

# Correlated equilibrium, conformity and stereotyping in social groups\*

Edward Cartwright  
Department of Economics  
Keynes College, University of Kent,  
Canterbury CT2 7NP, UK  
E.J.Cartwright@kent.ac.uk

Myrna Wooders  
Department of Economics, Vanderbilt University  
Nashville, TN 37235 USA  
myrna.wooders@vanderbilt.edu

This version; March 2010

*Keywords:* correlated equilibrium, conformity, stereotyping, social groups, within-group fairness, homophily, social norms

---

\*This paper is an outgrowth of working papers Cartwright and Wooders (2005, 2006) and Cartwright (2003). We are grateful to participants for comments at presentations of this work at the European Meetings of the Econometric Society, Budapest, 2007; Hebrew University Center for Rationality, 2006; Stony Brook Game Theory International Conference, 2006; SING (Spanish, Italian, Netherlands Game Theory) Meeting, (plenary lecture), 2006; Society for Economic Design conference 2004; Northwestern University Kellogg Centre for Game Theory 2004; and at a number of other universities where these papers were presented. We thank Rod Garrett and John Wu for reading the paper.

**Abstract:** We demonstrate that correlated equilibrium can express conformity to norms and the coordination of behavior within social groups. Given a social group structure (a partition of players into social groups), we propose three properties that one may expect of a correlated equilibrium consistent with social group structures satisfying behavioral conformity. These are: (a) within-group anonymity (conformity within groups); (b) group independence (no conformity between groups), and (c) predictable social group behavior (ex-post stability). We also consider stereotyped beliefs – beliefs that all (other) players in a social group can be expected to behave in the same way. We demonstrate that:

- (1) Correlated equilibrium satisfying both (a) and (b) exist very generally;
- (2) If there are many players then a correlated equilibrium satisfying (a), (b) and (c) exists;
- (3) Stereotyping is not costly to the player who stereotypes.

# 1 Introduction

An individual conforms if he chooses to follow the norms of the society to which he belongs. The behavior prescribed by a norm can be quite subtle or unexpected. For example, Sugden (1989) describes an unwritten rule about the gathering of driftwood after a storm on the Yorkshire coast. The first person to find some driftwood can make a pile out of the driftwood and simply place two rocks on top the pile. These two rocks establish his property rights to the wood, even if left on the beach unattended, for up to two high tides. The crucial feature of this norm and many norms or conventions is that adherence to the norm requires *different individuals to take different actions* and for *actions to be conditional on some external signal*. Conditional on the random (or not necessarily random) event of who finds the driftwood first, one person will claim the driftwood and others will not. To the best of our knowledge, there is no game theoretic work that is able to capture such norms and conformity.<sup>1</sup> The main objective of this paper is to introduce *a model of conformity where, because actions are conditional on signals, individuals can perform different actions but still conform to the same norm*. As one application of the model, we introduce the notion of stereotyping and find that, although stereotyping may be costly to those being stereotyped, it need not be costly to the individual who stereotypes.

Our basic premise is that, with some additional properties, *correlated equilibrium* satisfying certain additional properties provides a natural way to capture conformity. This is because correlated equilibrium, in contrast to Nash equilibrium, allows player actions to be statistically dependent on some random signals external to the model (Aumann 1974, 1987). As suggested above, and noted by Sugden (1989) amongst others, such external signals are fundamental to many conventions and norms. Indeed, one interpretation of correlated equilibrium is that a mediator instructs players to take actions according to some commonly known probability distribution. In this paper we think of the mediator as a social norm that assigns roles to individual players in a society, where a role is equivalent to an action. If, assuming others will do so as well, it is in the interests of each player to assume the role assigned to him by the norm, then the probability distribution over roles is a correlated equilibrium (Aumann 1987; Forges 1986; Dhillon and Mertens, 1996).<sup>2</sup>

Given our interpretation of correlated equilibrium, we can think of every player as using the strategy “if told to play action  $x$  then play action  $x$ ”. Thus, a correlated equilibrium can capture the idea that players use the same strategy while potentially performing different actions. That actions are conditioned on signals or roles also recognizes how conformity can lead to coordinated actions

---

<sup>1</sup>Instead, it is typical to think of conformity as requiring individuals to choose the same actions (for example: Bernheim, 1994; Wooders, Cartwright and Selten, 2006).

<sup>2</sup>Given that roles may be correlated across individuals, the set of correlated equilibria is generally larger than the set of Nash equilibria. This is partly responsible for the set of correlated equilibria having many appealing properties; for example, the set of correlated equilibrium is nonempty, compact, convex and easy to describe (Aumann 1974, 1987). See also Hart (2005) for a discussion of recent work (in collaboration with Mas-Colell) on how adaptive learning leads to correlated equilibrium play.

within social groups (Johnson and Johnson 1977; Selten 1980). Clearly, however, some care is needed because merely using the concept of correlated equilibrium has created a setting where every player uses the same strategy. *We do need, therefore, to be sure that “play the same strategy” equates with “behaves in the same way” and “conforms”.* To do this we impose conditions on how the mediator or social norm distributes roles in order to guarantee that a correlated equilibrium can be interpreted as exhibiting behavioral conformity. We first assume that the population is partitioned into social groups. We then consider four conditions on a correlated equilibrium along with a partition of the set of players into social groups:

*Within-group anonymity* requires that any two individuals within the same group have the same probability of being assigned each role. Thus, not only will individuals in the same group use the strategy, “if told to play action  $x$  then play  $x$ ”, they will also have the same chance of being told to play action  $x$ . This means that, ex-ante, before they know their assigned roles, any two individuals in the same social group are expected to behave in the same way. For example, when two Yorkshire beachcombers independently choose to go to the beach in search of driftwood they may be equally likely to be the first to arrive at the right stretch of beach and will thus be assigned the role “claim the wood”.<sup>3</sup>

*Group independence* requires that the distribution of roles be statistically independent between different groups and thus rules out any correlation of actions across groups. Correlation between social groups is typically unlikely (Tajfel 1978) and so it seems important to rule this out.<sup>4</sup> More fundamentally, however, group independence draws the boundaries between social groups.

*Homophily* requires that individuals within a social group share similar preferences and that the actions of members of the same social group have similar effects on others. This is a property commonly observed in social groups (Brown 2000; Currarini, Jackson and Pin 2008). We stress that homophily may hold even though players within the same social group are not similar in all respects. Some individual characteristics directly affect payoffs; we call such characteristics the *attributes* of players. The attribute of a player is given by his preferences and how his choice of actions affects others. Other underlying player characteristics, which do not directly affect payoffs, may well serve as signals. For example, all automobile drivers may have similar preferences – to safely proceed through an intersection of two roads. The relative positions of drivers at the intersection can serve as a signal to correlate actions. If two drivers arrive at the same time then

---

<sup>3</sup>Within-group anonymity results in equity of opportunity, whereby the distribution of roles within a group can be seen as fair, and equity of expected payoff, whereby outcomes can be seen as fair. The importance of fairness and within group equity is well known (Johnson and Johnson, 1987; Tajfel, 1978; Rabin, 1993; Brown, 2000).

<sup>4</sup>Ruling out correlation between groups does not rule out coordination between groups.

the norm may be that “Left yields to Right.”<sup>5</sup> As another example, in some situations men and women may have similar preferences and their actions may have the same effect on other players. In such situations, gender may serve as a signal. For example, the norm may be that women cook and men wash the dishes. Gender is not directly relevant but may be a good signal with which to correlate actions so that dinner is cooked and the dishes washed.

*Predictable group behavior* requires that the number of individuals in each social group who will play each action is known ex-ante. This implies that the behavior of a social group is predictable even if the behavior of any individual in the group is not predictable. For example, we can predict that some beachcomber will claim a find of driftwood but it is not known which beachcomber will have that particular role. Predictable group behavior guarantees ex-post stability in the sense that no member of a social group would want to change his action once he knows his own role and the roles of others.

Behavioral conformity and consistency of a correlated equilibrium with a social norm, two of the main concepts of this paper, can now be discussed using the above properties.

**Behavioral conformity:** Together, within-group anonymity, group independence and homophily imply that individuals within the same group have similar attributes and can be expected to behave similarly, while individuals in different groups may behave in different ways. We say that a correlated equilibrium with these properties *satisfies behavioral conformity*. Our first result (Theorem 1) provides a family of games for which near to every Nash equilibrium there is a correlated equilibrium satisfying behavioral conformity. The only requirement we impose is continuity with respect to player attributes. This appears a very mild requirement and we therefore show that *a correlated equilibrium satisfying behavioral conformity exists very generally*.

**Consistency with a social norm:** Behavioral conformity is not sufficient for correlated equilibrium to satisfy the notion of a social norm. We propose that to interpret members of a social group as conforming to a norm requires that the aggregate behavior of the group membership be both predictable and stable. We impose predictability of group behavior and then demonstrate that, when this condition is satisfied, we interpret correlated equilibrium as *consistent with a social norm*. Our second main result (Theorem 2), demonstrates that, with

---

<sup>5</sup>For this example, we have in mind the following sort of strategic game, where  $W$  means “wait at the intersection for the other automobile” and  $G$  means “go through the intersection.” The positions of the drivers relative to each other play no direct role and thus can serve as a signal.

Driver 1 \ Driver 2	$W$	$G$
$W$	0, 0	0.9, 1
$G$	1, 0.9	-1, -1

a Lipschitz continuity condition on utility functions, and many players, near to any Nash equilibrium there is a correlated equilibrium that is consistent with a social norm (that is, a correlated equilibrium that satisfies behavioral conformity and predictable group behavior), even with the requirement that there be a relatively small number of social groups. The main additional requirement is that the game have many, anonymous players. We show, therefore, that a *correlated equilibrium consistent with a social norm exists for an interesting class of game*.

It is noteworthy that we are able to obtain the existence of a *correlated* equilibrium satisfying behavioral conformity under much milder conditions than we obtained the existence of a Nash equilibrium satisfying behavioral conformity. This is most apparent by comparing the analogous Theorem 2 of Wooders et. al. (2006) with Theorem 1 in this paper. Unlike the prior Theorem, our Theorem 1 does not require many players, nor does it require social groups to have a specific compositions. These are significant relaxations permitting, for example, an individual player to have a substantial influence on others. We are able to obtain these relaxed sufficient conditions because the mediator correlates actions. It is therefore important to recognize that individuals can and do correlate their behavior to mutual advantage, even if doing so requires different individuals to perform different actions and receive inequitable rewards (e.g. Schelling 1960; Hayek 1982; Sugden 1989; Friedman 1996; Van Huyck et al. 1997; Rapaport, Seale, and Winter 2002; Hargreaves-Heap and Varoufakis 2002). It has also been long recognized that conformity, often subconscious, to established rules and norms of behavior facilitates such correlation of behavior (Hayek 1960, 1982; Sherif 1966; Tajfel 1978; Johnson and Johnson 1987; Brown 2000).<sup>6</sup>

A further consideration is that *perceptions*, and not necessarily reality, may matter in terms of whether individuals think of outcomes as fair and are therefore willing to conform and play their assigned roles (Hogg and Vaughan 2005). For example, if an outcome is perceived as equitable, it may not matter whether it is in fact equitable. It also may be the case that individual players perceive all members of a social group as similar or the same and thus stereotype others according to their social group memberships. To take these two considerations into account we turn to *subjective* correlated equilibrium. A subjective correlated equilibrium extends the notion of correlated equilibrium by allowing players to have differing beliefs about the probability with which roles are distributed. We say that beliefs are stereotyped if each player expects players in the same social group to behave in the same way. Stereotyping may be interpreted as a form of bounded rationality and allows for simpler correlating devices. We demonstrate (Theorem 3) that stereotyping can be consistent with correlated equilibrium. We also demonstrate that a player who stereotypes (perhaps incorrectly) could not do better by having non-stereotyped beliefs. Thus, there is no incentive for players who stereotype to revise or correct their beliefs.

We proceed as follows: Section 2 introduces the model, Section 3 the proper-

---

<sup>6</sup>The economic literature on conformity includes Akerlof (1980), Elster (1989) and Bernheim (1994).

ties of social groups, Section 4 provides the main results and Section 5 concludes.

## 2 Model and notation

A *game*  $\Gamma$  is given by a triple  $(N, A, \{u_i\}_{i \in N})$  consisting of a finite *player set*  $N = \{1, \dots, n\}$ , a finite set of *K actions*  $A = \{1, \dots, K\}$ , and a set of *payoff functions*  $\{u_i\}_{i \in N}$ . An *action profile* consists of a vector  $\bar{a} = (\bar{a}_1, \dots, \bar{a}_n)$  where  $\bar{a}_i \in A$  denotes the *action of player i*. The set of action profiles is given by  $A^N$ . For each  $i \in N$  the payoff function  $u_i$  maps  $A^N$  into the real line  $\mathbb{R}$ .

A *strategy* in game  $\Gamma$  is given by a randomization  $\sigma = (\sigma(1), \dots, \sigma(K))$  over the set of actions where  $\sigma(k)$  denotes the probability that the player will play action  $k \in A$ . Let  $\Sigma = \Delta(A)$  denote the set of strategies. A *strategy profile* consists of a vector  $\bar{\sigma} = (\bar{\sigma}_1, \dots, \bar{\sigma}_n)$  where  $\bar{\sigma}_i \in \Sigma$  denotes the *strategy of player i*. We assume von-Neumann Morgenstern expected utility functions and, with a slight abuse of notation, denote by  $u_i(\bar{\sigma})$  the expected payoff to player  $i$  given strategy profile  $\bar{\sigma}$ . Strategy profile  $\bar{\sigma}$  is a *Nash  $\varepsilon$ -equilibrium* for some real number  $\varepsilon \geq 0$  if

$$u_i(\bar{\sigma}) \geq u_i(\sigma, \bar{\sigma}_{-i}) - \varepsilon$$

for all  $i \in N$  and  $\sigma \in \Sigma$ .

### 2.1 Pregames

We make use of a non-cooperative pregame, which allows us to consider families of games derived from a common underlying structure and to formalize a notion of similar games.<sup>7</sup> Informally, a non-cooperative pregame consists of a set of player attributes, a set of actions, and a universal preference function. Given a finite set of players, a game is induced by ascribing a point in attribute space to each player. The utility function of a player in the induced game is determined by the universal payoff function. As discussed by Wooders et. al. (2006), a pregame need not imply any assumptions on the induced games. It does, however, provide a useful framework in which to treat a family of games, all induced from a common strategic setting.

More formally, let  $\Omega$  be an *attribute space*, a compact, convex subset of a normed vector space equipped with a metric denoted by *dist*. Let  $A$  be a set of *actions*. Let  $W$  denote the set of all mappings with finite support from  $\Omega \times A$  into the non-negative set of integers  $\mathbb{Z}_+$ .<sup>8</sup> A member of  $W$  is called a *weight function* and, for any  $w \in W$ ,  $w(\omega, k)$  is interpreted as a number of players

<sup>7</sup>Noncooperative pregames appear in Cartwright and Wooders (2009) and Wooders et al. (2006). Similar sorts of concepts have a long history in economics and game theory. For example, in the context of an exchange economies, a “pre-economy” is a space of preferences and a set of possible endowment vectors. An economy is then determined by a set of economic agents and a function ascribing a preference relation and an endowment vector to each player. In cooperative games, a pregame is a set of player attributes and a function defining a payoff possibilities set for all possible finite sets of players described by their attributes.

<sup>8</sup>A function  $w$  from  $\Omega \times A$  has *finite support* if  $w(\omega, k) \neq 0$  for only a finite number of elements  $(\omega, k)$  in  $\Omega \times A$ .

with attribute  $\omega$  who choose action  $k$ . A *universal preference function*  $h$  is a mapping from  $\Omega \times A \times W$  into the set of real numbers  $\mathbb{R}$ . In interpretation  $h(\omega, k, w)$  is the payoff to a player of attribute  $\omega$  if he plays action  $k$  and the actions of all players are summarized by  $w \in W$ . A *non-cooperative pregame* is a triple  $\mathcal{G} = (\Omega, A, h)$ .

Given a pregame  $\mathcal{G} = (\Omega, A, h)$ , let  $N = \{1, \dots, n\}$  be a finite set and let  $\alpha$  be a mapping from  $N$  to  $\Omega$ , called an *attribute function*. The pair  $(N, \alpha)$  is a *population* where  $N$  is a set of players and  $\alpha$  provides a description of the players in terms of their attributes. Given action profile  $\bar{a} \in A^N$  we say that weight function  $w_{\alpha, \bar{a}} \in W$  is *relative to*  $\bar{a}$  if

$$w_{\alpha, \bar{a}}(\omega, k) = |\{i \in N : \alpha(i) = \omega \text{ and } \bar{a}_i = k\}|$$

for all  $k \in A$  and all  $\omega \in \Omega$ . An *induced game*  $\Gamma(N, \alpha)$  can now be defined:

$$\Gamma(N, \alpha) = (N, A, \{u_i^\alpha : A^N \longrightarrow \mathbb{R}_+\}_{i \in N})$$

where

$$u_i^\alpha(\bar{a}) \stackrel{\text{def}}{=} h(\omega, \bar{a}_i, w_{\alpha, \bar{a}})$$

for all  $\omega \in \Omega$  and  $\bar{a} \in A^N$ . We note that players who are ascribed the same attribute have the same payoff function.

## 2.2 The mediator

Given an induced game  $\Gamma(N, \alpha)$  we think of a *mediator* who assigns (or suggests) a role from set  $A$  to each player. The player then chooses an action that may or may not be consistent with his assigned role. In interpretation we shall equate the mediator with “society” that prescribes roles to players according to societal norms; players then choose to conform or deviate from their prescribed role. Returning to a earlier examples, society may suggest that “Left” yields to “Right” or that the way to lay claim to a find of driftwood is to pile up the driftwood and place two rocks on top of the pile.

The mediator is modeled by a correlating device given by a *probability distribution*  $p$  over action profiles. Thus,  $p(\bar{a})$  denotes the probability that players will be assigned roles consistent with action profile  $\bar{a}$ . We shall denote by  $p(\bar{a}_{-i}|\bar{a}_i)$  the probability of role assignments being consistent with  $\bar{a}$ , conditional on player  $i$  having role  $\bar{a}_i$ . We shall denote by  $p_i$  the marginal distribution of  $p$ , where  $p_i(k)$  denotes the probability that player  $i$  is assigned role  $k$ .<sup>9</sup> Let  $P$  denote the set of possible correlating devices.

The correlating device  $p$  transforms game  $\Gamma(N, \alpha)$  into a *game with roles*, denoted by  $\Gamma^p(N, \alpha)$ .<sup>10</sup> In game  $\Gamma^p(N, \alpha)$ , action choice can be made conditional on assigned role. A *behavioral rule for player*  $i \in N$  in game  $\Gamma^p(N, \alpha)$  is a

<sup>9</sup>Formally,  $p_i(k) = \sum_{\bar{a} \in A^N : \bar{a}_i = k} p(\bar{a})$  and, when  $p_i(\bar{a}_i) > 0$ , it holds that  $p(\bar{a}_{-i}|\bar{a}_i) = p(\bar{a})/p_i(\bar{a}_i)$

<sup>10</sup>Note that, although his payoff may indirectly depend on roles through the choice of action induced by a distribution of roles, a player’s payoff does not directly depend on his role or the roles of other players.

function  $b_i$  mapping the set of signals  $A$  to the set of actions  $A$ . In interpretation,  $b_i(k)$  is the action performed by player  $i$  if he is assigned role  $k$ . Of primary interest is the *conformist behavioral rule*  $b_i^c(k) = k$ , that is, the behavioral rule where player  $i$  plays the action consistent with his assigned role.

We shall assume for the present that correlating device  $p$  is common knowledge and players have consistent beliefs with respect to  $p$ . We relax this assumption in Section 4.2. Given correlating device  $p$ , we can now define a payoff function  $U_i^\alpha : P \rightarrow \mathbb{R}$  for each player  $i \in N$ , where

$$U_i^\alpha(p) = \sum_{\bar{a} \in A^N} p(\bar{a}) u_i^\alpha(\bar{a})$$

denotes the expected payoff of player  $i$  if roles are assigned according to distribution  $p$  and each player  $j \in N$  follows the conformist behavioral rule  $b_j^c$ .<sup>11</sup> Let

$$U_i^\alpha(p|\bar{a}_i) = \sum_{\bar{a}_{-i}} p(\bar{a}_{-i}|\bar{a}_i) u_i^\alpha(\bar{a})$$

denote the expected payoff of player  $i$  conditional on being assigned role  $\bar{a}_i$ .

### 2.3 Correlated equilibrium

A correlating device  $p$  is a correlated equilibrium if no player can do better by deviating from his assigned role. That is, knowing  $p$  and knowing his assigned role (and expecting all other players to conform to their assigned roles) each player does best by conforming to his assigned role. We shall be interested in an approximate correlated equilibrium where no player can gain more than some  $\varepsilon$  by not conforming. Formally, for any  $\varepsilon \geq 0$  we say that *correlating device*  $p$  is a *correlated  $\varepsilon$ -equilibrium of game  $\Gamma(N, \alpha)$*  if and only if

$$U_i^\alpha(p|\bar{a}_i) \geq \sum_{\bar{a}_{-i}} p(\bar{a}_{-i}|\bar{a}_i) u_i^\alpha(k, \bar{a}_{-i}) - \varepsilon \quad (1)$$

for all  $i \in N$ , any action  $k$  and any  $\bar{a}_i$ .<sup>12</sup> We refer to a correlated 0-equilibrium as a correlated equilibrium.<sup>13</sup> If  $\bar{\sigma}$  is a *Nash  $\varepsilon$ -equilibrium* of game  $\Gamma(N, \alpha)$  then there exists a correlated  $\varepsilon$ -equilibrium in which the mediator independently assigns player  $i$  action (or role)  $k$  with probability  $\bar{\sigma}_i(k)$  for each  $i \in N$ .

<sup>11</sup> In other words,

$$U_i^\alpha(p) = \sum_{\bar{a} \in A^N} p(\bar{a}) u_i^\alpha(\bar{b}_1^c(\bar{a}_1), \dots, \bar{b}_n^c(\bar{a}_n)).$$

<sup>12</sup>Formally, we only require (1) to hold for actions  $\bar{a}_i$  that player  $i$  can be assigned with positive probability.

<sup>13</sup>A correlated equilibrium thus defined is consistent with the usual definition of correlated equilibrium. An approximate correlated equilibrium equates to a natural approximation to a correlated equilibrium as usually defined, although ‘role’ is often termed ‘signal’ (Fudenberg and Tirole 1998). Note, however, that the use of the term  $\varepsilon$ -correlated equilibrium by Myerson (1986) has a different meaning to the one here.

### 3 Social groups

As discussed in the introduction, to make correlated equilibrium an expression of behavioral conformity within social groups and consistent with a social norm it is necessary to impose conditions on the probability distribution over roles. In this Section we shall define and motivate the properties that we shall use. We first require two definitions.

Given a game  $\Gamma(N, \alpha)$  a *social group structure* is given by a partition  $\Pi = \{N_1, \dots, N_G\}$  of the player set into  $G$  subsets for some integer  $G$ . We refer to each  $N_g$  as a *social group*. An important property of a social group structure will be the *diameter of the partition*,  $D_{\alpha, \Pi}$ , where

$$D_{\alpha, \Pi} = \max_{N_g \in \Pi} \max_{i, j \in N_g} \{dist(\alpha(i), \alpha(j))\}$$

is the maximum difference in attribute between two players in the same social group. In interpretation, if  $D_{\alpha, \Pi}$  is small, then players within each social group are all similar in terms of their attributes.

Interchanging the roles assigned to players in the same social group will play an important role in our work. We say that an action profile  $\bar{a}'$  is a *permutation* of another action profile  $\bar{a}$  if the number of players in each social group playing each strategy (or assigned each role) is the same.<sup>14</sup> Given social structure  $\Pi$  and action profile  $\bar{a}$  let  $\mathcal{P}^{\Pi}(\bar{a})$  denote the set of action profiles that are permutations of  $\bar{a}$ .

**Within-group anonymity:** Given population  $(N, \alpha)$ , social group structure  $\Pi$  and correlating device  $p$  we say that correlating device  $p$  satisfies within-group anonymity if  $p$  treats players from the same social group in an identical way. Formally, given any two action profiles  $\bar{a}$  and  $\bar{a}' \in A^N$ , if  $\bar{a}' \in \mathcal{P}^{\Pi}(\bar{a})$  then:

$$p(\bar{a}) = p(\bar{a}').$$

Within-group anonymity captures the notion that players within the same social group are expected to behave in the same way. In particular, within-group anonymity implies an equality of opportunity (or of responsibility) within groups because any two players belonging to the same social group have the same probability of being assigned each role within the group.<sup>15</sup> This can be seen as fair and, as we shall see below, results in an equality of expected payoff.

We now turn to our second property, group independence. This property dictates that the mediator not correlate the actions of players in one social group with those of another social group.

<sup>14</sup>More precisely, if  $\#^{\Pi}(\bar{a}, k, g) = |\{i \in N_g : \bar{a}_i = k\}|$  denotes the number of players in group  $N_g$  who play action  $k$  then action profile  $\bar{a}'$  is a *permutation* of  $\bar{a}$  if  $\#^{\Pi}(\bar{a}, k, g) = \#^{\Pi}(\bar{a}', k, g)$  for all  $k$  and  $N_g$ .

<sup>15</sup>For instance, if  $i, j \in N_g$ , then  $p_i(k) = p_j(k)$  for all  $k \in A$ .

*Group independence:* Given population  $(N, \alpha)$ , social group structure  $\Pi$  and correlating device  $p$ , the device  $p$  satisfies group independence if there is no correlation of roles between groups, that is, for any  $\bar{a} \in A^N$

$$p(\bar{a}) = p_1(\bar{a}_{\Pi_1})p_2(\bar{a}_{\Pi_2}) \cdots p_G(\bar{a}_{\Pi_G})$$

where  $p_g(\bar{a}_{\Pi_g})$  is the marginal probability that members of group  $\Pi_g$  have the roles given by  $\bar{a}$ .

If social groups are distinct then correlation of actions between groups may be difficult or unlikely; for example, players in different social groups may not as easily identify or communicate with each other as those in the same social group. (Note, however, that a lack of correlation does not imply a lack of coordination between groups as induced by an equilibrium). The main function of group independence is to define the boundaries between groups.

Our next property serves to further distinguish between groups. Specifically, it captures the notion that individuals within the same group are likely to have similar attributes.

*Homophily:* Given population  $(N, \alpha)$ , a social group structure  $\Pi = \{N_1, \dots, N_G\}$  satisfies *homophily* if the attribute space  $\Omega$  can be partitioned into  $G$  convex subsets  $\Omega_1, \dots, \Omega_G$  such that, for any  $i \in N$ , if  $i \in N_g$  then  $\alpha(i) \in \Omega_g$ .

Informally, homophily requires that any player whose attribute is intermediate between two members of a particular social group should also belong to that same social group. Note this does not necessarily imply that two players with similar attributes belong to the same social group. It does, however, suggest that any two players within a social group do have some similarity in attributes.

To briefly recap, within-group anonymity implies *similarity in behavior within groups*, group independence implies *boundaries between groups* while homophily implies that *group membership correlates with attribute*. Together, we shall think of these as the main criteria for behavioral conformity.

**Behavioral conformity.** Given population  $(N, \alpha)$ , we say that social group structure  $\Pi$  and correlating device  $p$  satisfy *behavioral conformity* if  $p$  satisfies within-group anonymity and group independence and  $\Pi$  satisfies homophily.

We now introduce our fourth property, which requires that the behavior of the social group be known ex-ante, even if the behavior of any one individual within the group is not known.

*Predictable group behavior:* Given population  $(N, \alpha)$ , social group structure  $\Pi$  and correlating device  $p$ , device  $p$  satisfies predictable group behavior if, for each group  $N_g \in \Pi$ , the number of players in the group who will play each action is known for sure ex-ante. Formally, for any action profiles  $\bar{a}$  and  $\bar{a}'$ , if  $p(\bar{a}), p(\bar{a}') > 0$  then  $\bar{a}' \in \mathcal{P}^\Pi(\bar{a})$ .

**Consistency with a social norm:** Given population  $(N, \alpha)$ , we say that social group structure  $\Pi$  and correlating device  $p$  are *consistent with a social norm* if  $p$  satisfies behavioral conformity and predictable group behavior.

The notion of a norm suggests that in some way the behavior of a social group should be predictable. Predictable group behavior imposes a strong form of predictability in which the behavior of the group is known for sure. Also, if a norm is to be followed over time (which seems an essential property of the notion of a norm) then it is preferably that players should have an incentive to conform ex-post, as well as ex-ante, to the assignment of roles. Predictable group behavior, as we shall see, results in ex-post stability, meaning that players would indeed want to conform even when they know the roles assigned to all other players.

Besides satisfying behavioral conformity and predictable group behavior there are a number of motivations to *bound the number of societies*. These are discussed at some length in Wooders et. al. (2006). We note here that the notion of a social group may already suggest that the group contains many members (in fact, bounding the number of social groups only implies that, if there are many players, some groups must contain many members). In addition, in games with many players it is desirable to have a relatively small number of social groups. This will allow simpler correlation devices and also gives more meaning and power to stereotyping and to social norms generally.

A related and interesting side issue is that of optimal social group size. Within-group anonymity and predictable group behavior require that players in the same social group behave identically and therefore suggests that small, homogenous, social groups are advantageous. In our framework, correlation of actions allows, however, as we shall see in Section 4.1, increased payoffs. In order to realize these gains and maintain group independence it would seem that larger social groups are advantageous. This creates countervailing gains and losses to larger social groups, which suggest an optimal group size determined by the heterogeneity of players and the potential gains from correlating actions.<sup>16</sup>

## 4 Correlated equilibrium, behavioral conformity and consistency with a social norm

In this Section we present our main results demonstrating that near to any Nash equilibrium there is a correlated equilibrium satisfying behavioral conformity and, if there are many players, then there are correlated equilibria that

---

<sup>16</sup>A related issue is the number of nations as modeled, for example, by Alesina and Spolaore (1997); larger countries benefit not only from greater internal efficiencies, security and ability to cope with external shocks but also may suffer from greater heterogeneity and thus a problem of 'keeping everyone happy'. Similar conditions arise in economies with clubs and/or local public goods; see, for example, the survey articles Conley and Smith (2005), Demange (2005) and Haimanko, Le Breton and Weber (2005).

are consistent with a social norm. One would expect that we need to impose some assumptions in order to obtain existence of such equilibria. A continuity condition, ensuring that players with similar attributes are similar as players in induced games, is required.<sup>17</sup> The necessity of such a condition is illustrated by Example 1 in Section 4.3. Following Wooders et. al. (2006) we introduce a Lipschitz continuity assumption:

**Continuity in attributes:** A pregame  $\mathcal{G} = (\Omega, A, h)$  satisfies *continuity in attributes* if, given any  $\varepsilon > 0$  and any two games  $\Gamma(N, \alpha)$  and  $\Gamma(N, \tilde{\alpha})$  for which  $\text{dist}(\alpha(i), \tilde{\alpha}(i)) < \varepsilon$  for all  $i \in N$ , for any  $j \in N$  and for any action profile  $\bar{a}$  it holds that  $|u_j^\alpha(\bar{a}) - u_j^{\tilde{\alpha}}(\bar{a})| < \varepsilon$ .

Continuity in attributes dictates that, given action choices, if the attribute function changes only slightly, then payoffs change only slightly.

Our first result demonstrates that, given continuity in attributes, near to any Nash equilibrium there exists a correlated equilibrium satisfying behavioral conformity. We say that correlated equilibrium  $p$  is  $\varepsilon$ -near to Nash equilibrium  $p^*$  if

$$\left| U_i^\alpha(p) - \frac{1}{|N_g|} \sum_{j \in N_g} U_j^\alpha(p^*) \right| \leq \varepsilon \quad (2)$$

for all  $i \in N_g$  and all  $N_g$ .<sup>18</sup> Thus, the expected payoff of each player, given  $p$ , is close to the average payoff of players in her group, given  $p^*$ . Note that nearness necessarily implies players within a social group have similar expected payoffs because  $|U_i^\alpha(p) - U_j^\alpha(p)| \leq 2\varepsilon$  for any  $i, j \in N_g$ . This follows from within-group anonymity and homophily.

**Theorem 1:** Consider a pregame  $\mathcal{G} = (\Omega, A, h)$  that satisfies continuity in attributes. Also consider any population  $(N, \alpha)$ , any social group structure  $\Pi$ , and any Nash equilibrium  $p^*$ . There exists a correlated  $2D_{\alpha, \Pi}$ -equilibrium  $p$  of game  $\Gamma(N, \alpha)$  such that  $p$  and  $\Pi$  satisfy behavioral conformity and  $p$  is  $D_{\alpha, \Pi}$ -near to  $p^*$ .

If a social group contains players who are substantially dissimilar, then  $D_{\alpha, \Pi}$  may be very large. Using, however, the compactness of the attribute space we can obtain existence of a correlated equilibrium satisfying behavioral conformity with a small value for  $D_{\alpha, \Pi}$ .

<sup>17</sup>See also Example 2 of Wooders et. al. (2006).

<sup>18</sup>Since a correlated equilibrium and a strategy profile are different mathematical objects, the use of  $U_j^\alpha(p^*)$  is a slight abuse of notation since  $p^* = (p_1^*, \dots, p_n^*) \in \Sigma^N$  while  $p$  is a probability distribution on  $A^N$ . But the meaning should be clear:

$$U_j^\alpha(p^*) = \sum_{\bar{a} \in A^N} (p_1^*(\bar{a}_1) \times \dots \times p_n^*(\bar{a}_n)) u_j^\alpha(\bar{a}).$$

Compare with the definition of  $U_j^\alpha(p)$ .

**Corollary 1:** Consider a pregame  $\mathcal{G} = (\Omega, A, h)$  that satisfies continuity in attributes. For any real number  $\varepsilon > 0$ , a bound  $G(\varepsilon)$  can be put on the number of social groups such that, for any population  $(N, \alpha)$ , there exists a social group structure  $\Pi$  with no more than  $G(\varepsilon)$  groups and a correlated  $\varepsilon$ -equilibrium  $p$  of induced game  $\Gamma(N, \alpha)$  that satisfy behavioral conformity.

One noteworthy point about Theorem 1 is that the social group structure can be taken as exogenous and independent of the attribute function. Specifically, Theorem 1 states that given *any social group structure*, in which players in the same social group have sufficiently similar attributes, and near to *any Nash equilibrium*, there is a correlated equilibrium and social group structure that satisfy behavioral conformity. Example 2, in Section 4.3, illustrates why this is not possible when looking for pure strategy Nash equilibria, as in Wooders et al. (2006), rather than correlated equilibria.

That it is possible to obtain an equilibrium satisfying behavioral conformity for *any* social group structure is important, beyond mere technicalities. To see why, it is easiest to consider the implications were this not to be the case. Then any change, or difference, in the nature of interaction (i.e. the game) may require a change in social group structure in order to maintain an equilibrium consistent with behavioral conformity. We have shown that correlated equilibrium allows a greater flexibility than this, whereby the population would be able to adapt to changes in the game, without changes in the structure of social groups.

A final point we note is that Corollary 1 only demonstrates the existence of an *approximate* correlated equilibrium. To see why, see Example 3 in Section 4.3.

#### 4.1 Consistency with a social norm and ex-post stability

While Theorem 1 requires behavioral conformity it does not require predictable group behavior. Thus, the equilibrium considered in the Theorem need not be ex-post stable (as formulated in Kalai 2004, for example); even if an equilibrium satisfies behavioral conformity, ex-post, once a player has observed the roles assigned to others he may have an incentive to change his action.<sup>19</sup> If so, the equilibrium is not ex-post stable and is inconsistent with the notion of a social norm. This is exemplified by Example 4 in Section 4.3.

In many contexts, ex-post instability is an unavoidable property of Nash and correlated equilibrium. When we wish to interpret the mediator as reflecting societal norms, however, then there is something slightly worrying about ex-post instability. To allow use to demonstrate that there are correlated equilibria

---

<sup>19</sup>To use a Rawlsian thought experiment one can see that within-group anonymity results in players in the same group expecting to get the same payoff. This would be an acceptable social contract under Rawls's reasoning (Rawls 1972). The criticism often made, however, of the Rawls notion of social contract (e.g. Binmore 1989) is that ex-post outcomes need be neither fair nor individually rational, leading to questions of whether such a notion represents an appropriate form of social contract.

consistent with a social norm, we impose the following continuity property from Wooders et. al. (2006).

**Global Interaction:** The pregame  $\mathcal{G} = (\Omega, A, h)$  satisfies *global interaction* when, for any  $\varepsilon > 0$ , any game  $\Gamma(N, \alpha)$  and any two action profiles  $\bar{a}$  and  $\bar{m}$ , if

$$\frac{1}{|N|} \sum_k \sum_{\omega \in \alpha(N)} |w_{\alpha, \bar{a}}(\omega, k) - w_{\alpha, \bar{m}}(\omega, k)| < \varepsilon$$

then  $|u_j^\alpha(\bar{a}) - u_j^\alpha(\bar{m})| < \varepsilon$  for any  $j \in N$  where  $\bar{a}_j = \bar{m}_j$ .

Global interaction implies that no individual player can have a significant effect on the payoff of any other player in large games. Note that this is a continuity assumption but it requires that payoffs change only slightly when, on average, actions change only slightly. When global interaction is satisfied we find that, in games with sufficiently many players, near to any Nash equilibrium there is a correlated equilibrium satisfying behavioral conformity and predictable group behavior and thus consistent with a social norm.

**Theorem 2:** Let  $\mathcal{G} = (\Omega, A, h)$  be a pregame satisfying continuity in attributes and global interaction. Consider any population  $(N, \alpha)$ , any social group structure  $\Pi$  and any Nash equilibrium  $p^*$ . For any real number  $\varepsilon > 0$  there is a real number  $\eta(\varepsilon)$  such that if  $|N| > \eta(\varepsilon)$  there exists a correlated  $2D_{\alpha, \Pi} + \varepsilon$ -equilibrium  $p$  of induced game  $\Gamma(N, \alpha)$  such that  $p$  is  $D_{\alpha, \Pi}$ -near to  $p^*$ . Moreover,  $p$  and  $\Pi$  can be chosen to satisfy both behavioral conformity and predictable group behavior and thus are consistent with a social norm.

Theorem 2 and compactness imply that in games with sufficiently many players there exists an approximate correlated equilibrium consistent with a social norm.

**Corollary 2:** Consider a pregame  $\mathcal{G} = (\Omega, A, h)$  that satisfies continuity in attributes and global interaction. For any real number  $\varepsilon > 0$  there is a real number  $\eta(\varepsilon)$  and an integer  $G(\varepsilon)$  such that, for any population  $(N, \alpha)$  where  $|N| > \eta(\varepsilon)$ , there exists a social group structure  $\Pi$ , with no more than  $G(\varepsilon)$  social groups, and a correlated  $\varepsilon$ -equilibrium  $p$  of induced game  $\Gamma(N, \alpha)$  that satisfy behavioral conformity and predictable group behavior.

Thus, in sufficiently large games there exists a correlated equilibrium in which aggregate behavior can be known ex-ante even if individual actions are not known.<sup>20</sup> Many papers have considered the properties of equilibrium in large games (Khan and Sun 2002). The paper most relevant for our purposes is Kalai (2004). Using the law of large numbers, Kalai demonstrates that in large

---

<sup>20</sup>Predictable group behavior, and Theorem 2, could be extended by relaxing predictable group behavior to require only that aggregate behavior be approximately predictable. To do so, however, would require cumbersome notation and further approximation arguments so we prefer the current predictable group behavior condition.

games all Nash equilibria are highly likely to be (approximately) ex-post stable because ex-post outcomes are likely to be similar to what was expected ex-ante. This happens despite all players choosing their actions independently. In Theorem 2 we obtain an equilibrium in which ex-post outcomes are *exactly* what was expected ex-ante; this is possible because of the correlation of actions. The correlating of actions thus eliminates any uncertainty over aggregate actions. Example 5 in Section 4.3 illustrates this. The example also illustrates that without correlation, it may be impossible to obtain an equilibrium that satisfies both within-group anonymity and predictable group behavior.

## 4.2 Subjective beliefs and stereotyping

The assumption that players know the correlating device  $p$  shall now be relaxed. Instead, players are modeled as having subjective beliefs about the device. Specifically, there is a given *set of beliefs*  $\{\beta_i\}_{i \in N}$ , where  $\beta_i$ , given by a probability distribution over the set of action profiles, denotes the beliefs of player  $i$ . Thus,  $\beta_i(\bar{a})$  denotes the probability that player  $i$  puts on players having been assigned roles according to action profile  $\bar{a}$  and  $\beta_i(\bar{a}_{-i}|\bar{a}_i)$  denotes the probability that player  $i$  puts on roles being assigned according to action profile  $\bar{a}$  given that he is assigned role  $\bar{a}_i$ . Note that this definition of beliefs can be given a more general interpretation than beliefs about the correlating device. Beliefs basically capture what players expect other players to do.

We say that the set of beliefs  $\{\beta_i\}_{i \in N}$  constitutes a *subjective correlated  $\varepsilon$ -equilibrium* if

$$\sum_{\bar{a}_{-i}} \beta_i(\bar{a}_{-i}|\bar{a}_i) u_i(\bar{a}) \geq \sum_{\bar{a}_{-i}} \beta_i(\bar{a}_{-i}|\bar{a}_i) u_i(k, \bar{a}_{-i}) - \varepsilon$$

for each  $i \in N$  and  $\bar{a}_i, k \in A$ . This revises the definition of a correlated equilibrium (as given by (1)) in the natural way by requiring no individual  $i$  to expect a payoff gain from changing strategy *given his beliefs*  $\beta_i$ .

It is well known that once subjective beliefs are allowed it becomes difficult to tie down the set of correlated equilibria (Aumann 1974, 1987; Brandenburger and Dekel 1987). A framework of social identity, however, suggests certain properties, including stereotyping, that one might expect beliefs to satisfy. We propose a definition of stereotyped beliefs in which a player expects players in the same social group to behave identically – roughly, the player expects within-group anonymity to hold. We do assume, however, that a player does not stereotype himself and this requires a slight reformulation of within-group anonymity.<sup>21</sup>

Consider permutations of an action profile  $\bar{a}$  for which player  $i$ 's action does not change. More precisely, given game  $\Gamma(N, \alpha)$ , social group structure  $\Pi$ , action profile  $\bar{a}$  and player  $i$  (and the set  $\mathcal{P}^\Pi(\bar{a})$  of action profiles that are permutations of  $\bar{a}$ ) let  $\mathcal{P}_i^\Pi(\bar{a})$  denote the subset of  $\mathcal{P}^\Pi(\bar{a})$  where  $\bar{a}'_i = \bar{a}_i$ . We can now define stereotyped beliefs.

<sup>21</sup>In earlier versions of this work we referred to this as “other-stereotyping.”

**Stereotyping:** Given population  $(N, \alpha)$ , social group structure  $\Pi$ , player  $i$  and beliefs  $\beta_i$  we say that beliefs  $\beta_i$  are *stereotyped* if  $\beta_i(\bar{a}) = \beta_i(\bar{a}')$  for any two action profiles  $\bar{a}$  and  $\bar{a}'$  where  $\bar{a}' \in \mathcal{P}_i^\Pi(\bar{a})$ .

It is a simple extension of Theorem 1 to show that a subjective correlated  $\varepsilon$ -equilibrium exists. More interesting is whether stereotyping involves costs to players. In order to judge this we need to know the actual device used by the mediator. In other words, suppose that there is a mediator that distributes roles using device  $p$  and furthermore suppose that this device is a Nash equilibrium. This device may or may not satisfy within-group anonymity but suppose that a player has stereotyped beliefs. The following result shows that a player's payoff will be approximately the same whether or not he stereotypes, and stereotyping is consistent with equilibrium. Note that this is conditional on the actual device being a Nash equilibrium but clearly if the device is not a Nash equilibrium there is no reason to expect that stereotyping would be consistent with equilibrium.

**Theorem 3:** Consider a pregame  $\mathcal{G} = (\Omega, A, h)$  that satisfies continuity in attributes. Also consider any population  $(N, \alpha)$ , any social group structure  $\Pi$  and any Nash equilibrium  $p^*$ . If each  $\beta_i$  are stereotyped then  $\{\beta_i\}_{i \in N}$  is a subjective  $2D_{\alpha, \Pi}$ -correlated equilibrium and  $|U_i^\alpha(p|\bar{a}_i) - U_i^\alpha(\beta_i|\bar{a}_i)| \leq D_{\alpha, \Pi}$  for all  $i \in N$ .

Stereotyping can therefore be consistent with equilibrium and not change the expected payoff of the player who is stereotyping.

One point worth highlighting is that while stereotyping may not change the expected payoff of the person stereotyping it can change the payoff of the stereotyped. The following example illustrates this point. This is illustrated by Example 6 in Section 4.3.

### 4.3 Examples

Our first example shows that without continuity in attributes, it may not be possible to satisfy homophily with a small number of social groups.

**Example 1:** *Homophily is consistent with an equilibrium that satisfies within-group anonymity only if the number of social groups is as large as the number of players.*

Players choose between locations  $B$  and  $C$ . The attribute space is  $[0, 1]$ . Consider population  $(N, \alpha)$  where, without loss of generality, players are ordered so that  $\alpha(i) < \alpha(i+1)$  for all  $i$ . Player 1 (the player with the 'smallest' attribute) gets a payoff 1 if he chooses location  $B$  and 0 if he chooses location  $C$ . Any other player  $i \geq 2$  gets payoff 1 if he chooses a different location to player  $i - 1$  and payoff 0 if he chooses the same location as  $i - 1$ . Clearly, the unique Nash equilibrium is one in which player 1 chooses  $B$ , player 2 chooses  $C$ , player 3 chooses  $B$ , and so on. From this it is simple to argue that, for  $\varepsilon$  small, there exists no correlated  $\varepsilon$ -equilibrium that satisfies within-group anonymity

and homophily unless the number of social groups is as large as the number of players. To see why, consider players 1 and 2. To obtain an approximate Nash equilibrium, player 1 must choose  $B$  with high probability and, given this, player 2 must choose  $C$  with high probability. If players 1 and 2 are in the same social group then within-group anonymity will clearly not hold because the mediator needs to tell player 1 to play  $B$  more often than player 2. If player 1 is in a social group containing him alone and plays  $B$ , then we can then repeat the argument with players 2 and 3, and so on. One partial solution is to put all odd numbered players in a ‘play  $B$ ’ social group and all even numbered players in a ‘play  $C$ ’ social group but this clearly does not satisfy homophily.♦

Example 1 does not satisfy continuity in attributes because a slight change in attributes can alter the ordering of players by attribute, and therefore significantly affect payoffs.

Our next Example illustrates that we cannot replace correlated equilibrium by Nash equilibrium in Theorem 1.

**Example 2:** *Correlated equilibrium is consistent with any social group structure but Nash equilibrium is not.*

Players choose between locations  $B$  and  $C$ . The attribute space is  $[0, 1]$ . Given game  $\Gamma(N, \alpha)$  let  $\bar{b}$  denote the proportion who choose  $B$  and  $\bar{c}$  the proportion who choose  $C$ . A player prefers living in a different location to the majority. More precisely, if  $\bar{b} \leq \bar{c}$  a player gets payoff 1 if he chooses  $B$  and 0 if he chooses  $C$ . Similarly, if  $\bar{b} > \bar{c}$  a player gets payoff 1 if he chooses  $C$  and 0 if he chooses  $B$ . Consequently, Nash equilibrium requires ‘half’ the players to choose  $B$  and half  $C$ . Indeed, if the number of players is even, there exists a pure strategy Nash equilibrium in which half choose  $B$  and half  $C$ . Suppose that we try to partition the attribute space  $[0, 1]$  into convex subsets such that any two players in the same subset play the same pure strategy. This is simply not possible in a way that guarantees to obtain a *Nash* equilibrium, for any possible population. For example, if all players have the same attribute, then all players must belong to the same group, and we clearly cannot have that all players play the same pure strategy while half choose  $B$  and half  $C$ . Even if the attributes of players are distributed over  $[0, 1]$  we can only determine an appropriate partition into social groups once we know the attribute function, because we need to guarantee that half the players fall into one group and half in another. If, however, we wish to obtain a *correlated* equilibrium then we have no such problems. This is because we can simply have one social group and consider the correlated equilibrium that assigns roles so as exactly half the players have role  $B$  and half have  $C$ .♦

The following example shows that we cannot take  $\varepsilon = 0$  in Corollary 1.

**Example 3:** *There need not exist a correlated 0-equilibrium consistent with behavioral conformity.*

Players choose between locations  $B$  and  $C$ . The attribute space is  $[0, 1]$ . Again, without loss of generality, consider populations  $(N, \alpha)$  where players are ordered so that  $\alpha(i) < \alpha(i+1)$  for all  $i$ . If a player of attribute  $\omega$  chooses  $B$  (or  $C$ ) then his payoff is  $|\omega - \omega'|$  where  $\omega'$  is the attribute of the ‘nearest’ player who chooses  $C$  (or  $B$ ). Furthermore, player 1 (with the ‘smallest’ attribute) gets an extra  $2|\omega - \omega'|$  from choosing  $B$  where  $\omega'$  is the attribute of player 2 (with the second ‘smallest’ attribute). Consider populations where player  $i - 1$  has the attribute nearest to that of player  $i$  for all  $i > 2$ . It can easily be checked that the unique Nash equilibrium requires player 1 choose  $B$ , player 2 choose  $C$ , player 3 choose  $B$  and so on. Following the same reasoning as used in Example 1 it can then be checked that, unless the number of social groups is the same as the number of players, there exists no correlated equilibrium satisfying behavioral conformity. Example 3 does, however, satisfy continuity in attributes and so Theorem 1 can be applied to show the existence of an *approximate* correlated equilibrium satisfying behavioral conformity. To do so the attribute space can be partitioned into convex subsets  $[0, \varepsilon]$ ,  $(\varepsilon, 2\varepsilon]$ ,  $(2\varepsilon, 3\varepsilon]$  and so on. Equating subsets of  $\Omega$  with social groups, so that  $i \in N_g$  if and only if  $\alpha(i) \in ((g-1)\varepsilon, g\varepsilon)$ , it is apparent that homophily is satisfied. Consider a correlating device  $p$  such that in any social group with at least two players there will always be at least one player who plays  $C$  and one player who plays  $B$ . This can be done in such a way as to satisfy within-group anonymity and group independence. If we let a player in a one-member social group choose an optimal strategy then we have a correlated  $\varepsilon$ -equilibrium as desired.  $\blacklozenge$

**Example 4:** Players have to choose between two locations  $B$  and  $C$ . The attribute space is given by  $\{X, R\}$  where a player with crowding type  $X$  is a celebrity and a player with crowding type  $R$  an ‘ordinary’ member of the public. We suppose that there is only one celebrity. Members of the public like living in the same location as the celebrity. Thus, the payoff of a player with attribute  $R$  is equal to 1 if he matches the choice of the celebrity and 0 otherwise. The celebrity, by contrast, prefers to avoid the public and thus his payoff is equal to the proportion of members of the public whose choice of location he mismatches. Theorem 1 applies and so we can construct a correlated equilibrium. But any correlated equilibrium of this game has the celebrity mixing between  $B$  and  $C$  and ordinary members of the population ‘in aggregate’ mixing between  $B$  and  $C$ . Ex-post, once every player has chosen a location there must be at least one player who would wish to change his location.  $\blacklozenge$

There is nothing wrong with one player constituting a social group, as in Example 4, but it is difficult to interpret the actions of the celebrity in terms of a social norm. This is primarily because social norms suggest predictability in aggregate behavior, which does not hold for this example. Example 4 makes clear that additional restrictions are required to obtain predictable group behavior.

**Example 5:** *Correlation allows consistency with a social norm without loss of payoff.*

Players have to choose between two locations  $B$  and  $C$ . There is a unique attribute. Given game  $\Gamma(N, \alpha)$  let  $\bar{b}$  denote the proportion who choose location  $B$  and let  $\bar{c}$  denote the proportion who choose  $C$ . A player's payoff is given by  $-\bar{b}$  if he chooses  $B$  and  $-\bar{c}$  if he chooses  $C$ . For simplicity assume an even number of players. There exists a pure strategy Nash equilibrium in which half of the players choose  $B$  and the other half  $C$ . There also exists a mixed strategy Nash equilibrium in which all players randomly choose between  $B$  and  $C$  with equal probability. The first equilibrium does satisfy predictable group behavior but not within-group anonymity. The second equilibrium satisfies within-group anonymity but not predictable group behavior. *There is no Nash equilibrium that satisfies both within-group anonymity and predictable group behavior. There does exist, however, a correlated equilibrium that satisfies both within-group anonymity and predictable group behavior.* This is simply that the mediator randomly picks amongst the action profiles in which half of the players play  $B$  and the other half of the players play  $C$ . Ex-ante no player would know whether he will get role  $B$  or  $C$  but they do know that, if everyone conforms, an equal number will end up choosing locations  $B$  and  $C$ .

Kalai (2004) demonstrate that in games with many players the mixed strategy Nash equilibrium is approximately ex-post stable. This means that aggregate behavior can be predicted ex-ante with some precision. Because, however, choices are made independently, it is possible that realized aggregate behavior is not as predicted. For example, all players could randomly choose location  $B$  (no matter how unlikely this is). One consequence of this is that ex-ante expected payoffs are lower with the mixed strategy Nash equilibrium than with any pure strategy Nash equilibria. Thus, to obtain a Nash equilibrium satisfying behavioral conformity payoffs need be lower than they could be. With a correlated equilibrium this is not the case, because predictable group behavior means that payoffs are as high as with any pure strategy Nash equilibrium.♦

One point to highlight from Example 5 is the necessity of correlation of actions in order to obtain an equilibrium that satisfies both within-group anonymity and predictable group behavior. While both within-group anonymity and predictable group behavior can be achieved on their own through uncorrelated actions they are only simultaneously possible with correlation. If players have desires for within-group anonymity (because of fairness) and predictable group behavior (because of ex-post stability) then this suggests that they may want to be able to correlate their actions. Of course, one may question whether correlation is possible, and we shall discuss this in more detail in the Conclusion. But, a mixed strategy equilibrium where all players randomize can be seen as one extreme with no correlation, while a device that guarantees exactly half will choose  $B$  and half will choose  $C$  can be seen as another extreme of perfect correlation. One may expect reality to lie somewhere between these two extremes.

**Example 6:** *Stereotyping changes the payoff of the stereotyped.*

Players have to choose between locations  $B$  and  $C$ . The attribute space is  $[0, 1] \times \{X, R\}$  where, as before,  $X$  denotes ‘celebrity’ and  $R$  denotes ‘ordinary member of the public’. We suppose that there is only one celebrity. Every member of the public gets payoff 1 if the celebrity chooses location  $B$  (which may afford the celebrity less privacy, for example) and 0 if the celebrity chooses location  $C$ . Clearly members of the public want the celebrity to choose  $B$ . Member  $i$  of the public has attribute  $\alpha(i) = (\omega, R)$  where  $\omega \in [0, 1]$  is a measure of his charmingness. Let  $\bar{\omega}_B$  and  $\bar{\omega}_C$  denote the average charmingness of players in locations  $B$  and  $C$ . The celebrity likes to have charming neighbors but has a slight preference for  $C$  over  $B$ ; his payoff is  $\bar{\omega}_B$  if he chooses location  $B$  and  $\bar{\omega}_C + \delta$  for some small  $\delta$  if he chooses location  $C$ .

The Nash equilibria of most interest are those where the most charming members of the public choose location  $B$  in order that the celebrity will choose location  $B$ . In this case all members of the public get payoff 1. If, however, the celebrity stereotypes then it may be that she would choose location  $C$  and all members of the public get payoff 0. To provide a specific example, suppose that player 1 is the celebrity, player 2 has charm 0.5, players 3, ...,  $n$  have charm 0.49 and  $\delta = 0.005$ . There exists a Nash equilibrium where players 1 and 2 choose location  $B$  and all others choose location  $C$ . Suppose, however, that player 1 has stereotyped beliefs. This would mean that player 1 expects one member of the public to choose location  $B$  but each member of the public is considered equally probable to be this player. The expected average charm of players in location  $B$  and  $C$  is  $0.49 + \frac{0.05}{n-1}$  and so player 1 should choose location  $C$ . ♦

In this example the celebrity is not significantly affected by the fact that she stereotypes. This is because she stereotypes players that are actually similar. That the celebrity stereotypes can, however, result in a change in incentives that may lead her to change her action. In particular, the celebrity is basically indifferent between the locations but this means her actual choice could be sensitive to stereotyping. If she changes her action this may not significantly affect her payoff but may dramatically affect the payoffs of others.

#### 4.4 The set of correlated equilibria satisfying behavioral conformity

Take as given a game  $\Gamma(N, \alpha)$  and social group structure  $\Pi$ . We know, from Theorem 1, that near to any Nash equilibrium there is an approximate correlated equilibrium that satisfies behavioral conformity. Lemma 4 suggests that a stronger result could be obtained, because it shows that near to *any approximate correlated equilibrium* there is an approximate correlated equilibrium satisfying within-group *anonymity*. In general, however, it is not possible to obtain a stronger result. This is because there can be a correlated equilibrium for which there is no nearby approximate correlated equilibrium that satisfies group independence *and* within-group anonymity. An example illustrates:

**Example 7:** *Near to a correlated equilibrium there may not be a correlated*

*equilibrium satisfying behavioral conformity.*

Players choose between locations  $B$  and  $C$  and the set of attributes is  $\{X, R\}$ . Given game  $\Gamma(N, \alpha)$  let  $\bar{b}_\omega$  denote the proportion of those with attribute  $\omega$  who choose  $B$  and  $\bar{c}_\omega$  the proportion who choose  $C$ . If a player of attribute  $X$  chooses location  $B$  his payoff is  $\bar{c}_R$  and if he chooses  $C$  his payoff is  $9\bar{b}_R$ . Similarly, if a player of attribute  $R$  chooses  $B$  his payoff is  $9\bar{c}_X$  and if he chooses  $C$  his payoff is  $\bar{b}_X$ . There exists a correlated equilibrium where there is 0.5 probability all those of attribute  $X$  choose  $C$  and those of attribute  $R$  choose  $B$ , and a 0.5 probability all those of attribute  $R$  choose  $C$  and those of attribute  $X$  choose  $B$ . The expected payoff of each player given this correlated equilibrium is 5. There is, however, no correlated equilibrium satisfying behavioral conformity that gives an expected payoff of 5 for all players. If, for example, all players are in one social group and there is an unequal number of players with attribute  $X$  and  $R$  then it would not be possible to achieve such payoffs and satisfy within-group anonymity. If there are two social groups, split according to attribute, then it would not be possible to achieve such payoffs and satisfy group independence. ♦

Example 7 suggests that the set of approximate correlated equilibria satisfying behavioral conformity could seem ‘relatively small’ compared to the set of approximate correlated equilibria. It is also the case, however, as a final example will illustrate, that near to a correlated equilibrium satisfying behavioral conformity there need not be an approximate Nash equilibrium. In this sense the set of approximate correlated equilibria satisfying behavioral conformity could seem ‘relatively large’ compared to the set of approximate Nash equilibria.

**Example 8:** *Near to a correlated equilibrium satisfying behavioral conformity there may not be a Nash equilibrium .*

Players choose between locations  $B$  and  $C$  and there is a unique attribute. Given game  $\Gamma(N, \alpha)$  let  $\bar{b}$  denote the proportion who choose  $B$  and  $\bar{c}$  the proportion who choose  $C$ . A player gets payoff  $8\bar{b} + 1$  from choosing location  $B$  and payoff  $10\bar{b}$  from choosing  $C$ , irrespective of attribute. Any Nash equilibrium of this game essentially requires half the players to play  $B$  and half to play  $C$ . The maximum Nash equilibrium payoff, therefore, if there is one social group and within-group anonymity is 5, while the maximum possible Nash equilibrium payoff is 5.5. There exists, however, a correlated equilibrium where there is a  $\frac{1}{3}$  probability that all players are given role  $B$  and a  $\frac{2}{3}$  probability that half of players are given role  $B$  and half role  $C$ . The expected payoff of each player with this correlated equilibrium is  $6\frac{1}{3}$ . Clearly this payoff is significantly higher than that obtainable with a Nash equilibrium (in the sense of equation (2)). ♦

## 4.5 Proofs

The proofs of all the Theorems follow from the same simple arguments. In this section we shall talk through these arguments and provide all proofs. Throughout the following we take as given a pregame  $\mathcal{G} = (\Omega, A, h)$  that satisfies continu-

ity in attributes. Until otherwise stated we shall also take as given a population  $(N, \alpha)$  and social group structure  $\Pi = \{N_1, \dots, N_G\}$ .

A function  $\gamma$  mapping from  $N$  to  $N$  is said to be a *permutation of players* if  $\gamma$  is one-to-one and  $\gamma(i) \in N_g$  whenever  $i \in N_g$  for all  $i \in N$ . That is we permute the labels of players within the same society. Given a permutation of players  $\gamma$  and action profile  $\bar{a}$  we denote by  $\bar{a}^\gamma$  the action profile where  $\bar{a}_i^\gamma = \bar{a}_{\gamma(i)}$  for all  $i \in N$ . With this we can make the following observation which should require no proof.

**Lemma 1:** Consider any action profile  $\bar{a}$ . If  $\bar{a}' \in P^\Pi(\bar{a})$  then there exists a (not necessarily unique) permutation of players  $\gamma$  such that  $\bar{a}' = \bar{a}^\gamma$ .<sup>22</sup> Furthermore, if  $\gamma$  is a permutation of players then  $\bar{a}^\gamma \in P^\Pi(\bar{a})$ .

Thus, if action profile  $\bar{a}'$  is a permutation of  $\bar{a}$  then for every player  $i$  there exists some player  $\gamma(i)$ , who belongs to the same social group as  $i$ , such that  $i$ , according to  $\bar{a}'$ , plays the same action that  $\gamma(i)$  plays, according to action profile  $\bar{a}$ .

The following result, which is an application of continuity in attributes, shows an approximate equivalence between a permutation of actions and a permutation of utilities. A simple illustration is provided after the proof.

**Lemma 2:** Consider any action profile  $\bar{a}$  and permutation of players  $\gamma$ ,

$$\left| u_i^\alpha(k, \bar{a}_{-i}^\gamma) - u_{\gamma(i)}^\alpha(k, \bar{a}_{-\gamma(i)}) \right| < D_{\alpha, \Pi} \quad (3)$$

for any  $k \in A$ .

**Proof:** Given the population  $(N, \alpha)$  let  $(N, \tilde{\alpha})$  be the population in which  $\tilde{\alpha}(\gamma(i)) = \alpha(i)$  for all  $i$ . That is, attribute function  $\tilde{\alpha}$  assigns to player  $\gamma(i)$  the same attribute as  $\alpha$  assigns to  $i$ . By continuity in attributes

$$\left| u_{\gamma(i)}^{\tilde{\alpha}}(k, \bar{a}_{-\gamma(i)}) - u_{\gamma(i)}^\alpha(k, \bar{a}_{-\gamma(i)}) \right| < D_{\alpha, \Pi} \quad (4)$$

for all  $i \in N$  and any  $\bar{a} \in A^N$ . We know that  $\bar{a}_i^\gamma = \bar{a}_{\gamma(i)}$  for all  $i \in N$ . The inequality (3) follows. ■

To illustrate Lemma 2 consider a population  $(N, \alpha)$  with four players,  $N = \{1, 2, 3, 4\}$  and a social group structure  $\Pi$  consisting of  $N_1 = \{1, 2, 3\}$  and  $N_2 = \{4\}$ . Consider the permutation of players  $\gamma(1) = 2$ ,  $\gamma(2) = 3$ ,  $\gamma(3) = 1$  and  $\gamma(4) = 4$ . Given action profile  $\bar{a} = (\bar{a}_1, \bar{a}_2, \bar{a}_3, \bar{a}_4)$  we obtain that  $\bar{a}^\gamma = (\bar{a}_2, \bar{a}_3, \bar{a}_1, \bar{a}_4)$ . Observe that  $\bar{a}_{-1}^\gamma = (\bar{a}_3, \bar{a}_1, \bar{a}_4)$  and  $\bar{a}_{-\gamma(1)} = (\bar{a}_1, \bar{a}_3, \bar{a}_4)$ . Lemma 2 implies that

$$\left| u_1^\alpha(k, \bar{a}_{-1}^\gamma) - u_2^\alpha(k, \bar{a}_{-2}) \right| = \left| u_1^\alpha(k, (\bar{a}_3, \bar{a}_1, \bar{a}_4)) - u_2^\alpha(k, (\bar{a}_1, \bar{a}_3, \bar{a}_4)) \right| < D_{\alpha, \Pi}.$$

<sup>22</sup>That is,  $\bar{a}_i = \bar{a}_i^\gamma$  for all  $i \in N$ .

Note that there are two components to this inequality: (i) it compares the payoff of player 1 to that of player 2, when they play the same action, and (ii) the identity of the players in the social group playing each action has changed. This illustrates how players with a similar attribute need have both similar utility functions and effect others utility in a similar way.

Lemma 2 concerns a permutation of actions. The next step is to take this to a permutation of strategies and permuted correlating device. Given a correlating device  $p$  and permutation of players  $\gamma$  we denote by  $p^\gamma$  the correlating device where  $p^\gamma(\bar{a}^\gamma) = p(\bar{a})$  for all  $\bar{a} \in A^N$ . One way to interpret the device  $p^\gamma$  is that it randomly determines assigns roles  $\bar{a}$  according to the correlating device  $p$  but then assigns player  $i$  the role of player  $\gamma(i)$ , that is, it gives player  $i$  role  $\bar{a}_{\gamma(i)}$  instead of player  $\gamma(i)$ . This is a generalization of permuting strategies. The following result follows easily from Lemma 2 and shows that a permutation of roles leads to an approximate permutation of expected payoffs.<sup>23</sup>

**Lemma 3:** Let  $p$  be any correlating device,  $\gamma$  any permutation of players and  $i$  any player,

$$\left| \sum_{\bar{a}_{-i}} p^\gamma(\bar{a}_{-i} | \bar{a}_i^\gamma) u_i^\alpha(k, \bar{a}_{-i}) - \sum_{\bar{a}_{-\gamma(i)}} p(\bar{a}_{-\gamma(i)} | \bar{a}_{\gamma(i)}) u_{\gamma(i)}^\alpha(k, \bar{a}_{-\gamma(i)}) \right| < D_{\alpha, \Pi} \quad (5)$$

for any  $k \in A$  and  $\bar{a}$ .

**Proof:** Given that  $p^\gamma(\bar{a}^\gamma) = p(\bar{a})$  and  $\bar{a}_i^\gamma = \bar{a}_{\gamma(i)}$  for all  $\bar{a}$  and  $i$  we have that  $p^\gamma(\bar{a}_{-i}^\gamma | \bar{a}_i^\gamma) = p(\bar{a}_{-\gamma(i)} | \bar{a}_{\gamma(i)})$  for all  $\bar{a}$  and  $i$ . Thus,  $\sum_{\bar{a}_{-i}} p^\gamma(\bar{a}_{-i} | \bar{a}_i^\gamma) u_i^\alpha(k, \bar{a}_{-i})$  can be rewritten  $\sum_{\bar{a}_{-\gamma(i)}} p(\bar{a}_{-\gamma(i)} | \bar{a}_{\gamma(i)}) u_i^\alpha(k, \bar{a}_{-\gamma(i)})$ . Equation (5) can, therefore, be restated

$$\left| \sum_{\bar{a}_{-\gamma(i)}} p(\bar{a}_{-\gamma(i)} | \bar{a}_{\gamma(i)}) \left[ u_i^\alpha(k, \bar{a}_{-i}^\gamma) - u_{\gamma(i)}^\alpha(k, \bar{a}_{-\gamma(i)}) \right] \right| < D_{\alpha, \Pi}$$

and so applying Lemma 2 gives the desired result. ■

Now consider a correlated equilibrium  $p^*$  of game  $\Gamma(N, \alpha)$  and suppose that we permute the strategies of players. Specifically, let  $\gamma$  be a permutation of players and consider correlating device  $p^{*\gamma}$ . Lemma 3 implies that if player  $\gamma(i)$  had no incentive to deviate from his assigned role given correlating device  $p^*$  then player  $i$  could gain at most  $2D_{\alpha, \Pi}$  from deviating from his assigned role given correlating device  $p^{*\gamma}$ . This leads to the following result.

**Lemma 4:** Let  $p^*$  be any correlated  $\varepsilon$ -equilibrium of game  $\Gamma(N, \alpha)$  and let  $\gamma$  be any permutation of players. Correlating device  $p^{*\gamma}$  is a correlated  $2D_{\alpha, \Pi} + \varepsilon$ -equilibrium of game  $\Gamma(N, \alpha)$ .

<sup>23</sup>For example, Lemma 3 implies that  $\left| U_i^\alpha(p^\gamma | \bar{a}_i^\gamma) - U_{\gamma(i)}^\alpha(p | \bar{a}_{\gamma(i)}) \right| < \delta$ .

**Proof:** If  $p^*$  is a correlated  $\varepsilon$ -equilibrium then

$$\sum_{\bar{a}_{-\gamma(i)}} p^*(\bar{a}_{-\gamma(i)}|\bar{a}_{\gamma(i)})u_{\gamma(i)}^\alpha(\bar{a}) \geq \sum_{\bar{a}_{-\gamma(i)}} p^*(\bar{a}_{-\gamma(i)}|\bar{a}_{\gamma(i)})u_{\gamma(i)}^\alpha(k, \bar{a}_{-\gamma(i)}) - \varepsilon$$

for all  $i \in N$ ,  $k \in A$  and  $\bar{a}_{\gamma(i)}$ . Applying Lemma 3 implies that

$$\sum_{\bar{a}_{-i}} p^{*\gamma}(\bar{a}_{-i}|\bar{a}_i^\gamma)u_i^\alpha(\bar{a}) \geq \sum_{\bar{a}_{-i}} p^{*\gamma}(\bar{a}_{-i}|\bar{a}_i^\gamma)u_i^\alpha(k, \bar{a}_{-i}) - 2D_{\alpha, \Pi} - \varepsilon$$

for all  $i \in N$ ,  $k \in A$  and  $\bar{a}_i^\gamma$  as desired. ■

Thus, given a correlated equilibrium (or Nash equilibrium) we can permute players and obtain an approximate correlated (or Nash) equilibrium. With this we are basically done. All that remains is to determine the social group structure and obtain existence of a correlated equilibrium. It should be clear, however, that we could consider *any* social group structure in which players have sufficiently similar attributes and apply Lemma 4. Also, the existence of a correlated equilibrium is well known in games with a finite strategy set.

**Proof of Theorem 1:** Let  $\Lambda$  denote the set of permutations of players (consistent with  $\Pi$ ). By Lemma 4 we know that each  $p^{*\gamma}$  is a correlated  $2D_{\alpha, \Pi}$ -equilibrium. Consider correlating device  $p'$  given by

$$p'(\bar{a}) = \frac{1}{|\Lambda|} \sum_{\gamma \in \Lambda} p^{*\gamma}(\bar{a}). \quad (6)$$

It is well known that the set of correlated equilibria is convex. Extending such results to approximate correlated equilibria is simple. It follows that  $p'$  is a correlated  $2D_{\alpha, \Pi}$ -equilibrium. Also,  $p'$  satisfies within-group anonymity by construction and roles are distributed independently across players by device  $p^*$  and in every  $p^{*\gamma}$  which means that  $p'$  must satisfy group independence. Finally, Lemma 3 gives relation (2). ■

**Proof of Theorem 2:** For any  $\varepsilon > 0$ , Theorem 1 of Wooders et. al. (2006) demonstrates that in games with sufficiently many players there exists a Nash  $\varepsilon$ -equilibrium in pure strategies  $p^*$ . That  $p^*$  is an equilibrium in pure strategies implies  $p(\bar{a}^*) = 1$  for some action profile  $\bar{a}^*$ . Clearly  $p^*$  satisfies predictable group behavior. As in the proof of Theorem 1 let  $\Lambda$  denote the set of permutations of players (consistent with  $\Pi$ ) and let  $p'$  be defined as in (6). We can use the same arguments as in the proof of Theorem 1 to see that  $p'$  is a correlated  $\varepsilon$ -equilibrium satisfying behavioral conformity. It should be clear that the device  $p'$  also satisfies predictable group behavior. ■

**Proof of Theorem 3:** Fix a player  $i \in N$ . Let  $\Lambda_i$  denote the set of permutations of players  $\gamma$  in which  $i$  is not permuted, that is,  $\gamma(i) = i$ . Define beliefs  $\beta_i$  where

$$\beta_i(\bar{a}) = \frac{1}{|\Lambda_i|} \sum_{\gamma \in \Lambda_i} p^{*\gamma}(\bar{a}). \quad (7)$$

Beliefs  $\beta_i$  are stereotyped and following Lemma 3 and the logic of proof of Lemma 4,  $|U_i^\alpha(p^*|\bar{a}_i) - U_i^\alpha(\beta_i|\bar{a}_i)| \leq D_{\alpha,\Pi}$ . Further, constructing  $\beta_i$  for each  $i \in N$  we obtain a subjective  $2D_{\alpha,\Pi}$ -correlated equilibrium  $\{\beta_i\}_{i \in N}$ . ■

## 5 Concluding remarks

This paper models conformity and social norms in settings where different players can perform different actions but can still be seen as conforming to the same norm. We argued that correlated equilibrium is an appealing way to model such conformity. In doing so we proposed conditions one would want to impose on the nature of correlation – within-group anonymity, group independence and predictable group behavior– and have demonstrated the existence of a correlated equilibrium satisfying these properties. One way to interpret this result is that social interaction can act as a form of equilibrium-selection device that selects correlated equilibria satisfying certain properties.

One obvious question is: “If we recognize that there is no formal device telling people what to do, where does the correlation come from?”. It was not our intention in this paper to answer that question and so we have been quiet on this issue but now we will make a few remarks. First, in some social contexts there may be someone who indeed does tell people what to do and can directly correlate actions. Second, it has been observed that, in the absence of a formal device, correlation of actions can emerge spontaneously within groups (e.g. Schelling 1960; Hayek 1982; Sugden 1989; Van Huyck et al. 1997; Hargreaves-Heap and Varoufakis 2002). This can be achieved through conditioning actions on random signals such as gender, age, exam results etc. On a more theoretical level, Hart and Mas-Colell have shown how naive learning heuristics such as regret matching can lead to aggregate play corresponding to a correlated equilibrium (see Hart 2005). The approach of Hart and Mas-Colell is framed in a myopic setting in which correlation arises without any social context or social influence. It may be interesting to ask how learning dynamics would change if an element of social context, such as desires for within-group fairness, exists. This all suggests that correlation of actions within social groups is not unrealistic. In particular, while the correlation required by predictable group behavior or within-group anonymity may be asking too much it may be possible for social groups to obtain correlated equilibrium that approximates predictable group behavior and within-group anonymity. Example 5, stylized as it is, suggests why people would want to correlate actions, namely, that in doing so preferable outcomes can be obtained.

The possibility of subjective beliefs and stereotyping suggests an alternative interpretation of our results. If beliefs are subjective and stereotyped then there need not be any correlation of actions but just a belief that there is correlation. The focus, therefore, shifts from how actions could be correlated to whether it can be consistent with equilibrium for players to expect correlation even if there is none. We demonstrated that stereotyping, even if it causes erroneous beliefs, can be consistent with equilibrium. Furthermore, a player’s payoff is largely

invariant to whether he stereotypes. Stereotyping can, however, influence the payoffs of those being stereotyped. It should be emphasized that we obtain this result because a player only stereotypes members of groups consisting of similar players. This raises the question of how a player would form his beliefs about the actions expected of others.

At this point it may be worth relating our work to the concept of *rule rationality*, described by Aumann (2008). Rule rationality postulates that individuals follow rules and, although a rule is not necessarily optimal (or rational in the usual game-theoretic meaning of the term) in some situations in which an individual may find himself, the individual is well served by the rule *on average*. We might think of stereotyping and obeying a social norm as two rules that individuals may follow. Rule rationality does motivate the use of *approximate* equilibria with players being willing to sacrifice  $\varepsilon$  in any one game on the basis that in the long run the decision making costs of recovering this  $\varepsilon$  are not worthwhile. Beyond this, rule rationality is interesting when addressing the need for predictable group behavior and ex-post stability. More specifically, if there is not ex-post stability then it can be questioned whether individuals would conform over time (and this is why we looked for predictable group behavior), but a rule to obey the social norm may address this problem. Return again to Sugden's example of Yorkshire beachcombers. In the short run, it may have been beneficial for an individual who finds a pile of driftwood already claimed, with two rocks on top of the pile, to simply take the pile of driftwood. The established social norm may, however, be rule-rational for members of a coastal village. Sometimes an individual may find himself faring better than some individuals and sometimes he may fare worse, but on average, the social norm is fair. Interestingly, for this to work it is important that all members of the social group have the same rule, emphasizing that it is important to take rule rationality beyond the individual decision-making context emphasized by Aumann (2008).

One way to address some of these issues would be to make the role-assignment device endogenous, that is, to model how players can endogenously develop a coordinated way of recognizing and interpreting random signals from nature or pre-play communication. An endogenous role-assignment device would enable one to determine from the model whether correlation and behavioral conformity can be expected to emerge as properties. In doing so one would also like social groups to be endogenous. It may be possible to address this as a coalition-formation problem either in a noncooperative/cooperative framework such as in Perry and Reny (1996), or more recent work on economies with local public goods or many-to-any matching problems, such as Konishi and Unver (2006), or through a network approach similar to those described in Jackson (2005). Alternatively, evolutionary arguments as in Robson and Wooders (1997) may lead to the selection of social norms based on population growth. A related issue is to consider communication equilibrium as opposed to correlated equilibrium (Forges 1987). Communication equilibrium is the extension of correlated equilibrium to games in extensive form where communication and signals are possible, not only prior to play but also during play of the game. In endogeniz-

ing the assignment device and social group membership, it would be natural to model more explicitly the process of communication between players, not only before the game but during the play of the game (or, if thinking of repeated plays of a stage game, between plays of the stage game).

At this point, it may be worth remarking that our results are obtained under quite weak assumptions on the games. The pregame framework simply allows us to talk about similar players and to consider games that are close to each other (games with slightly different preferences of players). It may be interesting to ask whether further characterization results could be obtained in more restrictive environments. For example, environments with “intermediate preferences” may yield more insights, as in Carmona (2009).<sup>24</sup>

Finally, we conclude by relating this paper to our prior working papers, especially those dealing with conformity and stereotyping. In Cartwright and Wooders (2003) we raised the question of whether we could meaningfully extend the results of Wooders et. al. (2006) to situations where individuals in the same society could undertake different actions. In that paper, we treated these questions in the context of games with many players, as (in part) in this paper (in particular, the Lipschitz continuity condition and global interaction were both used). In an effort to simplify the results and bring into sharp focus the effects of most players having many close substitutes we took a different tack in Cartwright and Wooders (2005). In that paper we also introduced stereotyping and the question of whether stereotyping of others was harmful to an individual player (in other words, consistent with bounded rationality). These papers were widely presented and we have benefited from comments of participants in numerous conferences and seminars. The clarity and simplicity of our current paper is largely due to our prior work taking different approaches to make the same points. In the current paper, besides sharpening some of the prior results, we return to games with many players and, for the first time, introduce the property of predictable group behavior. What other natural behavioral properties of strategic games with many players can be obtained is an open question.

## References

- [1] Akerlof, G. A., (1980) “A theory of social custom of which unemployment may be one consequence,” *Quarterly Journal of Economics* 94: 749-75.
- [2] Akerlof, G. and R. Kranton, (2000) “Economics and identity,” *The Quarterly Journal of Economics* 115: 715-753.
- [3] Alesina, A. and E. Spolaore, (1997) “On the number and size of nations,” *The Quarterly Journal of Economics* 112: 1027-1056.

---

<sup>24</sup>Intermediate preferences were introduced by Grandmont (1978). See also Demange (1994) for another application and Wooders et.al (2006), which uses a stronger form of intermediate preferences than Carmona.

- [4] Aumann, R.J. (2008) "Rule rationality versus act rationality," Hebrew University of Jerusalem Center for the Study of Rationality Discussion Paper #497.
- [5] Aumann R.J. (1974) "Subjectivity and correlation in randomized strategies," *Journal of Mathematical Economics* 1, 67-96.
- [6] Aumann, R.J. (1987) "Correlated equilibrium as an expression of Bayesian rationality," *Econometrica* 55: 1-18.
- [7] Bernheim, B. D. (1994) "A theory of conformity," *The Journal of Political Economy*, Vol. 102, No. 5. (Oct., 1994), pp. 841-877.
- [8] Binmore, K., (1989) "Social contract I: Harsanyi and Rawls," *The Economic Journal* C: 84-102.
- [9] Brandenburger, A. and E. Dekel (1987) "Rationalizability and correlated equilibria," *Econometrica* 55:1391-1402.
- [10] Brown, R. (2000) "Social identity theory: past achievements, current problems and future challenges," *European Journal of Social Psychology* 30: 745-778.
- [11] Cartwright, E. (2003) "Bounded rationality, social conformity and large games," University of Warwick Ph. D. thesis.
- [12] Cartwright, E. and M. Wooders (2005) "Correlated equilibrium and behavioural conformity," The Warwick Economics Research Paper Series (TWERPS) 732.
- [13] Cartwright, E. and M. Wooders (2009) "On equilibrium in pure strategies in games with many players", *International Journal of Game Theory* 38: 137-153.
- [14] Conley, J. and S. Smith (2005) "Coalitions and clubs; Tiebout equilibrium in large economics," in *Group Formation in Economics; Networks, Clubs and Coalitions*, G. Demange and M. Wooders (eds.), Cambridge University Press, Cambridge UK.
- [15] Conley, J. and M. Wooders (2001) "Tiebout economics with differential genetic types and endogenously chosen crowding characteristics," *Journal of Economic Theory* 98: 261-294.
- [16] Currarini S., M. Jackson and P. Pin (2008) "An economic model of friendship: homophily, minorities and segregation," working paper.
- [17] Demange, G. (2005) "Group formation; The interaction of increasing returns and preference formation," in *Group Formation in Economics; Networks, Clubs and Coalitions*, G. Demange and M. Wooders (eds.), Cambridge University Press, Cambridge UK.

- [18] Demange G. (1994) "Intermediate preferences and stable coalition structures," *Journal of Mathematical Economics*, 1994, 45-48.
- [19] Dhillon A., and J. F. Mertens, (1996) "Perfect correlated equilibria," *Journal of Economic Theory* 68: 279-302.
- [20] Elster, J. (1989) "Social norms and economic theory," *Journal of Economic Perspectives* 3: 99-117.
- [21] Forges, F. ( 1986) "An approach to communication equilibria," *Econometrica* 54: 1375-1385.
- [22] Fudenberg, D. and J. Tirole (1998) *Game Theory*, MIT Press.
- [23] Friedman, D. (1996) "Equilibrium in evolutionary games: Some experimental results," *Economic Journal* 106: 1-25.
- [24] Grandmont, J. (1978) "Intermediate preference and the majority rule," *Econometrica* 46, 317-330.
- [25] Gravel, N. and S. Thoron (2007) "Does endogenous formation of jurisdictions lead to wealth-stratification?" *Journal of Economic Theory* 132: 569-583
- [26] Haimanko, O., M. Le Breton and S. Weber (2004) "Voluntary formation of communities for the provision of public projects," *Journal of Economic Theory* 115: 1-34.
- [27] Hayek, F.A., (1960) *The Constitution of Liberty* Routledge and Keegan Paul Ltd. London.
- [28] Hayek, F.A., (1982) *Law, Legislation and Liberty*, T.J. Press, Padstow.
- [29] Hargreaves-Heap, S. and Y. Varoufakis (2002) "Some experimental evidence on the evolution of discrimination, cooperation and perceptions of fairness," *The Economic Journal*, 112: 679-703.
- [30] Harsanyi, J.C. (1973) "Games with randomly distributed payoffs: a new rationale for mixed strategy equilibrium points." *International Journal of Game Theory* 2: 1-23.
- [31] Hart, S., (2005) "Adaptive heuristics," *Econometrica* 73: 1401-1430.
- [32] Hogg, M., D. Terry and K. White, (1995) "A tale of two theories: a critical comparison of Identity Theory with Social Identity Theory," *Social Psychology Quarterly* 58: 255-269.
- [33] Hogg, M.D., and G.M. Vaughan, (2005) *Social Psychology*, Pearson/Prentice Hall, Marlow.

- [34] Jackson, M. (2005) "A survey of models of network formation: Stability and efficiency," Chapter 1 in *Group Formation in Economics; Networks, Clubs and Coalitions*, edited by G. Demange and M. Wooders, Cambridge University Press, Cambridge U.K.
- [35] Johnson, D.W., and F. Johnson, (1987) *Joining Together: Group Theory and Group Skills*, Prentice Hall, London.
- [36] Khan, A. and Y. Sun (2002) "Noncooperative games with many players," *Handbook of Game Theory*, R. Aumann and S. Hart, eds. North Holland.
- [37] Konishi, and M. Utku Unver (2006), "Credible group-stability in many-to-many matching problems," *Journal of Economic Theory* 129: 57-80.
- [38] Kalai, E. (2004) "Large robust games," *Econometrica* 72: 1631-1666.
- [39] Myerson, R.B. (1986) "Acceptable and predominant correlated equilibria," *International Journal of Game Theory* 15: 133-154.
- [40] Rabin, M. (1993) "Incorporating fairness into game theory and economics," *American Economic Review* 83: 1281-1302.
- [41] Rapoport, A, D. Seale and E. Winter (2002) "Coordination and learning behavior in large groups with asymmetric players," *Games and Economic Behavior* 39, 111-136
- [42] Rawls, J., (1972) *A Theory of Justice* Oxford University Press.
- [43] Robson, A. and M. Wooders (1997) "The growth-maximizing distribution of income," *International Economic Review* 38, 511-526.
- [44] Sherif, M. (1966) *In Common Predicament: Social Psychology of Intergroup Conflict and Cooperation* Houghton Mifflin: Boston.
- [45] Schelling, T. C. (1960) *The Strategy of Conflict*. Cambridge, Harvard University Press, 1960.
- [46] Selten, R. (1980) "A note on evolutionary stable strategies in asymmetric animal conflicts", *Journal of Theoretical Biology* 84: 93-101.
- [47] Sugden, R. (1989) "Spontaneous order," *Journal of Economic Perspectives* 3, # 4, 85-97.
- [48] Tajfel, H. (1978) *Differentiation Between Social Groups: Studies in the Social Psychology of Intergroup Relations*, Academic Press, London.
- [49] Van Huyck, J., R.C. Battalio and F.W. Rankin (1997) "On the origin of convention: evidence from coordination games," *Economic Journal*, 107: 576-597.
- [50] Wooders, M., E. Cartwright and R. Selten (2006) "Behavioral conformity in games with many players," *Games and Economic Behavior*, 57: 347-360