ANNALS OF ECONOMICS AND FINANCE **14-1**, 111–128 (2013)

Categorical Segregation from a Game Theoretical Approach*

Antoni Rubí-Barceló

Universitat de les Illes Balears, Edifici Jovellanos, Cra. Valldemossa km 7.5; 07122 Palma de Mallorca (Spain)
E-mail: antoni.rubi@uib.eu

This paper exploits a coalition formation game with incomplete information to illustrate the causal relationship between categorical thinking and segregation. This causality was suggested by Fryer and Jackson (2008). The present model shows how societies can be segregated even when its self-interested members have no a priori motivation to discriminate by social identity; consequently, this paper supports the argument that segregation may not be malevolent in origin.

Key Words: Categorization; Segregation; Incomplete information; Cooperative games.
JEL Classification Numbers: J15, J71, C71.

1. INTRODUCTION

Segregation is a pervasive and persistent sociological phenomenon that emerges along many different lines (race, sex, religion, or language among others) and appears in a wide range of situations (residential or occupational segregation, for example). The set of explanatory factors possibly influencing segregation that have been explored by the literature is large. In a labor market environment, the main theories of segregation are based upon taste (Becker, 1971) or statistical discrimination by employers (Phelps, 1972). Papers about residential segregation have pointed out many differ-

* I am grateful to Antonio Cabrales, Antoni Calvó-Armengol and Joan Esteban for helpful comments and suggestions. Thanks also to Daniel Cardona, Aniol Llorente, Rahmi Ilkiliç, Coralio Ballester, Frédéric Deroïan, and Joan de Martí for helpful conversations. This paper have also benefited from the comments of two anonymous referees. Financial support from Fundación Ramón Areces, Universitat Pompeu Fabra and Ministerio de Clencia y Tecnología through grant ECO2011-23934 is gratefully acknowledged. A version of this work has been previously published as working paper 547 of the Collection of Working Papers of the *Fundación de las Cajas de Ahorros (FUNCAS)*. All errors are of course mine.

1529-7373/2013

ent factors but individual's preferences have been single out as a critical variable both by economists and sociologists (Clark, 1991). Schelling's neighborhood segregation model is the most widely cited contribution in this literature.

In a series of pioneering works, Schelling (1969, 1971a,b, 1972, 1978) introduced a self-organization model system of two distinguishable types of agent with discriminatory individual preferences for certain neighborhood compositions and then explored the dynamics of this model system. The results show that "micromotives at the local level give rise to macrobehavior at the aggregate (global) level" (p.2 in Pancs and Vriend, 2007) and "even quite color-blind individual preferences produce quite segregated neighborhoods" (p.3 in Epstein and Axtell, 1996).

Nevertheless, most studies investigating the causes of segregation agree that "classical" theories such as taste or statistical discrimination cannot alone explain this sociological condition. The present paper formally analyzes an alternative and complementary cause of segregation based on informational asymmetries across the different social groups. In this model, these informational asymmetries are grounded by *categorization*, a cognitive process largely analyzed in social psychology. To the best of my knowledge this paper constitutes the first formal attempt to motivate this type of segregation from a game theoretical approach.

The central idea of categorization is that human mind stores past experiences in a finite set of "folders" or categories and that the number of categories is limited. In social psychology a large list of contributions demonstrate that agents process information with the aid of categories (see Fryer and Jackson (2008) for a review). An extensive list of authors in this field treats some biases such as stereotyping or prejudice as inevitable consequences of categorization (see Allport (1954), Hamilton (1981), Tajfel (1969), Fiske (1998), Markman and Gentner (2001), and Macrae and Bodenhausen (2000)). Fryer and Jackson (2008) presents a formal model justifying that:

"[...] types of experiences and objects that are less frequent in the population tend to be more coarsely categorized and lumped together. As a result, decision makers make less accurate predictions when confronted with such objects".

This lower accuracy with respect to less "frequent" agents is what may explain prejudices or discrimination against minority members. Evidence for a coarser sorting of blacks by employers can be found in Jowell and Prescott-Clarke (1970), Hubbick and Carter (1980), Brown and Gay (1985), and Bertrand and Mullainathan (2003)[1].

[1]There is also a literature on racial and ethnic differences in facial recognition (see Sporer (2001) for a detailed review) showing that individuals who interact more frequently

In this model agents are featured by a physical characteristic (in general, any observable feature such as race, sex, or language) which is assumed to be payoff irrelevant and by a "qualitative" aspect (productivity, talent or human quality among other interpretations) that might be ignored by others and whose value affects others' payoff. Based on categorization evidence, it is assumed that agents' pattern of social interactions affects the accuracy of their predictions about others' qualitative feature. Specifically, this accuracy is lower when predicting the qualitative feature of those agents whose physical characteristic is less frequent in the set of agents who socially interact with the predictor. In this paper, these social interactions are formalized through coalitions or communities. Thus, the information a player has about others' qualitative aspect would depend on the social structure of her own community.

In many social, economic, and political problems individuals organize themselves in communities. Examples include social clubs, firms, teams, or faculties. In this paper, each player contributes to the production of a local public good according to her qualitative feature, say productivity. Then, each agent consumes the public good produced by the community to which she belongs. As a leading interpretation of this model, one can think on communities as social groups or clubs that carry out some lobbying activity or produce some club good (*i.e.* a local public good). What is feasible to produce for a community depends on the aggregate productivity of its members. Once a community is constituted, players can determine the aggregate public good production. Thus, we are talking about a cooperative game.

In this case, players' preferences over alternative organizations are hedonic,[2] *i.e.* players' utility depends only on the composition of their own community. The stability of coalition partitions where players have hedonic preferences has been analyzed in a large number of models with local public goods, as in Guesnerie and Oddou (1981), Greenberg and Weber (1986) and Demange (1994), or with some sort of political interaction, as in Greenberg and Weber (1993) and Banerjee et al. (2000)[3]. Nevertheless, incomplete information has been rarely treated in coalition formation games. This is true despite the empirical work which shows that the information of individuals in a social structure is limited (see Laumann (1969), Friedkin (1983), Kumbasar, Romney, and Batchelder (1994), Bondonio (1998), and Casciaro (1998)).

---

with members of a given racial group recognize members of this group better than members of other ethnic groups. This data is consistent with the model of categorization of Fryer and Jackson (2008).

[2]This terminology follows Drèze and Greenberg (1980).

[3]Purely hedonic games are studied in Bogomolnaia and Jackson (2002).

In this paper, communities are interpreted as a compendium of players and social connections where all members of a community are (directly or indirectly) connected to each other and disconnected to all the rest. Under this interpretation, two members of different communities can merge their groups by creating a social link between them. In the present paper, only pairwise movements are allowed, *i.e* at most two individuals consider changing the community structure at a time. Such stability tests make sense if players are small relative to the size of communities, or if the cost of coordinating movements to form new communities is high. Following this motivation, it is also reasonable to assume that any individual considering forming a social connection with another agent needs her acceptance. For these reasons, the equilibrium concept used in this paper is the Conjectural Pairwise Nash Equilibrium (CPNE). This concept generalizes the Pairwise Nash Equilibrium (PNE) concept[4] by relaxing the restriction that individual's beliefs must be correct. Specifically, CPNE concept allows players to have incorrect conjectures in equilibrium, so long as they have no information to contradict those conjectures. Additionally, this paper investigates the effect on stability of a refinement that has been considered in the literature: common knowledge of rationality[5].

Section 2 presents the basic setting and formally defines the equilibrium concepts. Following two different frameworks, Section 3 presents the results showing that (i) categorical thinking may cause segregation and (ii) common knowledge of rationality reduces but does not eliminate the possibilities of segregation. This paper introduces and applies a new stability concept, CPNE, to the study of coalition stability with incomplete information. Nevertheless, its main contribution relates to the implications of categorical thinking on segregation; unlike taste discrimination models, this paper shows that segregation may emerge even when agents' preferences are completely color-blind. This supports the argument that segregation may not be malevolent in origin.

2. MODEL

2.1. Set up

Consider a set of players $N = \{1, \dots, n\}$. Let graph g be a collection of direct links that represent pairwise and non-directed relations between the respective two agents. The subset of N containing i and j is denoted by ij

[4]A partition is PNE if no player has incentives to unilaterally deviate and no mutually beneficial link is left aside. See Goyal and Joshi (2006), Calvó-Armengol (2004), and Bloch and Jackson (2006, 2007) for definitions and applications of the PNE concept to network formation games.

[5]This refinement was first considered by Rubinstein and Wolinsky (1994) and Gilli (1999) with respect to the Conjectural Equilibrium concept.

and is referred to as the link ij. Players i and j are directly connected if and only if $ij \in g$, and indirectly connected if and only if there is a *path* in g connecting i and j, where a path is a set of distinct players $\{i_1, i_2, \ldots, i_m\} \subset N$ such that $\{i_1 i_2, i_2 i_3, \ldots, i_{m-1} i_m\} \subset g$. A *coalition* or *community* is a set of players directly or indirectly connected to each other and it is denoted by $S_k \subset N$. A *community partition* is a set $\pi = \{S_k\}_{k=1}^{K}$ that partitions N and Π is the set of all possible partitions. Thus, communities are disjoint and $\bigcup_{k=1}^{K} S_k = N$. The size of community S_k is denoted by s_k. Let N_i denote the set of all possible communities containing player i.

Players self-organize forming their social connections and, consequently, constituting communities for production purposes. Each player, endowed with an inherent productivity, contributes to the production of a local public good according to her productivity and then consumes this public good. Let t_i denote the productivity of agent i which can be high or low, *i.e.* $t_i \in \{H, L\}$. Players' preferences over alternative community partitions are entirely determined by the community they belong to, so this model considers a purely hedonic setting. In particular, preferences are lexicographic and they are affected by the size of the community, s_i, and by its proportion of highly productive members, h_{S_i}. If h_{S_i} exceeds a certain threshold, $\alpha_{t_j} \in [0, 1]$, player j considers that S_i is a good community; otherwise, S_i is a bad community for j. Good communities are preferred to bad ones. Moreover, bigger communities are preferred among the good ones whereas size is a negative factor among bad communities. Formally, for any two communities $S_j, S_k \in N_i$, $S_j \succ_i S_k$ if and only if (i) $s_j > s_k$ when $h_{S_j}, h_{S_k} \geq \alpha_{t_i}$ or (ii) $h_{S_j} \geq \alpha_{t_i} > h_{S_k}$ or (iii) $s_j < s_k$ when $h_{S_j}, h_{S_k} < \alpha_{t_i}$. Finally, $S_j \sim_i S_k$ if and only if $s_j = s_k$ and $h_{S_j} = h_{S_k}$. Notice that player i's preferences establish an order $\succeq_i$ (a complete, reflexive and transitive binary relation) over the set N_i. A community/coalition formation game $(N, \succ)$ is a set of players and a profile of binary relations. Let $y_i(s_i, t_{S_i})$ denote the payoff function representing the preferences described above, thus:

$$y_i(s_k, t_{S_k}) > y_i(s_j, t_{S_j}) \Leftrightarrow S_k \succ_i S_j$$

where t_{S_i} denotes the vector of productivities of community S_i members and $\mathcal{T}_{S_i}$ is the set of all possible productivity vectors of community S_i. As commented in the introduction, hedonic preferences have been motivated and analyzed in a large number of models of coalition formation.

Apart from productivity, players are also characterized by their social identity, which is assumed to be either red (R) or blue (B). This can be interpreted as a perfectly observable physical feature such as race, sex, or language among others. So, overall, players come in four flavors: red-high, red-low, blue-high, and blue-low.[6] Players' social identity is assumed to be payoff irrelevant, so agents' preferences are color-blind.

Unlike social identity, players' productivity might be generally unobserved by others. According to categorical thinking, players' capacity to perceive others' productivity is affected by their social interactions. Partition π represents the pattern of these interactions; agents only interact (directly or indirectly) with the members of their respective community. From these social interactions, players extract information so that the productivity of the members of a community is known by all its members. Based on categorization evidence, these social interactions also constitute the experiences that agents store in their mental categories and use to make conjectures about the productivity of other communities' members. The accuracy of those conjectures is affected by that categorization. As explained in the introduction, the information obtained by agent i from the members of the less frequent social identity in their community (say red) will be more coarsely categorized and, consequently, i's predictions about the productivity of red agents in other communities will be less accurate. In this paper, it is assumed that the informational content of S_i members' signal/message about the vector of productivities of community S_j, $m_{S_i}(t_{S_j})$, is as follows:

(A1) If S_i is a single-agent community then $i \in S_i$ will not be able to detect others' productivity. Otherwise, if the proportion of blue (red) members of S_i exceeds 0.5 then S_i members will not be able to detect the actual productivity of red (blue) players of S_j, for any $S_j \neq S_i$;[7] the productivity of all other players will be perfectly detected.[8]

In what follows this informational structure is referred to as "categorical information". Formally, this informational structure involves incomplete information since agent $i \in S_i$ might not be able to fully observe t_{S_j}, for any $S_j \neq S_i$. In general, it might be that $m_{S_i}(t_{S_j}) = m_{S_i}(t'_{S_j})$ for $t_{S_j} \neq t'_{S_j}$. Following the previous intuition, notice that all the members of a community will receive the same signals.

A community is *completely segregated* if it contains agents of only one social identity. In general, segregation is measured in terms of the absolute

[6]In order to make an examination of segregation non-trivial, it is assumed that there is some agent in each of these four types.

[7]In case a community has the same number of reds and blues, any tie-break rule can be assumed. The effect of this tie-break rule on the results is marginal. In this paper it is assumed that blues are observed in this case.

[8]This assumption is consistent with categorical thinking in cases where communities are big and highly segregated. In such communities agents of one social identity hardly interact with different-identity agents whereas they "store" many past experiences with members of one social identity. The results of this paper must be thought in this context. More sophisticated informational structures are left for future research.

deviation from a 50-50 community, *i.e.* a community containing equal number of reds and blues. A community is *h-partially segregated* (or *l-partially segregated*) if it contains high (or low) productivity agents of only one social identity. A society is *completely segregated* if all its communities are completely segregated.

2.2. Equilibrium concepts

In this game, a typical strategy of player i consists of a set of intended social links. Nevertheless, players' payoff is determined by the community partition; thus, this can be considered a coalition formation game. As in Jackson and Wolinsky (1996), a link between two individuals (i) can be severed unilaterally but (ii) can only be created by mutual consent of the two involved agents. The requirement of mutual consent in the link creation combined with the multidimensional strategy space (players can announce any combination of links they wish) involve that this kind of games display a multiplicity of Nash equilibria. For instance, the partition with n isolated players is always a Nash Equilibrium. If players are allowed to coordinate bilaterally, instead, no mutually beneficial link is left aside and the multiplicity is reduced. The Nash equilibrium outcomes that fulfill this added (coalitional move) requirement are called pairwise-Nash equilibria (PNE).

Implicit in the PNE concept is that each individual has complete information about other players' types. Thus, the PNE concept is not appropriate if individuals have categorical information. Bayesian concepts allow for the analysis of incomplete information situations, but they assume that players commonly know a prior probability distribution over types. Instead of forcing this convergence in beliefs, I follow the steps of McBride (2006, 2006a) and use an adaptation of the Conjectural Equilibrium concept (which was designed for games with imperfect monitoring) to the case of incomplete information.[9] The resulting concept is the *Conjectural Pairwise Nash Equilibrium (CPNE)* and it is defined below[10]. Let $p_{S_i} : \mathcal{T}_{S_k} \rightarrow [0,1]$ be the subjective probability distribution of members of S_i over the possible productivity vectors of coalition S_k. The element $p_{S_i}(t_{S_k})$ denotes the subjective probability that members of S_i are assigning to the productivity vector t_{S_k}.

[9]Notice that McBride (2006, 2006a) apply this concept to a stability network analysis.

[10]Pairwise-stability is another equilibrium concept that has been extensively used for positive purposes due to its computational (relative) simplicity, and to its ability to generate sharp predictions in many contexts. Nevertheless, Pairwise-Stability is too weak for the purposes of the present work.

Definition 2.1. A partition π is a *CPNE* if there do not exist a player $i \in S_i$ and a coalition $S_k \in \pi \cup \{\emptyset\}$ such that

$$\sum_{t'_{S_i \cup S_k} \in \mathcal{T}_{S_i \cup S_k}} p_{S_i}(t'_{S_i \cup S_k}) y_i(s_i + s_k, t'_{S_i \cup S_k}) > y_i(s_i, t_{S_i})$$

and

$$\sum_{t'_{S_i \cup S_k} \in \mathcal{T}_{S_i \cup S_k}} p_{S_k}(t'_{S_i \cup S_k}) y_j(s_i + s_k, t'_{S_i \cup S_k}) > y_j(s_k, t_{S_k}) \text{ for some } j \in S_k$$

where $m_{S_i}(t'_{S_j}) = m_{S_i}(t_{S_j})$ for any $t'_{S_j} \in \mathcal{T}_{S_j}$ such that $p_{S_i}(t'_{S_j}) > 0$, for any S_i.

In words, a community partition is CPNE if no pair of players of different communities believe that they will be better off by merging their communities (this can be done by creating a link between them) and no player benefits from isolating. Moreover, no player's beliefs should be contradicted by her signal.

Notice that CPNE does not require that probabilities attributed to each state of the world are justified. Instead, it only requires that these probabilities are not contradicted by the observed signal/message. Rubinstein and Wolinsky (1994) and Gilli (1999) acknowledged this drawback for the Conjectural Equilibrium concept. They consider imposing common knowledge of rationality as a way to refine players' beliefs. Common knowledge of rationality can be imposed by assuming that players commonly know (a) the message function and (b) that everyone plays a best response to her conjectures. The CPNE partitions that fulfill this requirement are referred to as *rationalizable* CPNE and are formally defined below. Let $p_{S_i,S_j}(t_{S_k})$ denote the subjective probability that members of S_j assign to the productivity vector of community S_k according to S_i members' beliefs.

Definition 2.2. A partition π is a *rationalizable* CPNE if for any $i \in S_i$ there do not exist a player $j \in S_j$ and a coalition $S_k \in \pi \cup \{\emptyset\}$ such that

$$\sum_{t'_{S_j \cup S_k} \in \mathcal{T}_{S_j \cup S_k}} p_{S_i,S_j}(t'_{S_j \cup S_k}) y_j(s_j + s_k, t'_{S_j \cup S_k}) > \sum_{t'_{S_j} \in \mathcal{T}_{S_j}} p_{S_i,S_j}(t'_{S_j}) y_j(s_j, t'_{S_j})$$

and

$$\sum_{t'_{S_j \cup S_k} \in \mathcal{T}_{S_j \cup S_k}} p_{S_i,S_k}(t'_{S_j \cup S_k}) y_k(s_j + s_k, t'_{S_j \cup S_k}) > \sum_{t'_{S_k} \in \mathcal{T}_{S_k}} p_{S_i,S_k}(t'_{S_k}) y_k(s_k, t'_{S_k})$$

for some $k \in S_k$ where $m_{S_j}(t'_{S_k}) = m_{S_j}(t_{S_k})$ for any $t'_{S_k} \in \mathcal{T}_{S_k}$ such that $p_{S_i,S_j}(t'_{S_k}) > 0$.

Imposing common knowledge of rationality involves making signal functions (not actual types) common knowledge. So, according to this definition S_i members' beliefs about S_j members' beliefs cannot contradict the signal observed by S_j members. Additionally, rationalization of S_i members' beliefs involves that S_j members' actions should be optimal responses to those beliefs.

EXAMPLE 2.1. Imagine a society of n agents that contains members of two different social identities (blues and reds). Assume that all players are high-type agents and $\alpha_H, \alpha_L \in (0,1)$. Let us analyze a particularly interesting structure. Imagine a completely segregated society in which there are two communities, S_b and S_r, each containing all agents of one social identity. Under full information this partition is not an equilibrium because a pair of players of different communities will have incentives to merge their groups by creating a link between them. Contrarily, this partition can be sustained as a CPNE partition under categorical information when $s_i < \alpha_H(s_r + s_b)$ for $i = b, r$.[11] In this case, agents will not be able to observe the type of any of the members of the other community, so blue agents do not know the type of any red agent and viceversa. Consequently, blue and red agents can believe that all members of the other community are low-type players because the received signal does not contradict that conjecture. Under this conjecture, no pair of agents of different communities would have incentives to merge their communities because both believe that the resulting community $S_b \cup S_r$ would hold $h_{S_b \cup S_r} < \alpha_H$. Therefore, a completely segregated society can be sustained as a CPNE partition. Can this partition be rationalized? Stability requires that for any pair of high-type players of two different communities, at least one of them believes that the proportion of highs in the other community is sufficiently low. Rationalizing these beliefs involves that players' actions should constitute a best response to the believed state of the world. Notice that if the proportion of highs in a community was lower than α_H then some member of this community would have incentives to isolate[12]. Thus, no player can rationalize that the proportion of highs in the other community is lower than α_H so that any pair of players of two different communities would have incentives to merge them by creating a link between them. Therefore complete segregation cannot be rationalized in this case.

[11] This can only hold when $\alpha_H > 0.5$.

[12] If this community contains some high-type player then she will isolate. Otherwise, a low-type agent will isolate because $\alpha_L > 0$.

Before moving to the results, notice that the size of the set of CPNE partitions depends on the informational content of messages as follows:

Remark 2.1. If the messages $\{m'_{S_i}\}_{S_i \in \pi}$ contain more information than $\{m_{S_i}\}_{S_i \in \pi}$ then a CPNE partition under $\{m'_{S_i}\}_{S_i \in \pi}$ is also a CPNE partition under $\{m_{S_i}\}_{S_i \in \pi}$, but the converse is not necessarily true.

Intuitively, the higher is the informational content of messages, the lower is the number of unrealistic beliefs that an agent can conjecture about the actual state of the world. Thus, if the partition meets the stricter requirements for CPNE under $\{m'_{S_i}\}_{S_i \in \pi}$, it will certainly meet the requirements for CPNE under $\{m_{S_i}\}_{S_i \in \pi}$. For this reason, the equilibrium partitions of the present model could also be sustained in other environments with less informative messages.

3. RESULTS

In this section the CPNE concept is used to analyze the effects of categorical thinking on segregation. For the sake of comparison, results are presented gradually: first, I characterize equilibrium partitions under complete information; second, "categorical information" is introduced, and finally I analyze the effects of imposing "common knowledge of rationality". This will allow us to see the effects on segregation of each of these factors separately. Additionally, the analysis is divided into two different frameworks in order to illustrate how the assumptions about preferences affect the results.

3.1. Framework 1: $\alpha_H > 0$ and $\alpha_L = 0$

$\alpha_L = 0$ implies that low productivity players always prefer bigger communities, independently of the productivity of their members. In consequence, segregation by social identity among low productivity agents will never exist in equilibrium because all of them will be members of the same community.

The set of CPNE partitions under full information is characterized below.

Proposition 1. *Consider $\alpha_H > 0$, $\alpha_L = 0$, and full information. If $h_N \geq \alpha_H$ the* grand coalition *is the unique CPNE partition. Otherwise, the population will be split into two communities completely segregated by productivity in equilibrium.*

Proof. Since $\alpha_L = 0$ all low productivity agents must be members of the same community, say S_l. Notice that high productivity agents cannot be scattered across different communities in equilibrium; if this was so then each of these communities (say S_k) would hold: $h_{S_k} \geq \alpha_H$. Thus, members of those communities would have incentives to merge them by creating a link between them. In consequence, there are only two possibilities: either all high productivity agents are in S_l (thus, there is a unique community conforming the grand coalition) or all high productivity agents constitute a different community S_h. Notice that this possibility cannot hold in equilibrium when $h_N \geq \alpha_H$ because both members of S_h and S_l will have incentives to merge their communities. ∎

For any given profile of qualities and social identities, there are exactly two types of partitions that could be sustained as CPNE under full information; society will either concentrate into a grand coalition or be completely segregated by productivity. Anyway, social identity does not play any role here in shaping the equilibrium partitions.

Incomplete information will widen the set of equilibria. Let r_{S_i} be the proportion of red players in community S_i.

PROPOSITION 2. *Consider $\alpha_H > 0$, $\alpha_L = 0$, and categorical information. Only if $h_N \geq \alpha_H$ the* grand coalition *is a CPNE partition. Low-type players will be always concentrated into a unique community, S_l. Moreover, high type players can be scattered across two (or more) different communities $S_1, S_2 \neq S_l$ as long as:*

(i) $r_{S_j} < \bar{r}_{S_j} = \frac{\alpha_h(s_i+s_j)-s_i h_{S_i}}{s_j}$, *if* $r_{S_i} > 0.5$,
(ii) $r_{S_j} > 1 - \bar{r}_{S_j}$, *if* $r_{S_i} \leq 0.5$, *or*
(iii) S_i *is a single-agent community,*

where $i, j \in \{1, 2\}$ *and* $i \neq j$.

Proof. Since $\alpha_L = 0$ all low productivity players must share the same community S_l. The *grand coalition* can be sustained in equilibrium as in the complete information case. However, even if $h_N \geq \alpha_H$, two different communities (S_i and S_j) without low-productivity members can coexist in equilibrium if for any pair of agents $i \in S_i$ and $j \in S_j$ at least one of them, say i, believes that the other community contains a proportion of high-productivity agents lower than α_H. This belief will not contradict messages if and only if:

$$\alpha_H > \frac{s_i h_{S_i} + s_j r_{S_j}}{s_i + s_j}, \text{ when } r_{S_i} > 0.5$$

or

$$\alpha_H > \frac{s_i h_{S_i} + s_j (1 - r_{S_j})}{s_i + s_j}, \text{ when } r_{S_i} \leq 0.5$$

These two conditions can be rewritten as in the statement of the proposition. Finally, notice that isolated agents have no information about others, so they can believe that the proportion of the remaining communities is arbitrarily low. ▌

Therefore, categorical information allows for the dispersion of high-productivity agents among different communities. This can happen when agents' information about other communities is sufficiently limited. In case of multi-agent communities, this information crucially depends on segregation levels. Specifically, if two communities have different majority social identities then stability requires sufficiently high levels of segregation. In particular, for any α_H, two completely segregated communities without low-productivity agents can always be sustained in equilibrium since any of their members can believe that all agents in the other community are low-productivity individuals. Therefore, segregation by social identity among high-productivity agents could be severe in this case.

The rationalizability refinement will narrow the set of equilibrium partitions as follows.

Proposition 3. *Consider $\alpha_H > 0$, $\alpha_L = 0$, and categorical information. A rationalizable CPNE network is constituted by either (i) a unique* grand coalition *(only if $h_N \geq \alpha_H$) or (ii) two communities S_l and S_h such that S_l concentrates all low-productivity players and does not contain any high-productivity agent of the majority social identity of S_h (S_l is h-partially segregated).*

Proof. Since $\alpha_L = 0$ all low-productivity agents must be in the same community S_l. Rationalizability involves that players cannot believe that low-productivity agents are scattered among two or more communities. This has two implications: (i) since agents are able to observe at least one low-productivity agent in S_l, this community will be always identified and, consequently, the maximum number of communities is two (say S_h and S_l) and (ii) if there are two communities, any member of S_l should know that there are not low-productivity agents in S_h, so she is always willing to create a link with a member of S_h. In consequence, a partition with these two communities can be sustained in equilibrium if members of S_h are not willing to merge their community with S_l because they believe that $S_h \cup S_l$ is worse than S_h. This belief can only be rationalized, when members of S_h do not observe any high-productivity agent in S_l. Notice that if S_l contains some high-productivity agent then it should be that $h_{S_l} \geq \alpha_H$. In such a case, $S_h \cup S_l$ would be better than S_h. Therefore, in equilibrium, agents in S_h should not observe any high-productivity agent in S_l; in other words, S_l does not contain any high-productivity agent of the majority social identity of S_h. Finally, the *grand coalition* can also be sustained under the same conditions of the complete information case. ▮

Thus, the maximum number of communities is two, as in the complete information case. However, categorical information allows for the existence of segregation by social identity among high-productivity agents even after imposing the rationalizability refinement. Notice that case (ii) of Proposition 3 can only be sustained if S_l is h-partially segregated. Notice also that in this case a completely segregated community S_h can also be sustained as part of a rationalizable CPNE. In consequence, high-productivity agents of the minority social identity in S_h might be systematically excluded from this community by its members as a consequence of being more coarsely sorted due to their scarcity in S_h. This possibility of discrimination affecting minorities is in line with the predictions of Fryer and Jackson (2008) with respect to the negative consequences of categorical thinking on minorities.

As commented above, since $\alpha_L = 0$ low-productivity agents will not be segregated in equilibrium. Next, the case where $\alpha_L > 0$ is considered. As expected, segregation among low-productivity agents in a CPNE will be sustainable, but this change generates interesting effects on the set of rationalizable CPNE social structures.

3.2. Framework 2: $\alpha_H, \alpha_L > 0$

As in the previous case, the complete information results are presented first.

Proposition 4. *Consider $\alpha_H, \alpha_L > 0$ and complete information. In any CPNE partition all high-productivity agents are members of the same community S_h. This community contains all agents if $h_N \geq \max\{\alpha_H, \alpha_L\}$. Otherwise, $h_{S_h} \geq \max\{\alpha_H, \alpha_L\} > h_{S_h}\frac{s_h}{s_h+1}$ and low-type players not included in S_h are isolated.*

Proof. As shown in the proof of Proposition 1 high-productivity agents must concentrate into a unique community S_h. The members of this community will form links with outsiders as long as $h_{S_h}\frac{s_h}{s_h+1} > \max\{\alpha_H, \alpha_L\}$. Since $\alpha_L > 0$, low-productivity agents not included in S_h prefer to stay alone rather than in a community with only low-productivity agents. ▮

Under complete information, the society would be constituted by a unique multi-agent community and, generally, low-type isolated players. Complete segregation by productivity can only be obtained under arbitrary high values of α_H or α_L. Otherwise, this multi-agent community would contain low-productivity agents. In any case, social identity does not play any role in this complete information setting.

Let us consider categorical information. Apart from the equilibrium structures described in the previous proposition, there are other equilibria. In a CPNE, high and low-productivity agents can be scattered across different communities whenever (i) for any (multi-agent) community S_k, $h_{s_k} \geq \alpha_{t_i}$ for all $i \in S_k$ and (ii) the distribution of social identities is such that, for any pair of agents of different communities, the message received by at least one of them (say i) does not contradict the belief that the proportion of highs in the other community is sufficiently low. Consequently, apart from having isolated high productivity agents, multiple partitions are sustainable in equilibrium and segregation can exist. In particular, a society completely segregated by social identity can be sustained as CPNE if the proportion of highs in each social identity group is high enough to hold condition (i) above. Thus, as expected, segregation may be even more severe than in framework 1 because low-productivity agents can also be segregated.

Again, the "common knowledge of rationality" requirement narrows the set of equilibrium partitions. By assuming $\alpha_L > 0$, low productivity agents face more demanding requirements to keep their community. That restricts the set of possible beliefs. In spite of that, there can be segregation in equilibrium.

Proposition 5. *Consider $\alpha_H, \alpha_L > 0$, and categorical information. In a rationalizable CPNE there is a unique multi-agent community S_h such that $h_{S_h} \geq \alpha_{t_i}$ for any $i \in S_h$. High-productivity agents may not be included in S_h whenever they belong to the minority social identity group of S_h. Agents not included in S_h, if any, must remain isolated.*

Proof. Any multi-agent community (say S_k) must contain some high-productivity agent and, consequently, $h_{S_k} \geq \alpha_H$. Agents' beliefs cannot contradict this requirement. Thus, a partition cannot contain two multi-agent communities because if this was so then some pair of high-productivity members would have incentives to create a link between them and merge these communities. The same argument applies if there is a multi-agent community S_h and an isolated high-productivity agent of the majority social identity of S_h. Consequently, high-productivity agents not included in S_h must belong to the minority social identity of S_h. Since $\alpha_H, \alpha_L > 0$, agents not included in S_h are isolated and believe that all remaining isolated players are low-type agents. ∎

Thus, when $\alpha_H, \alpha_L > 0$ the set of rationalizable CPNE allows for the possibility of segregation. As in framework 1 and in line with the arguments in Fryer and Jackson (2008), high-productivity agents of the minority social identity in S_h might be systematically excluded from this community.

4. CONCLUSION

This paper presents a simple coalition/community formation game with incomplete information to illustrate a new and complementary explanatory cause of segregation called categorization. According to this cognitive process, the information obtained by the observer from their (direct or indirect) social contacts is stored in a number of "folders" so scarce that force her to categorize more coarsely the experiences with the members of the less frequent social groups (see Fryer and Jackson, 2008). Consequently, the predictions about the hidden features of the minority members would be less accurate. In line with this general argument, this paper assumes a particular informational structure: any agent can perfectly observe the productivity of all linked agents, either directly or indirectly, but only the productivity of non-linked agents who belong to the social identity the observer is mostly interacting with. I take advantage of the simplicity of this informational structure to derive clear results concerning the possibilities of segregation in this context. More sophisticated informational structures could capture the essence of categorical thinking more accurately at the cost of complicating the analysis. The study of alternative informational structures consistent with categorization constitutes an interesting line for future research. For example, an agent could receive signals about outsiders' types whose precision depends on the social composition of her local community. Notice that in segregated societies, the informational content of messages would not vary so much under this alternative information structure (in a completely segregated community nobody is able to observe any individual of the other social identity, so this social identity would still be widely unknown). For this reason, as announced by footnote 8, the results of this article must be thought in the support of the argument defending that extreme segregation can arise from categorical thinking. The informational assumption of this model is consistent with categorical thinking in those extreme situations.

A general contribution of this paper is the application of the CPNE concept to the study of coalition/community formation games with incomplete information. This concept can be applied to other coalition formation games, thereby allowing researchers to study the relationship between efficiency, stability and information in other settings.

The results of the present paper show that CPNE social structures can be segregated by social identity as a consequence of categorical thinking. In framework 1, high-productivity agents can be scattered across different communities that can be completely segregated by social identity. Segregation can be even more severe in the second framework because it can also affect low-productivity agents. The extreme case of complete segregation (population is split in two communities according to social identity) is sustainable in that framework. The rationalizability refinement reduces the possibilities of segregation in equilibrium in both frameworks. To the extent that the "common knowledge of rationality" can be associated to a higher level of "intelligence" one would expect agents with lower "intelligence" to be more likely to organize themselves in segregated societies. The results of the present paper are in line with this general argument. Nevertheless, segregation can be sustained even in rationalizable equilibria; this reinforces the robustness of categorization as a cause of segregation. Minority social identity members are particularly affected. In line with the predictions of Fryer and Jackson (2008), the results show that these agents might be systematically excluded from the community including high-productivity agents as a consequence of being more coarsely sorted due to their scarcity.

The main contribution of this paper is to show, in a game theoretical framework, that segregation can arise among self-interested players even when they have no a prioriaim to discriminate by social identity, because preferences are not affected by this aspect. This contrasts with taste discrimination models. Equal opportunity laws are usually premised on the notion that intergroup bias is malevolent in origin. In line with Fryer and Jackson (2008) or Krieger (1995), the present paper would suggest that courts should reformulate doctrine to reflect the reality that discrimination and segregation can result from things other than discriminatory intent.

REFERENCES

Allport, G.W., 1954. *The Nature of Prejudice*, Reading, MA: Addison Wesley.

Banerjee, S., H. Konishi and T. Sönmez, 2001. Core in a simple coalition formation game. *Social Choice and Welfare* **18**, 135-153.

Becker, G. S., 1971. *The Economics of Discrimination*, Chicago, IL: The University of Chicago Press.

Bertrand, M. and S. Mullainathan, 2003. Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. NBER working paper number 9873.

Bloch, F. and M.O. Jackson, 2006. Definitions of equilibrium in network formation games. *International Journal of Game Theory* **34, 3**, 305-318.

Bloch, F. and M.O. Jackson, 2007. The formation of networks with transfers among players. *Journal of Economic Theory* **133**, 83-110.

Bogomolnaia, A. and M.O. Jackson, 2002. The Stability of Hedonic Coalition Structures. *Games and Economic Behavior* **38**, 201-230.

Bondonio, D., 1998. Predictors of Accuracy in Perceiving Informal Social Networks. *Social Networks* **20**, 201-220.

Brown, C. and P. Gay, 1985. *Racial Discrimination 17 Years after the Act*. London: Policies Studies Institute.

Calvó-Armengol, A., 2004. Job contact networks. *Journal of Economic Theory* **115, 1**, 191-206.

Casciaro, T., 1998. Seeing Things Clearly: Social Structure, Personality and Accuracy in Social Network Perception. *Social Networks* **20**, 331-351.

Clark, W. A. V., 1991, Residential preferences and neighborhood racial segregation: a test of the Schelling segregation model. *Demography* **28**, 1-19.

Demange, G., 1994. Intermediate Preferences and Stable Coalition Structures. *J. Math. Econ.* **23**, 45-58.

Drèze, J.H. and J. Greenberg, 1980. Hedonic Coalitions: Optimality and Stability. *Econometrica* **48, 4**, 987-1003.

Epstein, J. M. and R. Axtell, 1996. *Growing Artificial Societies: Social Science from the Bottom Up*, Cambridge, MA: MIT Press.

Fiske, S.T., 1998. Stereotyping, Prejudice, and Discrimination, in D.T. Gilbert, S.T. Fiske and G. Lindzey, eds, *Handbook of Social Psychology* **vol 2**. New York: Oxford University Press, 357-414.

Friedkin, N.E., 1983. Horizons of Observability and Limits of Informational Control in Organizations. *Social Forces* **62**, 54-77.

Fryer, R. and M.O. Jackson, 2008. A Categorical Model of Cognition and Biased Decision Making. *The B.E. Journal of Theoretical Economcis. Contributions* **8, 1**, Article 6.

Gilli, M., 1999. On Non-Nash Equilibria. *Games and Economic Behavior* **27**, 184-203.

Goyal, S. and S. Joshi, 2006. Unequal Connections. *International Journal of Game Theory* **34, 3**, 319-349.

Greenberg, J. and S. Weber, 1986. Strong Tiebout Equilibrium under Restricted Preference Domain. *J. Econ. Theory* **38**, 101-117.

Greenberg, J. and S. Weber, 1993. Stable Coalition Structures with Unidimensional Set of Alternatives. *J. Econ. Theory* **60**, 62-82.

Guesnerie, R., and C. Oddou, 1981. Second Best Taxation as a Game. *Journal of Economic Theory* **25**, 67-91.

Hamilton, D.L., 1981. *Cognitive Processes in Stereotyping and Inter-group Behavior*, Hillsdale, NJ: Erlbaum.

Hubbick, J. and S. Carter, 1980. *Half a Chance? A Report on Job Discrimination against Young Black Males in Nottingham*. London: Commission for Racial Equality.

Jackson, M.O. and A. Wolinsky, 1996. A strategic model of social and economic networks. *Journal of Economic Theory* **71**, 44-74.

Jowell, R. and P. Prescott-Clarke, 1970. Racial Discrimination and White-Collar Workers in Britain. *Race* **11**, 397-417.

Krieger, L.H., 1995. The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity. *Stanford Law Review* **47**, 1161-1248.

Kumbasar, E., A.K. Romney, and W. Batchelder, 1994. Systemic Biases in Social Perception. *American Sociological Review* **100**, 477-505.

Laumann, E., 1969. Friends of Urban Men: An Assessment of Accuracy in Reporting their Socioeconomic Attributes, Moral Choice, and Attitude Agreement. *Sociometry* **32**, 54-69.

Macrae, N. and G. Bodenhausen, 2000. Social Cognition: Thinking Categorically About Others. *Annual Review of Psychology* **51**, 93-120.

Markman, A.B. and D. Gentner, 2001. Thinking. *Annual Review of Psychology* **52**, 223-247.

McBride, M., 2006. Imperfect monitoring in communication networks. *J. Econ. Theory* **126**, 97-119.

McBride, M., 2006. Limited Observation in Mutual Consent Networks. *The B.E. Journal of Theoretical Economics. Advances* **6, 1**, Article 3.

Pancs, R. and N.J. Vriend, 2007. Schelling's spatial proximity model of segregation revisited. *Journal of Public Economics* **91**, 1-24.

Phelps, E. S., 1972, The Statistical Theory of Racism and Sexism. *American Economic Review* **62, 4**, 659-661

Rubinstein, A. and A. Wolinsky, 1994. Rationalizable Conjectural Equilibrium: Between Nash and Rationalizability. *Games and Economic Behavior* **6**, 299-311.

Schelling, T.C., 1969. Models of segregation. *American Economic Review, Papers and Proceedings* **59**, 488-493.

Schelling, T.C., 1971. Dynamic models of segregation. *Journal of Mathematical Sociology* **1, 2**, 143-186.

Schelling, T.C., 1971. On the ecology of micromotives. *The Public Interest* **25**, 61-98.

Schelling, T.C., 1972. A process of residential segregation: neighborhood tipping, in Pascal, A.H. (Ed.), *Racial Discrimination in Economic Life*. Lexington, MA: Lexington Books.

Schelling, T.C., 1978. *Micromotives and Macrobehavior*. New York: Norton Ed.

Sporer, S., 2001. Recognizing Faces of Other Ethnic Groups: An Integration of Theories. *Psychology, Public Policy, and Law* **7**, 36-97.

Tajfel, H., 1969. Cognitive Aspects of Prejudice. *Journal of Social Issues* **25**, 79-97.