small (e.g. in the case of $s_{BK} = 1\%$ or 2.5%), although there is a region in which the PA between BL and WL exists, the PA between WK and WL is still more plausible, by an argument analogous to that shown in Figure 2. This is consistent with the result we obtained previously when we assumed $s_{BK} = 0$.

Second, when the fraction of black capitalists is relatively large (e.g. $s_{BK} = 5\%$ or 10%), and if we focus on the case in which s_{WK} is very close to zero (i.e., at the bottom of the simplexes), the coalition between BK and BL is now more plausible. This is indeed an analogy of the first observation. The difference is that now s_{BK} is significantly larger than s_{WK}, approximating economies in which the elites are predominately black.

The third observation concerns the presence of a PA between WK and BK. Along each edge of each simplex (more prominent when s_{BK} becomes larger), there is a region in which such a PA exists. However, for the region along the left (right) edge, s_{BL} (s_{WL}) is rather small compared with s_{BK} (s_{WK}). Consider the case where $s_{BK} = 10\%$, for a PA to exist it between WK and BK requires that $s_{BL}/s_{BK} < 0.4$ (along the left edge) or that $s_{WL}/s_{WK} < 0.2$ (along the right edge). (Figure 4 plots such upper bounds in greater detail.) But the very notion that capitalists are a wealthy, minority group suggests that these circumstances are highly unlikely. Thus we maintain that a PA between WK and BK is unlikely to exist in a realistic environment.

Finally, one may wonder if the result differs once we relax Definition 1, say, to Definition 1a. Indeed, adopting the latter does not lead to significant changes. Our simulation has shown that the regions analogous to $C1$ and $C2$ in Figure 1 exist (albeit very small) when the fraction of black capitalists is extremely small (e.g. $s_{BK} = 0.1\%$), but disappear as soon as the fraction becomes larger (e.g. $s_{BK} \geq 0.2\%$). Therefore, Definition 1 can be used to study potential alliances without much loss of generality.

6 Extreme Identity

Thus far we have assumed moderate identity so that the total characteristic function is strictly superadditive. In case of stronger identity, the total characteristic function may

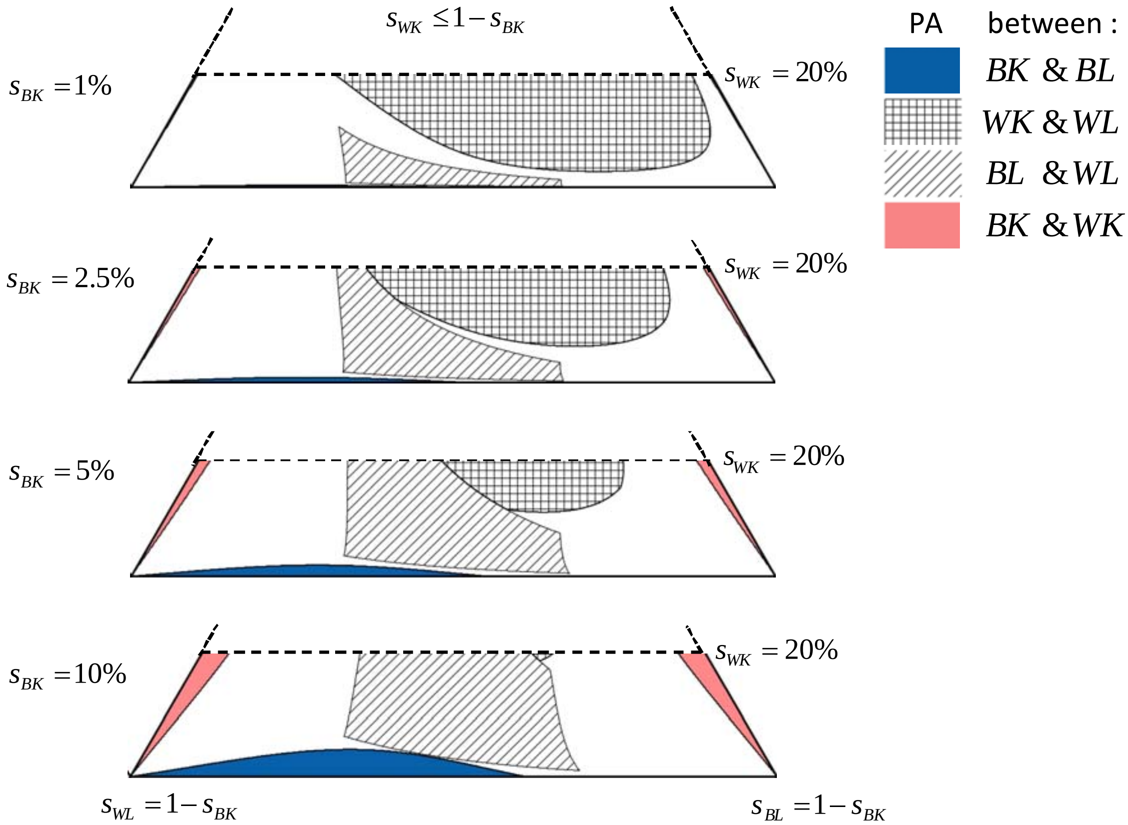

Figure 3: The simulation shows potential coalitions when s_{WK} is no larger than 20% and s_{BK} is set at four different levels. Regions with dark-blue color, grid pattern, diagonal pattern, and light-red color correspond respectively to the PA between BK and BL; WK and WL; BL and WL; and BK and WK.

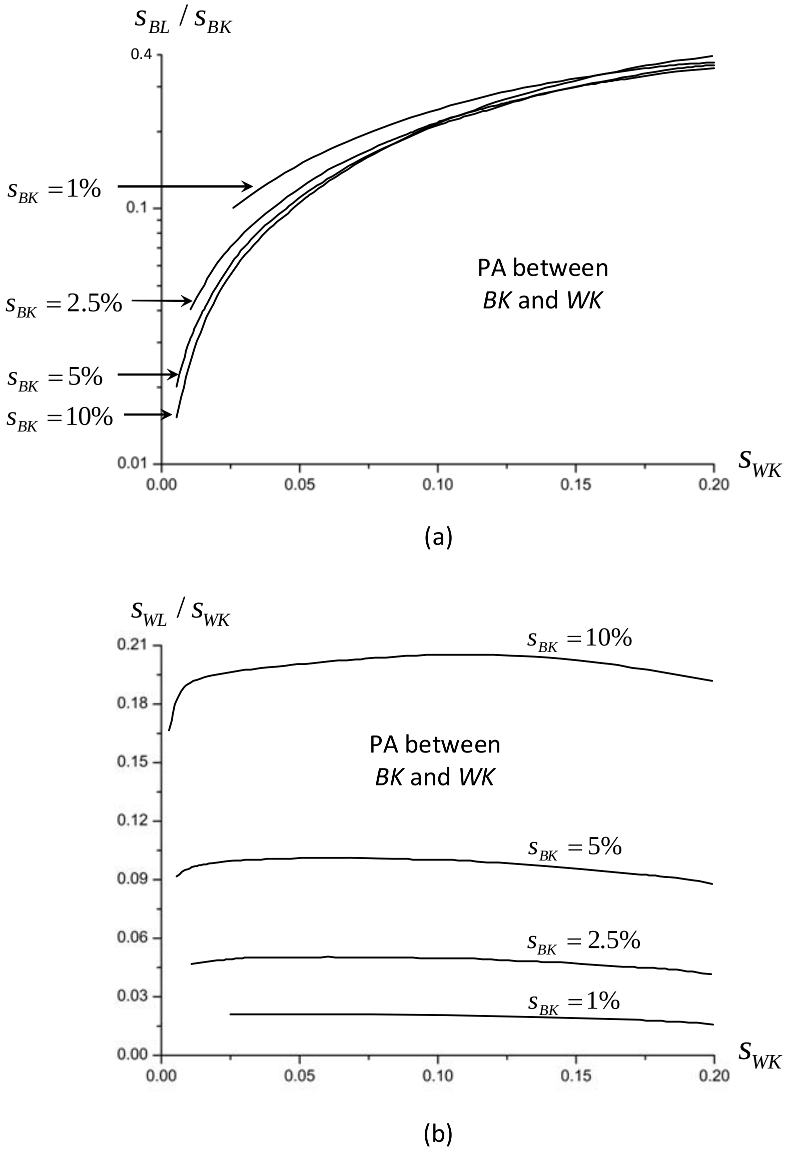

Figure 4: Panel (a) (Panel (b)) corresponds to the light-red region along the left (right) edge in Figure 3. Each curve is determined by a different value of s_{BK}. The PA between BK and WK exists only *under* each respective curve. The workers-capitalists ratio in the black (white) population has to be lower than 0.4 (0.2), which is highly unlikely in reality.

no longer be strictly superadditive, and some comparative statics obtained earlier may no longer hold. To see this, we adopt Assumption 3a.

Assumption 3a *The total characteristic function* $u(\cdot)$ *is otherwise strictly superadditive except that there exists a single pair of groups* j *and* k*, such that*

$$u(j \cup k) < u(j) + u(k).$$

Thus, group j and group k each prefers to work on its own, as opposed to working with each other. In this case, we say that j and k are *extremely hostile* to each other, or there is *extreme identity* or *extreme hostility.* The formation of the grand coalition is still efficient and so it will form.[19] Therefore, Assumption 3a is only a slight departure from Assumption 3. The case of extreme identity has implications in three aspects.

Material Foundation The following counter-intuitive result, in sharp contrast to Proposition 1.6, is obtained.

Proposition 4 *Under Assumptions 1, 2, and 3a, we have for* $n = 3$, $\partial\beta_i/\partial a_{jk} > 0$*; for any* $n \geq 4$,

$$\frac{\partial \beta_i}{\partial a_{jk}} \begin{cases} > 0 & \text{if} \quad (s_j + s_k)/N > \frac{n(n-1)(n-2)-6}{n(n-1)(n-2)}, \\ < 0 & \text{if} \quad s_i/N > \sqrt{\frac{6}{n(n-1)(n-2)-6}}. \end{cases}$$

The proposition says that i may benefit from increased intragroup amicability between j and k who are extremely hostile to each other. The case of $n = 3$ (i.e., with groups i, j, and k) is the easiest. Recall that each group's payoff is the group's average marginal contribution to the production of the total payoff when groups arrive sequentially in a random order. The presence of extreme identity between j and k now has an impact on i's payoff whenever group i is the last group to arrive. It turns out that an increase in the amicability between j and k now allows i to make a larger marginal contribution to the grand coalition, yielding a larger total payoff to i. The material payoff to i is also larger because its psychological payoff is independent of a_{jk} (or a_{kj}) and material payoff equals total payoff minus psychological payoff. For $n \geq 4$, this result still holds under certain conditions.

[19] When Assumption 3 does not hold, either the grand coalition does not form, or the grand coalition forms but a certain sub-coalition is not feasible in some intermediate sequence of group arrivals. In either case, the exact total payoffs that group i will get is different from (2). The details are relegated to Appendix A.

Alliance Formation in Class and Race Extreme identity also has implications for the incentive to form an alliance.

Proposition 5 *Suppose Assumption 1, 2, and 3a hold, there are four groups:* WK, WL*,* BK*, and* BL*, and* $N = 1$*. For all* (t, d)*, if* $s_{f_t(d')} = 0$ *and there is extreme identity between* $f_t(d)$ *and* $\tau = f_{f_t(d)}(d')$*, then we have* $B_t^I(d) = B_t^O(d)$ *given by*

$$B_t(d) \begin{cases} = 0 & \text{if} \quad s_{f_t(d)} = \{0, 1 - s_t\}, \\ < 0 & \text{if} \quad s_{f_t(d)} \in (0, 1 - s_t). \end{cases}$$

This result can be understood by letting $t = WK$, $d = R$, $f_t(d) = WL$, $\tau = BL$, and maintaining $f_t(d') = BK$ with $s_{BK} = 0$. Now, if there is an extreme identity between WL and BL, $B_{WK}(R) < 0$; enhancing amicability within the same race always reduces WK's material payoffs. This is in sharp contrast to Proposition 3.2, where such a change *could* benefit WK. With extreme identity, according to the manipulation interpretation, WK should now foster the intergroup amicability between different races, rather than exacerbating the differences.

Diversity and Economic Performance Easterly and Levine (1997) find that ethnic diversity can explain the low growth rates of African economies. Later work along this line supports the general relationship that having greater ethnic diversity lowers the growth rate, and that this negative effect is less pronounced in more developed economies (see, most notably, Alesina and La Ferrara 2000, 2005). Our framework can shed new light on this. Remember that in our framework the grand coalition need not form when the total characteristic function is not strictly superadditive. In that case, groups would trade material payoff for psychological payoff, and intragroup amicability and intergroup hostility could be conducive to poor economic performance.

To see this, let A be the coefficient to reflect the overall productivity of the economy. The characteristic function that describes the material output, v, can then be written as $v = A \times w$, where w is a normalized characteristic function and is strictly superadditive. Then the total welfare of any coalition R of groups becomes

$$u(R) = A \times w(R) + \sum_{j \subset R} \alpha_j(R),$$

which is strictly superadditive if and only if A is high enough. If A is low, $u(N)$ may not give the greatest total payoff (in this case, the extreme identity is not as mild as prescribed by Assumption 3a). Instead, there exists another coalitional structure, C^*, that maximizes the total payoff and will be formed. Hence, the actual GDP will be less than the potential GDP and we can define the efficiency loss as the ratio between this difference over the potential GDP.

Suppose starting from a very low initial coefficient of overall productivity $A = A_0$, A increases steadily over time. Then the efficiency loss will decrease over time[20] and the reason is that, as A increases, the economy transforms from a more segregated state to a less segregated state. Hence, this prediction is consistent with the aforementioned empirical findings.

In Easterly and Levine (1997) and Alesina and La Ferrara (2005), the cost of ethnic diversity is the inability to agree on, hence the under-provision of, common public goods. The benefit, however, lies in the variety of skills in production. Because the latter is more beneficial in more developed economies, the net, negative effect is thus smaller in those countries. In our framework, however, diversity is manifested in the partition of the population into groups, and the output gap is smaller in less segregated (hence more developed) economies. Despite similar implications, the underlying mechanism is rather different.

7 Concluding Remarks

In this paper, we have provided a simple theory of intergroup relationship in which material payoffs and psychological payoffs are substitutes. Whereas the concern about the former compels agents to work together, the concern about the latter strengthens their bargaining power in their share of output. This thus suggests that psychological preferences play a subtle role in affecting a group's material payoffs. Lacking a better name, we have called the shared psychological preferences in a group as the group's identity.

The main result is that there is indeed an economic or material foundation of such psychology. Generally speaking, the stronger a group's identity, the greater its material payoffs. We have derived interesting implications, in context of class and race, consistent

[20] The growth of A need not be exogenous. It is possible that a vicious cycle exists where A is too low for the economy to escape from its voluntary segregation. Then given two economies that are at the same initial productivity level to start with, the one that is more homogenous may be able to break free from the vicious cycle to become a developed state while the other that is plagued by strong identity may not.

with the observation that class has become a less relevant factor in politics. Capitalists, by definition, are a minority versus the working class. The two groups along the other, non-class dimensions (with different races or different religions) do not have this characteristic. It is this very subtle asymmetry between the class and non-class dimensions that leads to decline of class and salience of the other dimension.

This study has a few limitations. First, we have assumed that every agent in the economy belongs to some group and each group participates in the bargaining process; but in reality some groups are disenfranchised and their interests are not represented. But very often it is the presence of disenfranchised groups that motivate us to study the problem. Second, we do not study group formation or group dissolution; we simply assume that the memberships of all groups are exogenously given. While it is more difficult for an agent to change how she is viewed by other agents, she can always change how she views them. As Sen (2006) rightly points out, because of multiple identities, an individual has the freedom to choose which particular one to emphasize on. Finally, we have assumed that the interactions and surplus division between groups are determined through a random-order bargaining process, abstracting away political institutions as well as the possibility of violence. All these are interesting issues for future research.

References

[1] Akerlof, George A. and Rachel E. Kranton. (2000). "Economics and Identity," *Quarterly Journal of Economics*, 115(3): 715–753.

[2] Akerlof, George A. and Rachel E. Kranton. (2002). "Identity and Schooling: Some Lessons for the Economics of Education," *Journal of Economic Literature*, 40(4): 1167–1201.

[3] Akerlof, George A. and Rachel E. Kranton. (2005). "Identity and the Economics of Organizations," *Journal of Economic Perspectives*, 19(1): 9–32.

[4] Alesina, Alberto and Eliana La Ferrara. (2000). "Participation in Heterogeneous Communities," *Quarterly Journal of Economics*, 115(3): 847–904.

[5] Alesina, Alberto and Eliana La Ferrara. (2005). "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, 43(3): 762–800.

[6] Aumann, Robert J. and Roger M. Myerson. (1986). "Endogenous Formation of Links Between Players and of Coalitions: An Application of the Shapley Value," in *The Shapley value: Essays in Honor of Lloyd S. Shapley*, Alvin E. Roth, ed. Cambridge; New York: Cambridge University Press.

[7] Bénabou, Roland and Jean Tirole. (2006). "Belief in a Just World and Redistributive Politics," *Quarterly Journal of Economics*, 121(2): 699–746.

[8] Bénabou, Roland and Jean Tirole. (2010). "Identity, Morals and Taboos: Beliefs as Assets," *Quarterly Journal of Economics*, forthcoming.

[9] Berman, Eli. (2000). "Sect, Subsidy, and Sacrifice: An Economist's View of Ultra-Orthodox Jews," *Quarterly Journal of Economics*, 115(3): 905–953.

[10] Berman, Eli. (2005). "Hamas, Taliban and the Jewish Underground: An Economist's View of Radical Religious Militias," mimeo, UC San Diego.

[11] Bernheim, B. Douglas. (1994). "A Theory of Conformity," *Journal of Political Economy*, 102(5): 841–877.

[12] Bisin, Alberto, Eleonora Patacchini, Thierry Verdier, and Yves Zenou. (2010). "Bend It Like Beckham: Ethnic Identity and Integration," mimeo, NYU.

[13] Bisin, Alberto and Thierry Verdier. (2000). ""Beyond the Melting Pot": Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits," *Quarterly Journal of Economics*, 115(3): 955–988.

[14] Bisin, Alberto and Thierry Verdier. (2001). "The Economics of Cultural Transmission and the Dynamics of Preferences," *Journal of Economic Theory*, 97(2): 298–319.

[15] Bolton, Gary and Axel Ockenfels. (2002). "ERC: A Theory of Equity, Reciprocity and Competition," *American Economic Review*, 90(1): 166–193.

[16] Bowles, Samuel and Herbert Gintis. (2004). "Persistent Parochialism: Trust and Exclusion in Ethnic Networks," *Journal of Economic Behavior & Organization*, 55(1): 1–23.

[17] Brewer, Marilynn B. and Norman Miller. (1996). *Intergroup Relations*, Buckingham: Open University Press.

[18] Carmines, Edward G. and James A. Stimson. (1989). *Issue Evolution: Race and the Transformation of American Politics*. Princeton: Princeton University Press.

[19] Caselli, Francesco and Wilbur J. Coleman II. (2010). "On the Theory of Ethnic Conflict," mimeo, Duke.

[20] Charness, Gary, Luca Rigotti, and Aldo Rustichini. (2007). "Individual Behavior and Group Membership," *American Economic Review*, 97(4): 1340–1352.

[21] Chen, Yan and Sherry Xin Li. (2009) "Group Identity and Social Preferences," *American Economic Review* 99(1): 431–457.

[22] Choi, Jung-Kyoo and Samuel Bowles. (2007). "The Coevolution of Parochial Altruism and War," *Science*, 318(5850): 636–640.

[23] Currarini, Sergio, Matthew O. Jackson, and Paolo Pin. (2009). "An Economic Model of Friendship: Homophily, Minorities, and Segregation," *Econometrica*, 77(4): 1003–1045.

[24] Darity, William A., Patrick L. Masonc, and James B. Stewart. (2006). "The Economics of Identity: The Origin and Persistence of Racial Identity Norms," *Journal of Economic Behavior & Organization*, 60(3): 283–305.

[25] Dasgupta, Ani and Y. Stephen Chiu. (1998). "On Implementation via Simple Demand Commitment Games," *International Journal of Game Theory*, 27(2): 161–190.

[26] Dixit, Avinash K. and Joseph E. Stiglitz. (1977). "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67(3): 297–308.

[27] Drucker, Peter F. (1998). *Adventures of a Bystander*, New York: John Wiley.

[28] Easterly, William and Ross Levine. (1997). "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics*, 112(4): 1203–1250.

[29] Eckel, Catherine C. and Philip J. Grossman. (2005). "Managing Diversity by Creating Team Identity," *Journal of Economic Behavior & Organization*, 58(3): 371–392.

[30] Esteban, Joan and Debraj Ray. (2008). "On the Salience of Ethnic Conflict," *American Economic Review*, 98(5): 2185–2202.

[31] Fearon, James D. (1999). "Why Ethnic Politics and 'Pork' Tend to Go Together," mimeo, Stanford.

[32] Glaeser, Edward L.. (2005). "The Political Economy of Hatred," *Quarterly Journal of Economics*, 120(1): 45–86.

[33] Gul, Faruk. (1989). "Bargaining Foundations of Shapley Value," *Econometrica* 57(1): 81–95.

[34] Hart, Sergiu and Andreu Mas-Colell. (1996). "Bargaining and Value," *Econometrica*, 64(2): 357–380.

[35] Hobsbawn, Eric J.. (1994). *The Age of Extremes: A History of the World, 1914-1991.* New York: Pantheon Books.

[36] Huckfeldt, Robert R. and Carol W. Kohfeld. (1989). *Race and the Decline of Class in American Politics.* Urbana: University of Illinois Press.

[37] Iannaccone, Laurence R. (1992). "Sacrifice and Stigma: Reducing Free-riding in Cults, Communes, and Other Collectives," *Journal of Political Economy*, 100(2): 271–291.

[38] Kitschelt, Herbert. (1994). *The Transformation of European Social Democracy.* Cambridge; New York: Cambridge University Press.

[39] Laver, Michael and W. Ben Hunt. (1992). *Policy and Party Competition*. New York: Routledge.

[40] Lazarsfeld, Paul F. and Robert K. Merton. (1954). "Friendship as a Social Process: a Substantive and Methodological Analysis," in *Freedom and Control in Modern Society*, Monroe Berger, Theodore Abel, and Charles H. Page, ed. New York: Van Nostrand.

[41] Luttmer, Erzo F. P.. (2001). "Group Loyalty and the Taste of Redistribution," *Journal of Political Economy*, 109(3): 500–528.

[42] Maskin, Eric. (2003). "Bargaining, Coalitions, and Externalities," mimeo, Institute of Advanced Studies.

[43] Mui, Vai-Lam. (1999). "Information, Civil Liberties, and the Political Economy of Witch-Hunts," *Journal of Law, Economics and Organization*, 15(2): 503–525.

[44] Poole, Keith T. and Howard Rosenthal. (1991). "Patterns of Congressional Voting," *American Journal of Political Science*, 35(1): 228–278.

[45] Przeworski, Adam and John Sprague. (1986). *Paper Stones: A History of Electoral Socialism*. Chicago: University of Chicago Press.

[46] Putterman, Louis. (1997). "Why Have the Rabble Not Redistributed the Wealth? On the Stability of Democracy and Unequal Property," in *Property Relations, Incentives, and Welfare*, John E. Roemer, ed. London: Macmillan.

[47] Roemer, John E.. (1998). "Why the Poor Do Not Expropriate the Rich: An Old Argument in New Garb," *Journal of Public Economics*, 70(3), 399–424.

[48] Romer, Paul. (1995). "Preferences, Promises, and the Politics of Entitlement," in *Individual and Social Responsibility: Child Care, Education, Medical Care, and Long Term Care in America*, Victor R. Fuchs, ed. Chicago and London: University of Chicago Press.

[49] Segal, Ilya. (2003). "Collusion, Exclusion, and Inclusion in Random-Order Bargaining," *Review of Economic Studies*, 70(2): 439–460.

[50] Sen, Amartya. (2006). *Identity and Violence: the Illusion of Destiny*. New York: W.W. Norton & Company.

[51] Shapley, Lloyd S.. (1953). "A Value for n-Person Games," in *Contributions to the Theory of Games II*, Harold W. Kuhn and Albert W. Tucker, ed. Princeton, N.J.: Princeton University Press.

[52] Shayo, Moses. (2009). "A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution," *American Political Science Review*, 103(2): 147–174.

[53] Sumner, William G.. (1906). *Folkways*. New York: Ginn.

[54] Tajfel, Henri, M. G. Billig, R. P. Bundy, and Claude Flament. (1971). "Social Categorization and Intergroup Behaviour," *European Journal of Social Psychology*, 1(2): 149–178.

[55] Tajfel, Henri and John Turner. (1979). "An Integrative Theory of Intergroup Conflict," *The Social Psychology of Intergroup Relations*, William G. Austin and Stephen Worchel, ed. Monterey, CA: Brooks-Cole.

[56] Turner, John C., Michael A. Hogg, Penelope J. Oakes, Stephen D. Reicher, and Margaret S. Wetherell. (1987). *Rediscovering the Social Group: A Self-Categorization Theory*. New York: Basil Blackwell.

[57] Veblen, Thorstein. (1899). *The Theory of the Leisure Class: An Economic Study in the Evolution of Institutions*, New York: The Macmillan Company; London: Macmillan & Co., ltd..

Appendix A: Payoff Determination under More General Conditions

Here we present a generalization of the bargaining solution which deals with non-superadditive total utility function. This formulation, which is used in the part that deals with extreme identity, is reduced to the same formulation as stated in (2) when u is superadditive. Let R be any set of agents that are in a union of groups. Define $C \equiv \{R_1, \cdots, R_m\}$ to be a partition respecting group boundary if the following three properties hold: (i) $\cup_{i=1,\cdots,m} R_i = R$ (exhaustiveness); (ii) for all $i, j \in \{1, \cdots, m\}$ and $i \neq j$, $R_i \cap R_j = \emptyset$ (mutual exclusivity); (iii) for all $i \in \{1, \cdots, m\}$, R_i is a collection of groups (respecting group boundary). Let $\mathcal{C}(R)$ be the set of all such possible partitions given R. Define

$$\begin{aligned} C^*(R) &\equiv \arg\max_{C \in \mathcal{C}(R)} \sum_{R_j \in C} u(R_j) \\ \text{and} \qquad \omega(R) &\equiv \sum_{R_j \in C^*(R)} u(R_j). \end{aligned}$$

The intuition is that $C^*(R)$ is the coalition structure that maximizes the total utility given R, and $\omega(R)$ is the corresponding maximized total payoff. Then define each group's total payoff by

$$\phi_i(N) = \sum_{T \not\supseteq i} \frac{|T|!\,(n - |T| - 1)!}{n!} \left[w(T \cup i) - w(T)\right]. \tag{7}$$

If u is strictly superadditive, $\omega(R) \equiv u(R)$ and (7) reduces to (2). If strict superadditivity of $u(\cdot)$ holds except that, there exist two groups j and k such that $u(j \cup k) < u(j) + u(k)$ (this is Assumption 3a in the main text), then (7) has two implications. First, $C^*(N)$ continues to correspond to the formation of the grand coalition; in other words, according to (7), the sum of each group's total payoff is still equal to $u(N)$. Second, for all $i \neq j$ or k,

$$\begin{aligned} \phi_i(N) &= \sum_{T \not\supseteq i, T \neq j \cup k} \frac{|T|!\,(n - |T| - 1)!}{n!} \left[u(T \cup i) - u(T)\right] \\ &\quad + \frac{2}{n(n-1)(n-2)!} \left[u(j \cup k \cup i) - u(j) - u(k)\right], \end{aligned}$$

which is the same as (2), except for the second term on the RHS. For j,

$$\phi_j(N) = \sum_{T \not\supseteq j, T \neq k} \frac{|T|!\,(n - |T| - 1)!}{n!} \left[u(T \cup j) - u(T)\right] + \frac{u(j)}{n(n-1)},$$

which is the same as (2), except for the second term on the RHS.

Appendix B: Proofs

Proof of Proposition 1. Result 1 is obvious. We show result 2 at the end of the proof. For result 3, we show

$$\begin{aligned}\frac{\partial \phi_i}{\partial a_{ii}} &= \sum_{T \supsetneq\ i} \frac{|T|!\,(n-|T|-1)!}{n!}\left(\frac{\partial u\,(T\cup i)}{\partial a_{ii}} - \frac{\partial u(T)}{\partial a_{ii}}\right)\\ &= \sum_{T \supsetneq\ i} \frac{|T|!\,(n-|T|-1)!}{n!}\left(\frac{\partial u\,(T\cup i)}{\partial a_{ii}}\right)\\ &= \sum_{T\ \supsetneq i} \frac{|T|!\,(n-|T|-1)!}{n!}\left(\frac{s_i^2}{s_T+s_i}\right) > 0.\end{aligned}$$

Since $\partial\alpha_i/\partial a_{ii} = s_i^2/N$, we then have

$$\begin{aligned}\frac{\partial \beta_i}{\partial a_{ii}} &= \frac{\partial \phi_i}{\partial a_{ii}} - \frac{\partial \alpha_i}{\partial a_{ii}} = \sum_{T \supsetneq i} \frac{|T|!\,(n-|T|-1)!}{n!}\left(\frac{s_i^2}{s_T+s_i}\right) - \frac{s_i^2}{N}\\ &= \sum_{T \supsetneq i} \frac{|T|!\,(n-|T|-1)!}{n!}\left(\frac{s_i^2}{s_T+s_i} - \frac{s_i^2}{N}\right) > 0. \qquad (8)\end{aligned}$$

For result 4, we show

$$\begin{aligned}\frac{\partial \phi_i}{\partial a_{jj}} &= \sum_{T \supsetneq i, T\supseteq j} \frac{|T|!\,(n-|T|-1)!}{n!}\left(\frac{\partial u\,(T\cup i)}{\partial a_{jj}} - \frac{\partial u\,(T)}{\partial a_{jj}}\right)\\ &\quad + \sum_{T \supsetneq i,j} \frac{|T|!\,(n-|T|-1)!}{n!}\left(\frac{\partial u\,(T\cup i)}{\partial a_{jj}} - \frac{\partial u\,(T)}{\partial a_{jj}}\right)\\ &= \sum_{T \supsetneq i, T\supseteq j} \frac{|T|!\,(n-|T|-1)!}{n!}\left(\frac{\partial u\,(T\cup i)}{\partial a_{jj}} - \frac{\partial u\,(T)}{\partial a_{jj}}\right)\\ &= \sum_{T \supsetneq i, T\supseteq j} \frac{|T|!\,(n-|T|-1)!}{n!}\left(\frac{s_j^2}{s_T+s_i} - \frac{s_j^2}{s_T}\right) < 0.\end{aligned}$$

Noting that $\partial\alpha_i/\partial a_{jj} = 0$, we then get

$$\frac{\partial \beta_i}{\partial a_{jj}} = \frac{\partial \phi_i}{\partial a_{jj}} = \sum_{T \supsetneq i, T\supseteq j} \frac{|T|!\,(n-|T|-1)!}{n!}\left(\frac{s_j^2}{s_T+s_i} - \frac{s_j^2}{s_T}\right) < 0. \qquad (9)$$

For result 5, we show

$$\begin{aligned}\frac{\partial \phi_i}{\partial a_{ji}} &= \sum_{T \not\supseteq i, T \supseteq j} \frac{|T|!\,(n-|T|-1)!}{n!} \left(\frac{\partial u\,(T \cup i)}{\partial a_{ji}} - \frac{\partial u(T)}{\partial a_{ji}} \right) \\ &\quad + \sum_{T \not\supseteq i, j} \frac{|T|!\,(n-|T|-1)!}{n!} \left(\frac{\partial u\,(T \cup i)}{\partial a_{ji}} - \frac{\partial u(T)}{\partial a_{ji}} \right) \\ &= \sum_{T \not\supseteq i, T \supseteq j} \frac{|T|!\,(n-|T|-1)!}{n!} \left(\frac{\partial u\,(T \cup i)}{\partial a_{ji}} \right) \\ &= \sum_{T \not\supseteq i, T \supseteq j} \frac{|T|!\,(n-|T|-1)!}{n!} \left(\frac{s_i s_j}{s_T + s_i} \right) > 0.\end{aligned}$$

Since $\partial \alpha_i / \partial a_{ji} = 0$, we have

$$\frac{\partial \beta_i}{\partial a_{ji}} = \frac{\partial \phi_i}{\partial a_{ji}} = \sum_{T \not\supseteq i, T \supseteq j} \frac{|T|!\,(n-|T|-1)!}{n!} \left(\frac{s_i s_j}{s_T + s_i} \right) > 0. \tag{10}$$

For result 6, we show

$$\begin{aligned}\frac{\partial \phi_i}{\partial a_{jk}} &= \sum_{T \not\supseteq i, T \supseteq j,k} \frac{|T|!\,(n-|T|-1)!}{n!} \left(\frac{\partial u\,(T \cup i)}{\partial a_{jk}} - \frac{\partial u(T)}{\partial a_{jk}} \right) \\ &= \sum_{T \not\supseteq i, T \supseteq j,k} \frac{|T|!\,(n-|T|-1)!}{n!} \left(\frac{s_j s_k}{s_T + s_i} - \frac{s_j s_k}{s_T} \right) < 0.\end{aligned}$$

Noting that $\partial \alpha_i / \partial a_{jk} = 0$, we have

$$\frac{\partial \beta_i}{\partial a_{jk}} = \frac{\partial \phi_i}{\partial a_{jk}} = \sum_{T \not\supseteq i, T \supseteq j,k} \frac{|T|!\,(n-|T|-1)!}{n!} \left(\frac{s_j s_k}{s_T + s_i} - \frac{s_j s_k}{s_T} \right) < 0. \tag{11}$$

For result 7, we show

$$\begin{aligned}\frac{\partial \phi_i}{\partial a_{ij}} &= \sum_{T \not\supseteq i, T \supseteq j} \frac{|T|!\,(n-|T|-1)!}{n!} \left(\frac{\partial u\,(T \cup i)}{\partial a_{ij}} - \frac{\partial u\,(T)}{\partial a_{ij}} \right) \\ &\quad + \sum_{T \not\supseteq i, j} \frac{|T|!\,(n-|T|-1)!}{n!} \left(\frac{\partial u\,(T \cup i)}{\partial a_{ij}} - \frac{\partial u\,(T)}{\partial a_{ij}} \right) \\ &= \sum_{T \not\supseteq i, T \supseteq j} \frac{|T|!\,(n-|T|-1)!}{n!} \left(\frac{\partial u\,(T \cup i)}{\partial a_{ij}} \right) \\ &= \sum_{T \not\supseteq i, T \supseteq j} \frac{|T|!\,(n-|T|-1)!}{n!} \frac{s_i s_j}{s_T + s_i} > 0.\end{aligned}$$

Since $\partial\alpha_i/\partial a_{ii} = s_i^2/N$, we then have

$$\begin{aligned}
\frac{\partial\beta_i}{\partial a_{ij}} &= \frac{\partial\phi_i}{\partial a_{ij}} - \frac{\partial\alpha_i}{\partial a_{ij}} \\
&= \sum_{T\not\supseteq i, T\supseteq j} \frac{|T|!\,(n-|T|-1)!}{n!} \frac{s_i s_j}{s_T + s_i} - \frac{s_i s_j}{N}. \qquad (12)
\end{aligned}$$

Note that

$$\begin{aligned}
\sum_{T\not\supseteq i, T\supseteq j} \frac{|T|!\,(n-|T|-1)!}{n!} &= \sum_{|T|=1}^{n-1} \begin{pmatrix} n-2 \\ |T|-1 \end{pmatrix} \frac{|T|!\,(n-|T|-1)!}{n!} \\
&= \sum_{|T|=1}^{n-1} \frac{|T|}{n\,(n-1)} = \frac{1}{2},
\end{aligned}$$

We first derive a sufficient condition for $\partial\beta_i/\partial a_{ij}$ to be strictly negative. Note that the RHS of (12) can be written as

$$\begin{aligned}
& \frac{1}{n}\frac{s_i s_j}{N} + \sum_{T\not\supseteq i, T\supseteq j, T\neq N\setminus i} \frac{|T|!\,(n-|T|-1)!}{n!} \frac{s_i s_j}{s_T + s_i} - \frac{s_i s_j}{N} \\
< & \sum_{T\not\supseteq i, T\supseteq j, T\neq N\setminus i} \frac{|T|!\,(n-|T|-1)!}{n!} \frac{s_i s_j}{s_j + s_i} - \frac{n-1}{n}\frac{s_i s_j}{N} \\
= & \left(\frac{1}{2} - \frac{1}{n}\right) \frac{s_i s_j}{s_j + s_i} - \frac{n-1}{n}\frac{s_i s_j}{N}
\end{aligned}$$

which is strictly negative if and only if

$$\left(\frac{1}{2} - \frac{1}{n}\right) \frac{s_i s_j}{s_j + s_i} < \frac{n-1}{n}\frac{s_i s_j}{N} \Leftrightarrow \frac{s_i + s_j}{N} > \frac{1}{2}\frac{n-2}{n-1}.$$

Similarly, we can derive a sufficient condition for $\partial\beta_i/\partial a_{ij}$ to be strictly positive. The RHS of (12) equals

$$\begin{aligned}
& \frac{(n-2)!}{n!}\left(\frac{s_i s_j}{s_i + s_j}\right) + \sum_{T\not\supseteq i, T\supseteq j, T\neq j} \frac{|T|!\,(n-|T|-1)!}{n!} \frac{s_i s_j}{s_T + s_i} - \frac{s_i s_j}{N} \\
> & \frac{(n-2)!}{n!}\left(\frac{s_i s_j}{s_i + s_j}\right) + \sum_{T\not\supseteq i, T\supseteq j, T\neq j} \frac{|T|!\,(n-|T|-1)!}{n!} \frac{s_i s_j}{N} - \frac{s_i s_j}{N} \\
= & \frac{(n-2)!}{n!}\left(\frac{s_i s_j}{s_i + s_j}\right) + \left(\frac{1}{2} - \frac{(n-2)!}{n!}\right) \frac{s_i s_j}{N} - \frac{s_i s_j}{N}
\end{aligned}$$

which is strictly positive if and only if

$$\frac{(n-2)!}{n!}\left(\frac{s_i s_j}{s_i+s_j}\right) > \frac{1}{2}\frac{s_i s_j}{N} + \frac{(n-2)!}{n!}\frac{s_i s_j}{N} \Leftrightarrow \frac{s_i+s_j}{N} < \frac{2}{n(n-1)+2}.$$

When $n = 3$, the RHS of these two sufficient conditions coincide, which delivers $\partial\beta_i/\partial a_{ij} > 0$ if and only if $(s_i + s_j)/N < 1/4$.

Finally, we show result 2. For (i), note that if any of s_i, s_j, and s_k is zero, the relevant equations in (8)-(12) becomes zero too. To see (ii), consider the case of adding a new group $\ell \neq i, j, k$ with $s_\ell = 0$ into the economy (the case of removing one can be seen by symmetry). Suppose the grand coalition then changes from N to $N \cup \ell \equiv N'$. In all equations (8)-(12), any T in the original economy corresponds to two cases in the new economy, $T' = T$ or $T' = T \cup \ell$, with equal probability in random arrival. Whichever being true, we always have $s_{T'} = s_T$. Therefore, adding such group ℓ leads to no change in (8)-(12), which completes the proof. ■

Proof of Lemma 1. The RHS equals

$$\begin{aligned}
& -\sum_{j\neq i}\left[\sum_{T\not\ni i, T\ni j}\frac{|T|!\,(n-|T|-1)!}{n!}\left(\frac{s_i s_j}{s_T+s_i}\right) - \frac{s_i s_j}{N}\right] \\
= & -\sum_{T\not\ni i, T\neq\phi}\frac{|T|!\,(n-|T|-1)!}{n!}\frac{s_i s_T}{s_T+s_i} + \frac{s_i(N-s_i)}{N} \\
= & -\sum_{T\not\ni i, T\neq\phi}\frac{|T|!\,(n-|T|-1)!}{n!}\frac{s_i(s_T+s_i-s_i)}{s_T+s_i} + s_i - \frac{s_i^2}{N} \\
= & -\sum_{T\not\ni i, T\neq\phi}\frac{|T|!\,(n-|T|-1)!}{n!}s_i + \sum_{T\not\ni i, T\neq\phi}\frac{|T|!\,(n-|T|-1)!}{n!}\frac{s_i^2}{s_T+s_i} + s_i - \frac{s_i^2}{N} \\
= & -\frac{n!-(n-1)!}{n!}s_i + \sum_{T\not\ni i, T\neq\phi}\frac{|T|!\,(n-|T|-1)!}{n!}\frac{s_i^2}{s_T+s_i} + s_i - \frac{s_i^2}{N} \\
= & \ \frac{s_i}{n} + \sum_{T\not\ni i, T\neq\phi}\frac{|T|!\,(n-|T|-1)!}{n!}\frac{s_i^2}{s_T+s_i} - \frac{s_i^2}{N} \\
= & \sum_{T\not\ni i}\frac{|T|!\,(n-|T|-1)!}{n!}\frac{s_i^2}{s_T+s_i} - \frac{s_i^2}{N}
\end{aligned}$$

which is just the LHS. ■

Proof of Lemma 2. For result 1, by Proposition 1.2, it is w.l.o.g. to consider $n = 2$. By definitions, we then have

$$\begin{aligned}
\frac{\partial\beta_1}{\partial a_{1,1}} &= \frac{1}{2}\left(\frac{s_1^2}{0+s_1}-\frac{s_1^2}{N}\right)+\frac{1}{2}\left(\frac{s_1^2}{s_n+s_1}-\frac{s_1^2}{N}\right)=\frac{1}{2}s_1\left(1-\frac{s_1}{N}\right),\\
\frac{\partial\beta_1}{\partial a_{n,n}} &= \frac{1}{2}\left(\frac{s_n^2}{s_n+s_1}-\frac{s_n^2}{s_n}\right)=-\frac{1}{2}s_n\left(1-\frac{s_n}{N}\right).
\end{aligned}$$

It follows immediately that $\partial\beta_1/\partial a_{1,1}+\partial\beta_1/\partial a_{n,n}=0$.

For result 2, the case when either $s_{n-1}=0$ or $s_n=0$ is just result 1. If instead $s_{n-1},s_n>0$, it is again w.l.o.g. to consider $n=3$. By definitions, we have

$$\begin{aligned}
\frac{\partial\beta_1}{\partial a_{1,1}} &= \frac{s_1^2}{6}\left(\frac{2}{s_1}+\frac{1}{s_{n-1}+s_1}+\frac{1}{s_n+s_1}-\frac{4}{N}\right),\\
\frac{\partial\beta_1}{\partial a_{n-1,n}}+\frac{\partial\beta_1}{\partial a_{n,n-1}} &= \frac{2s_{n-1}s_n}{3}\left(\frac{1}{N}-\frac{1}{s_{n-1}+s_n}\right),\\
\frac{\partial\beta_1}{\partial a_{n-1,n-1}} &= \frac{s_{n-1}^2}{6}\left(\frac{1}{s_{n-1}+s_1}-\frac{1}{s_{n-1}}+\frac{2}{N}-\frac{2}{s_{n-1}+s_n}\right),\\
\frac{\partial\beta_1}{\partial a_{n,n}} &= \frac{s_n^2}{6}\left(\frac{1}{s_n+s_1}-\frac{1}{s_n}+\frac{2}{N}-\frac{2}{s_{n-1}+s_n}\right).
\end{aligned}$$

Summing all of them, substituting $s_{n-1}=N-s_1-s_n$, and manipulating yields

$$\frac{\partial\beta_1}{\partial a_{1,1}}+\sum_{i,j\in\{n-1,n\}}\frac{\partial\beta_1}{\partial a_{i,j}}=-\frac{s_1s_n\left(N-s_1-s_n\right)\left(N+s_1\right)}{3N\left(s_1+s_n\right)\left(N-s_n\right)}.$$

Since by assumption $0<s_n<1-s_1$, the RHS is strictly negative. ■

Proof of Lemma 3. Consider in general four groups t, $f_t(d)$, $f_t(d')$, and $\tau=f_{f_t(d)}(d')$. Then expanding terms in (5) and (6) yields

$$\begin{aligned}
B_t^I(d) &= \frac{\partial\beta_t}{\partial a_{t,t}}+\frac{\partial\beta_t}{\partial a_{t,f_t(d)}}+\frac{\partial\beta_t}{\partial a_{f_t(d),t}}+\frac{\partial\beta_t}{\partial a_{f_t(d),f_t(d)}}\\
&\quad+\frac{\partial\beta_t}{\partial a_{f_t(d'),f_t(d')}}+\frac{\partial\beta_t}{\partial a_{f_t(d'),\tau}}+\frac{\partial\beta_t}{\partial a_{\tau,f_t(d')}}+\frac{\partial\beta_t}{\partial a_{\tau,\tau}},\\
B_t^O(d) &= -\left(\frac{\partial\beta_t}{\partial a_{t,f_t(d')}}+\frac{\partial\beta_t}{\partial a_{f_t(d),f_t(d')}}+\frac{\partial\beta_t}{\partial a_{t,\tau}}+\frac{\partial\beta_t}{\partial a_{f_t(d),\tau}}\right.\\
&\quad\left.+\frac{\partial\beta_t}{\partial a_{f_t(d'),t}}+\frac{\partial\beta_t}{\partial a_{f_t(d'),f_t(d)}}+\frac{\partial\beta_t}{\partial a_{\tau,t}}+\frac{\partial\beta_t}{\partial a_{\tau,f_t(d)}}\right).
\end{aligned}$$

Hence, by Lemma 1, $B_t^I(d) = B_t^O(d)$ is shown by substituting

$$\begin{aligned}
\frac{\partial \beta_t}{\partial a_{t,t}} &= -\frac{\partial \beta_t}{\partial a_{t,f_d(t)}} - \frac{\partial \beta_t}{\partial a_{t,f_t(d')}} - \frac{\partial \beta_t}{\partial a_{t,\tau}}, \\
\frac{\partial \beta_t}{\partial a_{f_d(t),f_d(t)}} &= -\frac{\partial \beta_t}{\partial a_{f_d(t),t}} - \frac{\partial \beta_t}{\partial a_{f_t(d),f_t(d')}} - \frac{\partial \beta_t}{\partial a_{f_t(d),\tau}}, \\
\frac{\partial \beta_t}{\partial a_{f_t(d'),f_t(d')}} &= -\frac{\partial \beta_t}{\partial a_{f_t(d'),t}} - \frac{\partial \beta_t}{\partial a_{f_t(d'),f_t(d)}} - \frac{\partial \beta_t}{\partial a_{f_t(d'),\tau}}, \\
\frac{\partial \beta_t}{\partial a_{\tau,\tau}} &= -\frac{\partial \beta_t}{\partial a_{\tau,t}} - \frac{\partial \beta_t}{\partial a_{\tau,f_t(d)}} - \frac{\partial \beta_t}{\partial a_{\tau,f_t(d')}}
\end{aligned}$$

into $B_t^I(d)$. ■

Proof of Proposition 2. Result 1 being obvious, we show result 2 as follows. Consider again four groups t, $f_t(d)$, $f_t(d')$, and τ. It is more convenient to express $B_t(d)$ by $B_t^O(d)$ (as in the proof of Lemma 3). If $s_{f_t(d')} = 0$, by equations (9)-(12) and using Proposition 1.2, we have

$$B_t(d) = -\left(\frac{\partial \beta_t}{\partial a_{t,\tau}} + \frac{\partial \beta_t}{\partial a_{\tau,t}} + \frac{\partial \beta_t}{\partial a_{f_t(d),\tau}} + \frac{\partial \beta_t}{\partial a_{\tau,f_t(d)}}\right)$$

Expanding each term and using $s_{f_t(d')} = 0$, we obtain

$$\begin{aligned}
\frac{\partial \beta_t}{\partial a_{t,\tau}} + \frac{\partial \beta_t}{\partial a_{\tau,t}} &= \frac{s_t s_\tau}{3}\left(\frac{1}{s_t + s_\tau} - 1\right), \\
\frac{\partial \beta_t}{\partial a_{f_t(d),\tau}} + \frac{\partial \beta_t}{\partial a_{\tau,f_t(d)}} &= \frac{2 s_{f_t(d)} s_\tau}{3}\left(1 - \frac{1}{s_{f_t(d)} + s_\tau}\right).
\end{aligned}$$

Therefore, simplifying terms and substituting $s_\tau = 1 - s_t - s_{f_t(d)}$ yields

$$B_t(d) = \frac{s_t}{3(1-s_t)} \frac{s_{f_t(d)}\left(1 - s_t - s_{f_t(d)}\right)\left(1 + s_t - 2 s_{f_t(d)}\right)}{1 - s_{f_t(d)}}.$$

The result then follows immediately. If instead $s_\tau = 0$, the same result holds obviously by symmetry. ■

Proof of Lemma 4. Given $s_\tau = 0$, we use Proposition 2.2 (and its proof) to derive the difference $B_t(d) - B_t(d')$ as follows.

$$\begin{aligned}
B_t(d) - B_t(d') &= \frac{\partial \beta_t}{\partial a_{t,f_t(d)}} + \frac{\partial \beta_t}{\partial a_{f_t(d),t}} - \frac{\partial \beta_t}{\partial a_{t,f_t(d')}} - \frac{\partial \beta_t}{\partial a_{f_t(d'),t}} \\
&= \frac{s_t s_{f_t(d)}}{3}\left(\frac{1}{s_t + s_{f_t(d)}} - 1\right) - \frac{s_t s_{f_t(d')}}{3}\left(\frac{1}{s_t + s_{f_t(d')}} - 1\right) \\
&= -\frac{s_t s_{f_t(d)} s_{f_t(d')}}{3\left(s_t + s_{f_t(d)}\right)\left(s_t + s_{f_t(d')}\right)}\left(s_{f_t(d)} - s_{f_t(d')}\right).
\end{aligned}$$

Hence $B_t(d) - B_t(d')$ and $s_{f_t(d)} - s_{f_t(d')}$ always have the opposite sign. ■

Proof of Lemma 5. Result 1 is obvious. To see result 2.i, note that when $s_\tau = 0$, Proposition 2.1 implies that PA between τ and $f_\tau(d)$ makes $f_\tau(d)$ strictly worse off, and PA between τ and $f_\tau(d')$ makes $f_\tau(d')$ strictly worse off similarly. For 2.ii, if PA between t and $f_t(d)$ and PA between t and $f_t(d')$ co-exist, by Definition 1, the former requires $B_t(d) \geq B_t(d')$ while the latter requires $B_t(d') \geq B_t(d)$, which delivers $B_t(d) = B_t(d')$, hence $s_{f_t(d)} = s_{f_t(d')}$ by Lemma 4. ■

Proof of Proposition 3. By Lemma 5.2, we only need to consider the PA between WK and WL, or between BL and WL. We show the former; the latter follows from symmetry. According to Definition 1, four conditions are (i) $B_{WK}(R) \geq 0$, (ii) $B_{WL}(R) \geq 0$, (iii) $B_{WK}(R) \geq B_{WK}(C)$, (iv) $B_{WL}(R) \geq B_{WL}(C)$, and at least one of the first two is strict. In Figure 1, (i) and (ii) correspond to the area above Line bb and below Line aa, respectively. (iii) follows from Proposition 2.1. Finally, (iv) lies below Line ee. Region B is thus determined. ■

Proof of Proposition 4. By definition,

$$\begin{aligned}
\frac{\partial \beta_i}{\partial a_{jk}} &= \frac{2!\,(n-3)!}{n!}\frac{s_j s_k}{s_j + s_k + s_i} + \sum_{\substack{T \supsetneq i, T \supsetneq j,k \\ T \neq j \cup k}} \frac{|T|!\,(n - |T| - 1)!}{n!}\left(\frac{s_j s_k}{s_T + s_i} - \frac{s_j s_k}{s_T}\right) \\
&= \frac{2!\,(n-3)!}{n!}\frac{s_j s_k}{s_j + s_k + s_i} - s_j s_k s_i \sum_{\substack{T \supsetneq i, T \supsetneq j,k \\ T \neq j \cup k}} \frac{|T|!\,(n - |T| - 1)!}{n!}\frac{1}{(s_T + s_i)\, s_T}.
\end{aligned}$$

When $n = 3$, the following is straightforward for any distinct i, j, and k:

$$\frac{\partial \beta_i}{\partial a_{jk}} = \frac{2}{6}\frac{s_j s_k}{s_j + s_k + s_i} = \frac{1}{3}\frac{s_j s_k}{N} > 0.$$

When $n > 3$, we have

$$\begin{aligned}
\frac{\partial \beta_i}{\partial a_{jk}} &> \frac{2!\,(n-3)!}{n!}\frac{s_j s_k}{s_j+s_k+s_i} - s_i s_j s_k \sum_{\substack{T\supsetneq i, T\supseteq j,k \\ T\neq j\cup k}} \frac{|T|!\,(n-|T|-1)!}{n!}\frac{1}{(s_j+s_k+s_i)\,(s_j+s_k)} \\
&= \frac{2!\,(n-3)!}{n!}\frac{s_j s_k}{s_j+s_k+s_i} - \left(\frac{1}{3}-\frac{2!\,(n-3)!}{n!}\right)\frac{s_i s_j s_k}{(s_j+s_k+s_i)\,(s_j+s_k)}
\end{aligned}$$

which is positive if and only if

$$\begin{aligned}
\frac{2!\,(n-3)!}{n!} &> \left(\frac{1}{3}-\frac{2!\,(n-3)!}{n!}\right)\frac{s_i}{s_j+s_k} \\
\Leftrightarrow s_j+s_k &> \frac{n\,(n-1)\,(n-2)-6}{6}s_i.
\end{aligned}$$

A sufficient condition for this to hold is

$$s_j+s_k > \frac{n\,(n-1)\,(n-2)-6}{6}\,(N-s_j-s_k) \Leftrightarrow \frac{s_j+s_k}{N} > \frac{n\,(n-1)\,(n-2)-6}{n\,(n-1)\,(n-2)}.$$

This thus is the sufficient condition for $\partial\beta_i/\partial a_{jk} > 0$. Next we establish the counterpart for $\partial\beta_i/\partial a_{jk} < 0$ in a similar way.

$$\begin{aligned}
\frac{\partial \beta_i}{\partial a_{jk}} &< \frac{2!\,(n-3)!}{n!}\frac{s_j s_k}{s_j+s_k+s_i} - s_i s_j s_k \sum_{\substack{T\supsetneq i, T\supseteq j,k \\ T\neq j\cup k}} \frac{|T|!\,(n-|T|-1)!}{n!}\frac{1}{N^2} \\
&= \frac{2!\,(n-3)!}{n!}\frac{s_j s_k}{s_j+s_k+s_i} - s_i s_j s_k \left(\frac{1}{3}-\frac{2!\,(n-3)!}{n!}\right)\frac{1}{N^2},
\end{aligned}$$

which is strictly negative if and only if

$$\begin{aligned}
\frac{2!\,(n-3)!}{n!}\frac{1}{s_j+s_k+s_i} &< \left(\frac{1}{3}-\frac{2!\,(n-3)!}{n!}\right)\frac{s_i}{N^2} \\
\Leftrightarrow \frac{1}{s_j+s_k+s_i} &< \left(\frac{n\,(n-1)\,(n-2)}{6}-1\right)\frac{s_i}{N^2}.
\end{aligned}$$

A sufficient condition for it to hold is

$$\frac{1}{s_i} < \left(\frac{n\,(n-1)\,(n-2)}{6}-1\right)\frac{s_i}{N^2} \Leftrightarrow \frac{s_i}{N} > \sqrt{\frac{6}{n\,(n-1)\,(n-2)-6}}$$

which completes the proof. ■

Proof of Proposition 5. Consider four groups t, $f_t(d)$, $f_t(d')$, and $\tau = f_{f_t(d)}(d')$, with extreme identity between $f_t(d)$ and τ. If $s_{f_t(d')} = 0$, we have

$$\begin{aligned}
\hat{B}_t^I(d) &= \frac{\partial \beta_t}{\partial a_{t,t}} + \frac{\partial \beta_t}{\partial a_{t,f_t(d)}} + \frac{\partial \beta_t}{\partial a_{f_t(d),t}} + \frac{\partial \beta_t}{\partial a_{f_t(d),f_t(d)}} + \frac{\partial \beta_t}{\partial a_{\tau,\tau}}, \\
\hat{B}_t^O(d) &= -\left(\frac{\partial \beta_t}{\partial a_{t,\tau}} + \frac{\partial \beta_t}{\partial a_{\tau,t}} + \frac{\partial \beta_t}{\partial a_{f_t(d),\tau}} + \frac{\partial \beta_t}{\partial a_{\tau,f_t(d)}}\right).
\end{aligned}$$

with the hat representing the case with extreme identity. To establish equality, we verify the following three equations by definitions. Note that, because of extreme identity, they do not immediately follow from Lemma 1.

$$\begin{aligned}
\frac{\partial \beta_t}{\partial a_{t,t}} &= -\frac{\partial \beta_t}{\partial a_{t,f_t(d)}} - \frac{\partial \beta_t}{\partial a_{t,\tau}}, \\
\frac{\partial \beta_t}{\partial a_{f_t(d),f_t(d)}} &= -\frac{\partial \beta_t}{\partial a_{f_t(d),t}} - \frac{\partial \beta_t}{\partial a_{f_t(d),\tau}}, \\
\frac{\partial \beta_t}{\partial a_{\tau,\tau}} &= -\frac{\partial \beta_t}{\partial a_{\tau,t}} - \frac{\partial \beta_t}{\partial a_{\tau,f_t(d)}}.
\end{aligned}$$

In the first equation, each term is the same as with moderate identity. Hence it does follow from Lemma 1. For the second equation, note that $\partial \beta_t / \partial a_{f_t(d),t}$ stays the same as with moderate identity, while both $\partial \beta_t / \partial a_{f_t(d),f_t(d)}$ and $-\partial \beta_t / \partial a_{f_t(d),\tau}$ decreases by $\frac{1}{3}\frac{s_{f_t(d)} s_\tau}{s_{f_t(d)} + s_\tau}$ (from definitions and by Assumption 3a). Hence the equation still holds as with moderate identity. Finally, the third equation holds by symmetry. The inward- and outward-looking changes are thus still equivalent.

Hence, $\hat{B}_t(d)$ decreases by $\frac{2}{3}\frac{s_{f_t(d)} s_\tau}{s_{f_t(d)} + s_\tau}$ compared to $B_t(d)$, and can be written as

$$\hat{B}_t(d) = \frac{s_t}{3(1 - s_t)} \frac{s_{f_t(d)}\left(1 - s_t - s_{f_t(d)}\right)\left(1 + s_t - 2 s_{f_t(d)}\right)}{1 - s_{f_t(d)}} - \frac{2}{3}\frac{s_{f_t(d)} s_\tau}{s_{f_t(d)} + s_\tau}.$$

Substituting $s_\tau = 1 - s_t - s_{f_t(d)}$ into the equation and rearranging terms will yield

$$\hat{B}_t(d) = \frac{s_t}{3(1 - s_t)} \frac{s_{f_t(d)}(1 - s_t - s_{f_t(d)})\left[(\mu - 2) s_{f_t(d)} + s_t - (\mu - 1)\right]}{1 - s_{f_t(d)}}$$

where $\mu = 2/s_t > 2$. Some manipulation then delivers the result. ■