ENDOGENOUS CATEGORIZATION AND NEIGHBORHOOD EFFECTS: EMPIRICAL EVIDENCE FROM THREE BRAZILIAN METROPOLITAN AREAS[1]

Michel ALEXANDRE

PhD candidate in Economics, University of Siena (Italy) & Central Bank of Brazil.

Abstract: This paper raises the hypothesis that neighborhood effects are weaker under endogenous categorization circumstances. If categorization is essentially endogenous, changes in individual behavior can make her depart from the group prescribed behavior. It increases the probability of this individual be no longer seen as a member of the group and, thus, her actions will have low or null influence on other members' behavior. This hypothesis is empirically tested using Brazilian data. It is analyzed the influence of adult behavior on teenager school attainment, considering two potential sources of neighborhood effects: physical (residential) proximity and identity, there is, membership in a common social category. Two social categories are being considered: gender and race. In two of three analyzed Metropolitan Areas (São Paulo, Rio de Janeiro and Salvador), this hypothesis is corroborated, as the neighborhood effects stemming from gender (the exogenous category) showed to be stronger than those stemming from race (the endogenous category). It has important implications for antipoverty policies.

Resumo: Este estudo propõe que a endogeneidade da categorização enfraquece os efeitos de vizinhança. Se a categorização é essencialmente endógena, alterações no comportamento individual podem ser vistas como fuga das prescrições comportamentais do grupo. Isto aumenta a probabilidade de o indivíduo não ser mais categorizado como um membro do grupo e, deste modo, suas ações deixariam de influenciar os demais membros da categoria. Esta hipótese é testada a partir de dados do Brasil. É analisado o impacto do comportamento de adultos nas decisões educacionais de adolescentes, considerando-se duas fontes potenciais de efeitos de vizinhança: proximidade física (residencial) e identidade, isto é, pertencimento à mesma categoria social. Duas categorias sociais estão sendo consideradas: gênero e raça. Em duas das três Regiões Metropolitanas estudadas (São Paulo, Rio de Janeiro e Salvador), esta hipótese é corroborada, uma vez que os efeitos de vizinhança oriundos de gênero (categorização exógena) mostraram-se mais robustos que aqueles oriundos de raça (categorização endógena). Isto traz importantes implicações para políticas anti-pobreza.

Key-words: education, neighborhood effects, social categorization.

Palavras-chave: educação, efeitos de vizinhança, categorização social.

JEL Classification: Z13; I21; I38.

Área ANPEC: Área 11 – Economia Social e Demografia Econômica.

Endogenous categorization and neighborhood effects: empirical evidence from three Brazilian metropolitan areas

1. Introduction

A vigorous literature is dedicated to the empirical study of neighborhood effects. These effects are shown to be relevant in the explanation of a wide set of individual behaviors, ranging from crime to school attainment.

There is not a unique, well defined concept of neighborhood. Some studies consider as neighborhood a set of geographically close individuals. On the other hand, other studies incorporate insights from other social sciences, as psychology and sociology, and assume identity as an important component of neighborhood formation. Thus, neighbors would be those who identify with the same group or culture, as an ethnic group.

In this sociological concept of neighborhood, a crucial point is how individuals place themselves and others in social categories. Some categorizations are essentially exogenous. The more obvious example of this kind of categorization is gender grouping. Individuals follow very unambiguous and easily observable exogenous characteristics to map themselves and others as men or women. In this case, neighborhood effects tend to be very strong, since a woman, for example, will be always correctly identified as so and will be followed by other women in her actions.

Nevertheless, some categorizations are essentially endogenous. This is the case of ethnic/racial categorization. How an individual is racially placed depends on her behavior. Groups have behavioral prescriptions associated with it. In highly endogenous categorizations, the more an individual is close to the behavioral prescription of some group, the more likely she will be classified by others as a member of this group. Thus, there is a situation of multiple determination: the behavior of an individual influences the actions that will be chosen by the members of her reference group, but also influence if she will be considered or not as a member of this group. Therefore, in the case of endogenous categorizations, neighborhood effects tend to be weaker, given the possibility that an individual be "expelled" from some reference group according to her attitude.

In this paper I test empirically this hypothesis, comparing the neighborhood effects stemming from an exogenous categorization (gender) with neighborhood effects stemming from an endogenous categorization (race). The utilized data are the 2000 Census from three Brazilian Metropolitan Areas: São Paulo, Rio de Janeiro and Salvador. The results corroborate this hypothesis, as the estimated neighborhood effects originated from gender are much stronger than those originated from race.

Beside this introduction, this paper has more four parts. Section 2 discusses the theoretical framework relative to neighborhood effects and the endogeneity of some categorizations, notably the racial categorization. Section 3 sets the model. The data and the regression results are presented in Section 4. Some conclusions take the last part.

2. Theoretical framework

Neighborhood effects

Neighborhood effects arise when individual decisions are influenced by the actions taken by other members of some reference group. In general they are justified along psychological and sociological lines (Durlauf, 2004)[2]. Using criminal behavior as reference, Glaeser et al (2001) cite some channels thorough which neighborhood effects can take place, "ranging from pure physical externalities (while one person is being arrested, the police find it harder to arrest someone else), to learning from one's neighbors, to stigma (the more people who are committing a particular crime--the less likely is that crime to be a negative signal) to pure taste externalities (individuals just enjoy imitating others)" (Glaeser et al, 2001: 1).

An extensive literature is dedicated to the empirical analyses of neighborhood effects on a wide range of actions, as crime (Glaeser et al, 1996), student outcomes (Sacerdote, 2001), school attainment (Datcher, 1982) and teenage pregnancy (Evans et al, 1992). The methodologies applied to the measurement of these effects, especially those relying on micro data, face three main problems: the reflection problem (individual is influenced by others, but also influence them), the presence of an exogenous characteristic of some area leading all the habitants of this area to behave in a similar way (for example, the bad quality of local schools can lead almost all local teenagers to drop out school) and the endogeneity of interactions (individuals with the same preferences prefer to interact among them) [3].

Some studies present empirical evidence on the importance of neighborhood effects. Case & Katz (1991), analyzing young men in poor Boston neighborhoods, found significant peer effects on criminal behavior, drug and alcohol use, church attendance and labor market activity. Crane (1991) concludes that the percentage of workers with professional or managerial job in PUMS neighborhoods affects 16-19 years old women behavior regarding dropping out of high school and fertility. However, the importance of neighborhood effects is not consensual among empirical studies. Evans et al (1992), for instance, studying teenager behavior regarding dropping out of high school and pregnancy, found no evidence of neighborhood effects once endogeneity of neighborhood is controlled for by instrumental variables.

Identity and neighborhood effects

The basic idea of this approach is that one's actions are influenced by the behavior of one's neighbors, but there is not a unique, well defined concept of neighborhood. In most of literature, a neighborhood is formed by those who share some physical space, as a residential area or a classroom. In Evans et al (1992), for example, neighbors are the students of some school.

Nevertheless, a widespread literature, bringing insights from other social sciences, notably social psychology and sociology, suggests that social influence goes beyond geographic proximity. Rather, it has to do with how individuals are located in what Akerlof (1997) calls "social space". This perception of neighborhood is supported by a sociological view of social interactions, based on "concepts that play little or no role in modern economics: class, community, culture, influence, status, gender roles, and so on" (Manski,

---

[2] An important exception concerns local public finance of education. In this case, children of poor neighborhoods are affected by the low level of local per pupil expenditures. To more on this, see Hussar & Sonnenberg (2001).

[3] To more on this, see Manski (1993).

2000: 12). That is, neighborhood effects arise between individuals linked by some kind of social glue – a common sense of *identity*.

The study of Akerlof & Kranton (2000) brings an excellent survey of this literature. Individuals belong to one or more *social category*. A universally familiar example of social categorization concerns gender. Within this kind of categorization, there are two social categories, man and woman. To each social category, it is associated a set of *behavioral prescriptions*, which can be interpreted as the ideal behavior for each category. Following in this example, there are behavioral prescriptions for gender categories regarding dressing (e.g., just women are supposed to wear dresses), jobs (e.g., firefighters are men, while nurses are women), family roles (the father is responsible for providing the financial resources, while the mother should stay at home taking care of the children), among other aspects.

The process of commitment with the behavioral prescriptions of some social category is called in psychology *internalization of rules of behavior*: "Identification is a critical part of this internalization process: a person learns a set of values (prescriptions) such that her actions should conform with the behavior of some people and contrast with that of others" (Akerlof & Kranton, 2000: 728). The violation of these behavioral prescriptions is costly both to oneself – in this case, the person feels what psychologists calls *anxiety* – and to others. For example, men in predominantly male jobs can be hostile to female presence because it is a threat to their status as members of an occupational group (Goldin, 2002).

In this approach, neighborhood effects are closely related to the endogeneity of behavioral prescriptions. An example concerns the feminist movement and consequent change in gender occupational segregation. Popularly, the feminist movement was launched in U. S. by the book *The Feminine Mystique*, written by Betty Friedan in 1963. After that, there was a great incursion of females in predominantly male occupations. In U. S., between 1970 and 1990, there was an increase of the percent female in virtually all 45 Census occupations, except in some traditionally female jobs (as dietitians and speech therapists). Moreover, some former predominantly male occupations turned to be predominantly female, as underwriters and psychologists (Blau et al, 1998). It represents a change in female behavioral prescriptions concerning family roles (weakening of the housewife stereotype), as well as those related to occupations.

There are few empirical studies devoted to analyze the impact of identity in neighborhood effects. For instance, Borjas (1992) stress the importance of the so-called "ethnic capital", defined as the average skills of the ethnic group in the parents' generation, on children skills. It was shown that ethnic capital matters even among individuals who grow up in the same neighborhood (Borjas, 1995), suggesting that ethnic identity is an important channel of neighborhood effect. Another example is the study of Aizer & Currie (2004), which found evidence that the use of prenatal care public programs is highly correlated within groups defined using race/ethnicity and neighborhoods.

Endogenous categorization

In this sociological concept of neighborhood, a crucial point is to which social category(s) the individual belongs to. What individuals will be considered as members of one's reference group and whose actions will, thus, influence one's choices? Regarding this point, there are two dimensions of categorization which are very important: the way individuals group themselves and how individuals place others – self-categorization and heterocategorization, respectively.

Some categorizations are essentially exogenous. An example of this first type of categorization is gender categorization. In this case, individuals generally group themselves and others in one of the two categories – males and females – in an unambiguous way driven by easily observable exogenous characteristics.

Notwithstanding, other types of categorization are at least partially endogenous. Ethnic/racial categorizations are good examples of this second kind of grouping. In this case, categorization is driven not just by exogenous individual characteristics – as phenotype – but also by endogenous elements regarding the individual (e.g. social status) and the environment (e.g. social and historical context).

The literature brings abundant evidence that self-categorization may be essentially endogenous. For example, Miguel & Posner (2006), analyzing data from twelve African countries, concluded that ethnic identification is positively related to employment in non-traditional economic sectors and to the proximity of the survey to a competitive national election. Working with U.K. data, Bisin et al (2006) found that the main determinants of ethnic identity include past racial harassment experiences, language spoken at home and with friends, quality of housing, structure of the family and degree of neighborhood segregation.

Also heterocategorization is influenced by endogenous elements. Bowen (1996), for instance, relates that, in Rwanda, those with lots of cattle were classified as a Tutsi, and the poorer ones, as Hutu. Even a country as United States, in which there already was a legal definition of race – and, thus, supposed to have a more exogenous racial categorization – presents a significant degree of racial endogeneity[4]. In the U. S. prior to the Civil War, marked by racist rules, African descendents, especially mulattoes, could be "white" by behavior and reputation (Bodenhorn & Ruebeck, 2003).

This categorization endogeneity can undermine, or even completely neutralize, neighborhood effects. Suppose some group has a behavioral prescription regarding some action (for example, share of time devoted to work). If categorization is essentially endogenous, membership is considered on the basis of commitment with the prescribed behavior, rather than on some exogenous attribute (as skin color or gender). It means that, the closer an individual action is from the prescribed behavior, the more likely she will be considered a member of this group. In this circumstance, if an individual change their choice, in such a way that it departs from the prescribed behavior, she may be no more seen as a member of the group. Thus, her change will have a meager, or even null, impact on other members' choices. In other words, there is a *trade-off between endogeneity of categorization and endogeneity of behavioral prescriptions*.

It has strong implications regarding the persistence of group inequality. Sociological theories of poverty, as the social dominance theory (Pratto et al, 2006), points that dominated groups follow as behavioral prescription low levels of mobility-enhancing behavior, as schooling. For example, some studies point that African-American youth, recognizing societal iniquity in confront with their group, may come to feel education as of little usefulness for their economic and social mobility. Thus, adolescents identified with the Black culture may present targets of low academic achievement (Chavous et al, 2003). In contrast, dominant groups are attached to pro-mobility strategies. Thus, the erosion of group differences is jeopardized in contexts of high endogenous categorization.

The following model formalizes this idea.

---

[4] In this country, "the records of governmental specification of rules of racial identity associated with peoples of African descent were established both through legislation and court decisions (Darity Jr. et al, 2006: 288)".

3. The model

Consider a population composed by a large number of individuals equal to n. They maximize their respective utility function by choosing some variable X (e.g. years of schooling). The expected value of X is $\overline{X}$. There are S possible social categories in this society to which individuals can belong to. The choices made by the other members of one's social categories affect one's utility, in the sense that individuals are better-off if they choose a level of X close to that of their neighbors, understood here as the members of the same social category.

Suppose an individual a, member of this population, belongs to J social categories. Her utility can be represented by the following equation:

$$U(X_a) = f(X_a) - \sum_{j=1}^{J} \frac{1}{2} \beta_j (X_a - \hat{X}_j)^2$$

In the equation above, f(X) represents some direct utility generated by X, being f_X non-negative, β_j is a non-negative parameter representing the weight attributed to the conformity to group j and $\hat{X}_j$ is the average X observed among other members of group j. Maximizing,

$$X_a = \frac{f_X + \sum_{j=1}^{J} \beta_j \hat{X}_j}{\sum_{j=1}^{J} \beta_j}$$

Thus, individual a will choose a level of X driven by the direct utility generated by X and the observed choices of X made by the members of the categories she belongs to. If she perceives an increase in the average X chosen by some of the categories she identifies with, she will augment her chosen X too, and similarly in case of a decrease in the observed X.

The question is how a choice is affected by the choice made by other individuals. The key point is that heterocategorization is endogenous, in the sense that the probability of an individual be seen by a as a member of some group j is influenced by her choice of X. Specifically,

$$P_i^j = P(\alpha^j, X_i, X^j, \gamma_i)$$

In the equation above, α^j is the endogeneity of category j, there is, how much heterocategorization in group j depends on the endogenous element (the choice of X). X^j is

6

the prescribed X to group j, linked to the concept of behavioral prescription discussed above, and γ_i is a set of exogenous individual characteristics, as gender or skin color.

The way X_i affects P_i^j depends on X^j. What matters is the distance between these values, that is, the more the X chosen by the individual is far from the prescribed X to j, the smaller is her probability of be classified as member of j. The intensity of this change is positively correlated with α^j: for higher levels of endogeneity, deviations from (approximations to) X^j will cause higher decrease (increase) in P_i^j. Thus, the derivative of P_i^j with relation to the distance between X_b and X^j is negative and, as this distance is always non-negative, the derivative of P_i^j with relation to α^j is non-positive. Finally, how γ affects P_i^j depends on the category and the exogenous characteristic in question. For example, a dark skin color increases one's chance of be classified as Black.

Since P_i^j and the real X chosen by each individual, it is possible to calculate the expected $\hat{X}_j$, that will be equal to

$$E(\hat{X}_j) = \frac{\sum_{i=1}^{n} X_i P_i^j}{\sum_{i=1}^{n} P_i^j}$$

The question now is to see how $E(\hat{X}_j)$ will vary in response to a change in the X chosen by some individual k. The derivative of $E(\hat{X}_j)$ with respect to X_k is

$$\frac{\partial E(\hat{X}_j)}{\partial X_k} = \frac{(X_k P_X^j + P_k^j)\sum_{i=1}^{n} P_i^j - P_X^j \sum_{i=1}^{n} X_i P_i^j}{(\sum_{i=1}^{n} P_i^j)^2} = \frac{P_X^j(X_k \sum_{i=1}^{n} P_i^j - \sum_{i=1}^{n} X_i P_i^j) + P_k^j \sum_{i=1}^{n} P_i^j}{(\sum_{i=1}^{n} P_i^j)^2},$$

where P_X^j is the derivative of P_k^j with respect to X_k. The second term of the numerator and the denominator of the equation above are clearly non-negative. However, it is necessary more information to determine the sign of the first term of the numerator. Suppose, for example, that X_k is sufficiently low. In this case, the expression between brackets in the first term of the numerator would be negative. On the other hand, in this case P_X^j would probably be positive[5] and the product between the two terms would be negative. An opposite situation is possible for a sufficiently large value of X_k.

---

[5] To see this, suppose that X_k is below the prescribed X. Thus, an increase in X_k would reduce the absolute difference between this value and the prescribed X. Adopting an X closer to X prescribed to group j, the individual k would be considered as a member of this group with a greater probability.

Thus, the overall effect of a variation in the chosen X of a potential member of j in the observed X of group j is impossible to determine given the available information. However, we know that $E(X) = \overline{X}$. Making $X_k = X_i = \overline{X}$ the term between brackets goes to zero and

$$E(\frac{E(\partial \hat{X}_j)}{\partial X_k}) = \frac{P_k^j}{\sum_{i=1}^{n} P_i^j}$$

Thus, the expected variation is positive. Note that the expression above do not reduce to $1/n$ because P_k^j depends also on individual exogenous characteristics. Therefore, even if all individuals choose the same level of X, they would present different probabilities of be classified as members of j.

This influence is decreasing in the degree of categorization endogeneity. It can be seen by taking the derivative with respect to α of the equation above, which is equal to

$$E(\frac{\partial^2 E(\hat{X}_j)}{\partial X_k \partial \alpha}) = \frac{P_\alpha^j (\sum_{i=1}^{n} P_i^j - P_k^j)}{(\sum_{i=1}^{n} P_i^j)^2},$$

where P_α^j is the derivative of P^j with respect to α, which is negative. The term between brackets is clearly positive; hence the expression above is negative. Thus, the more endogenous is some categorization, the weaker is the expected influence of someone belonging to such category.

4. Data and estimation

It was used data from the 2000 Census relative to the Metropolitan Areas of São Paulo, Rio de Janeiro and Salvador (Brazil). I am interested in analyzing neighborhood effects on teenagers' school attainment. The dependent variable is effective years of schooling of 10-17 years old adolescents. Beside their gender, it was taken into consideration the teenagers' self-declared race: White, Black, *Pardo*[6], *Yellow*[7] and Indian. Individuals with ignored race were excluded from the sample. The general characteristics of the samples in the three Metropolitan Regions are presented in Table 1. Other independent variables include dwelling's per capita income and years of effective schooling of the dwelling's responsible, as shown in Table 2.

The crucial variables are GEN and RAC, which are suppose to capture the neighborhood effects. Doing so, I am close to the sense of neighborhood present in the role

---

[6] *Pardo*, in general, is a category chosen by mixed-race individuals.
[7] *Yellow*, in Brazil, refers to Asian (mainly Japanese) descendents.

models (Streufert, 2000), according to which adult behavior in some community affect teenagers' choices. Beside residential proximity, the concept of neighborhood used here is constructed also along identity lines, considering two types of categories: gender and race. Note that the variable RAC is very similar to the concept of "ethnic capital" defined by Borjas (1992, 1995).

Brazil is a suitable society to make the comparison I am interested in. Racial heterocategorization in this country is essentially endogenous and influenced by school attainment. Researches driven by the sociologist Edward Telles (Telles, 2002; Telles & Lim, 1998), since a nationwide survey conducted in 1995 by one of the main Brazilian newspapers, show that self-classified Blacks are more likely to be heteroclassified as Whites the more educated they are. Analyzing a sample of 243 students in a public school in São Paulo City, Carvalho (2005) concludes that teachers tend to "whiten" children with better grades.

The estimation results are presented in Table 3. It can be seen that in São Paulo and Salvador the gender neighborhood effect coefficients are slightly higher than those related to race. By contrast, in Rio de Janeiro, gender effects are negative and race effects are positive. Thus, in two of three Metropolitan Areas, the basic hypothesis presented in this study is empirically corroborated, that is, neighborhood effects stemming from the more exogenous social category (gender) appear to be stronger than those stemming from the more endogenous social category (race).

Racial neighborhood effects are stronger in São Paulo than in the other Metropolitan Areas. It can be explained by the smaller fraction of self-classified pardos in this region, how can be seen in Table 1. Corroborating this view, the Metropolitan Area with the weakest racial neighborhood effect (Salvador) has also the higher fraction of self-classified pardos (60%). Under miscegenation, it is more difficult to sort individuals through their phenotypic characteristics. But individuals feel an intrinsic necessity of group each other. When this is not possible through phenotypic characteristics, categorization is rooted in other elements as, for example, behavioral checks (Humphreys et al, 2002). Thus, racial endogeneity tends to be higher in more mixed areas.

Another explanation to this fact can be the higher presence of self-classified Whites in São Paulo. As argued earlier, dominant groups tend to have among their behavioral prescriptions high levels of mobility-enhancing attitudes. In this case, the probability that the level of schooling chosen by some individual be below the prescribed schooling attainment is higher. By increasing her schooling attainment, this individual will get closer the prescribed behavior and will be seen, with a higher probability, as a member of that group. Thus, she will be imitated by other individuals that identify with this category with a greater chance.

5. Concluding remarks

This paper suggests that there is a trade-off between endogeneity of categorization and endogeneity of behavioral prescriptions. In highly endogenous categorizations, members departing from the group's behavioral prescriptions will be no longer seen as members of the group, and their idiosyncratic behavior will be not followed by other members of the category. Thus, neighborhood effects stemming from endogenous categories tend to be weaker than those stemming from exogenous ones.

This hypothesis is empirically tested. The used data is the 2000 Census from three Brazilian Metropolitan Areas: São Paulo, Rio de Janeiro and Salvador. It was tested if

teenager school attainment is influenced by the behavior of adults living in the same sub district, considering two sources of identity: gender and race. In two Metropolitan Areas (São Paulo and Salvador), this basic hypothesis is corroborated: individuals are affected by those belonging to the same category, but this influence is stronger when the more exogenous category (gender) is considered rather than the more endogenous one (race). Moreover, racial neighborhood effects seem to be stronger in São Paulo than in the other two Metropolitan Areas, what can be explained by the smaller miscegenation (what implies smaller racial endogeneity) and the higher presence of Whites (attached to higher educational prescriptions) in this region.

These results bring additional implications for public policy. In the presence of neighborhood effects, antipoverty policies would be more efficient, for example, if resources were concentrated in more interacting individuals (Durlauf, 2000). Nonetheless, as already heralded by the current literature, neighborhood effects go beyond physical proximity and depends on identity ties. The novelty of this paper consists in suggesting that neighborhood effects are stronger when such ties are rooted in exogenous elements. If the insights presented in this study are correct, it is true, for example, that affirmative action policies are more efficient in more exogenous racial categorization contexts.

References

Aizer, A. & J. Currie (2004). "Networks or Neighborhood? Correlations in the Use of Publicly-Funded Maternity Care in California". *Journal of Public Economics*, 88 (12): 2573-2585.

Akerlof, G. A. (1997). "Social distance and Economic Decisions". *Econometrica*, 65 (5): 1005-1027.

Akerlof, G. A. & R. E. Kranton (2000). "Economics and Identity." *The Quarterly Journal of Economics*, 115 (3): 715-753.

Bisin, A., E. Patacchini, T. Verdier & Y. Zenou (2006). "'Bend it like Beckham:' Identity, Socialization, and Assimilation". *CEPR Discussion Paper* 5662.

Blau, F. D., P. Simpson & D. Anderson (1998). "Continuing Progress? Trends in Occupational Segregation in the United States over the 1970's and 1980's". *NBER Working Paper no. 6716*.

Bodenhorn, H. & C. S. Ruebeck (2003). "The Economics of Identity and the Endogeneity of Race". Unpublished manuscript.

Borjas, G. J. (1992). "Ethnic Capital and Intergenerational Mobility". *The Quarterly Journal of Economics*, 107: 123-150.

_____. (1995). ''Ethnicity, Neighborhoods, and Human Capital Externality." *American Economic Review*, 85 (3): 365–390.

Bowen, J. R. (1996). "The Myth of Global Ethnic Conflict." *Journal of Democracy*, 7: 3-14.

Carvalho, M. (2005). "Quem é Negro, Quem é Branco: Desempenho Escolar e Classificação Escolar de Alunos". *Revista Brasileira de Educação*, 28: 77-95.

Case, A & L Katz (1991). "The company you keep: the effects of family and neighborhood on disadvantaged families". *NBER Working Paper no 3705*.

Chavous, T. M., D. H. Bernat, K. Schmeelk-Cone, C. H. Caldwell, L. Kohn-Wood & M. A. Zimmerman (2003). "Racial Identity and Academic Attainment among African American Adolescents". *Child Development*, 74: 1076-1090.

Crane, J (1991). "The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing", *American Journal of Sociology*, 96:1226-1259.

Darity Jr., W. A., P. L. Mason & J. B. Stewart (2006). "The Economics of Identity: The Origin and Persistence of Racial Identity Norms." *Journal of Economic Behavior & Organization*, 60: 283-305.

Datcher, L., (1982). "Effects of Community and Family Background on Achievement". *Review of Economics and Statistics*, 64 (1): 32-41.

Durlauf, S. N. (2000). "The Membership Theory of Poverty: The Role of Group Affiliations in Determining Socio-Economic Outcome" Mimeo, University of Wisconsin-Madison.

_____ (2004). "Neighborhood Effects". In J. V. Henderson & J. F. Thisse (eds.), *Handbook of Regional and Urban Economics*. Amsterdam: Elsevier/North Holland.

Evans, W. N., W. E. Oates & R. M. Schwab (1992). "Measuring peer group effects: a study of teenage behavior". *Journal of Political Economy*, 100 (5): 966-991.

Glaeser, E. L., B. Sacerdote & J. A. Scheinkman (1996). "Crime and social interactions". *The Quarterly Journal of Economics*, 111 (2): 507-548.

Glaeser, E. L., & J. A. Scheinkman (2001). "Measuring social interactions". In S. N. Durlauf & P. Young (eds.), *Social dynamics* (pp. 83–131). Cambridge, MA: MIT Press.

Goldin, C. (2002). "A pollution theory of discrimination: male and female differences in occupations and earnings". *NBER Working Paper no 8985*.

Humphreys, M., D. N. Posner & J. M. Weinstein (2002). "Ethnic Identity, Collective Action and Conflict: An Experimental Approach". *APSA Meeting*, Boston, September.

Hussar, W. and W. Sonnenberg, (2001). "Trends in Disparities in School District Level Expenditures per Pupil". Mimeo, National Center on Education Statistics.

Manski, C. F. (1993). "Identification of Endogenous Social Effects: The Reflection Problem". *Review of Economic Studies*, 60: 531-542.

_____. (2000). "Economic Analysis of Social Interactions". *NBER Working Paper no 7580*.

Miguel, E. & D. N. Posner (2006). "Sources of Ethnic Identification in Africa". Mimeo, U. C. Berkeley.

Pratto, F., J. Sidanius & S. Levin (2006). "Social Dominance Theory and the Dynamics of Intergroup Relations: Taking Stock and Looking Forward". *European Review of Social Psychology*, 17: 271-320.

Streufert, P. (2000). "The Effect of Underclass Isolation on School Choice," *Journal of Public Economic Theory*, 2 (4): 461-482.

Telles, E. E. (2002). "Racial ambiguity among the Brazilian population." *Ethnic and Racial Studies*, 25 (3): 415-441.

Telles, E. E. & N. Lim (1998). "Does it Matter Who Answers the Race Question? Racial Classification and Income Inequality in Brazil." *Demography*, 35 (4): 465-474.

Table 1: Basic sample characteristics

	São Paulo	Rio de Janeiro	Salvador
# of observations	266,474	149,043	49,229
# of sub districts	157	89	26
Gender:			
% Male	50.19	50.47	49.74
% Female	49.81	49.53	50.26
Race:			
% White	62.71	48.67	18.53
% Black	4.58	10.46	20.20
% Pardo	31.38	40.44	60.18
% Yellow	1.17	0.19	0.32
% Indian	0.16	0.23	0.77

Table 2: Independent variables

Variable symbol	Variable description
AGE	Age in years
FEM	1 if female
BLA	1 if self-identified Black
PAR	1 if self-identified Pardo
YEL	1 if self-identified "Yellow"[1]
IND	1 if self-identified Indigenous
RES	Effective years of schooling of the responsible for the dwelling
PCI	Per capita income in the dwelling
SON	1 if living with at least one of the parents

	in the household[2]
GEN	Average effective years of schooling of 25-50 adults of the same gender and living in the same sub district
RAC	Average effective years of schooling of 25-50 adults of the same race and living in the same sub district
SUD	Set of dummy variables relative to sub districts

(1): Yellow, in Brazil, refers to the Asian (mainly Japanese) descendents.

(2): In Brazilian Census definitions, a dwelling can be composed by more than one household.

Table 3: Regression results

Variável	São Paulo	Rio de Janeiro	Salvador
AGE	0.767	0.744	0.645
FEM	0.274	0.382	0.413
BLA	-0.278	-0.578	-0.419
PAR	-0.174	-0.297	-0.245
YEL	0.111	-0.100*	-0.304
IND	-0.278	-0.372	-0.314
RES	0.068	0.113	0.130
PCI	0.017	0.011	0.050
SON	0.602	0.374	0.496
GEN	0.096	-0.144	0.421
RAC	0.050	0.017	0.003*

(*): Not sgnificant at 5%.