BEHAVIORAL PROPERTIES OF CORRELATED EQUILIBRIUM; SOCIAL GROUP STRUCTURES WITH CONFORMITY AND STEREOTYPING

by

Edward Cartwright and Myrna Wooders

Behavioral properties of correlated equilibrium; Social group structures with conformity and stereotyping *

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; June 2008

Abstract: We explore the potential for correlated equilibrium to 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 social stability. These are within-group anonymity (conformity of behavior within groups), group independence (no conformity between groups), and predictable social group behavior (stability). As an expression of bounded rationality in the presence of social group structures 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 (a) correlated equilibrium required only to satisfy the two properties of within-group anonymity and group independence exists. If, in addition, there are many players then (b) a correlated equilibrium satisfying all three of the above properties exists and, with probability one, is ex-post stable (c) a player who stereotypes other players cannot do better with correct beliefs and thus stereotyping is not costly to the player who stereotypes.

1 Introduction

Individuals belonging to the same society or social group typically share and conform to common social and behavioral norms and customs. This motivates the question of whether behavioral conformity can be consistent with self-interested behavior. In Wooders, Cartwright and Selten (2006), henceforth WCS, we argued that this consistency requires the existence of an approximate Nash equilibrium that induces a partition of players into 'relatively few' societies, where individuals within the same society are similar and play the same or similar strategies. The Nash equilibrium captures a notion of self interested behavior while the limit on the number of societies captures behavioral conformity by forcing large groups of players to conform. WCS provide a family of games where the desired equilibrium exists. In this paper we follow the approach of WCS but with a different notion of 'play the same or similar strategies' and hence a different notion of behavioral conformity. In doing so we provide results complementary to those of WCS and, in particular, demonstrate that behavioral conformity can be consistent with Nash equilibrium for a wider class of games than suggested by WCS.

But what does it mean to 'play the same strategy'? WCS equate playing the same strategy with taking the same action. Often, however, we may think of individuals as conforming to the same norm of behavior even if they perform different actions. For example, if a husband has a paid job and a wife does housework then, while they perform different actions, we can think of them as conforming to the same norm. Similarly, if two cars meet on a narrow road and the first to arrive does not give way but the second to arrive does give way then the two drivers perform different actions but may still be conforming to the same norm of behavior. That individuals can coordinate their behavior to mutual advantage even if doing so requires different individuals to perform different actions and receive inequitable rewards is well documented (e.g. Schelling 1960, Hayek 1982, Sugden 1989, Friedman 1996, Van Huyck et al. 1997, Rapaport, Seale, and Winter 2001, Hargreaves-Heap and Varoufakis 2002). It has also been long recognized that conformity, often subconscious, to established rules and norms of behavior facilitates such coordination of behavior (Hayek 1960, 1982, Sherif 1966, Tajfel 1978, Johnson and Johnson 1987, Akerlof and Kranton 2000, Brown 2000).[1]

Our objective in this paper is to investigate a broader class of behavioral properties of equilibrium and to allow players within social groups to conform to social norms or standards of behavior but still take different actions. This requires an appropriate conception of conformity. To capture those aspects of conformity that we view as fundamental to conformity to social norms, we use the concept of correlated equilibrium. In contrast to Nash equilibrium, correlated equilibrium allows player actions to be statistically dependent on some random event external to the model (Aumann 1974, 1987). This could be, for example, whether the player is male or female, first to arrive or second

---

[1] The economic literature on conformity includes Akerlof (1980), Elster (1989) and Bernheim (1994).

to arrive at a road junction. More generally, we can imagine a mediator (or some device) that instructs players to take actions according to some commonly known probability distribution. In this paper we think of the mediator as distributing roles. If it is in the interests of each player to assume the role assigned to him by the mediator, then the probability distribution over roles is a correlated equilibrium (Aumann 1987, Forges 1986, Dhillon and Mertens 1996).[2]

An appealing interpretation of the concept of correlated equilibrium is that every player is 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 but potentially perform different actions. Because actions are conditioned on signals or roles, correlated equilibrium also recognizes how conformity can lead to coordinated actions within social groups (Johnson and Johnson 1977, Selten 1980, Sugden 1989, Hogg et al. 1995). For example, it is no accident that one car will give way and one will not if both drivers follow a 'second to arrive gives way norm'. Care is needed, however, in modelling behavioral conformity with correlated equilibrium. This is evident by the fact that merely using the concept of correlated equilibrium has created a setting where every player can use the same strategy. We need, therefore, to be sure that 'play the same strategy' equates with 'behaves in the same way' or 'conforms'. To do this we need to impose conditions on how the mediator distributes roles in order to guarantee that a correlated equilibrium can be interpreted as consistent with behavioral conformity. We shall consider three conditions: within group anonymity, group independence and predictable group behavior.

Informally, *within-group anonymity* (WGA) requires that any two individuals within the same group have the same probability of being allocated 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 roles, any two individuals in the same group are expected to behave in the same way. For example, as two cars drive along the road it may be a 50-50 chance which arrives at a narrow section of road first and will thus is assigned the role "do not give way." This 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.[3] *Group independence* (GI) requires that the distribution of roles is statistically independent between different groups and thus rules out any correlation of actions across groups. Correlation between social groups is typically unlikely (Tajfel 1978, Hogg and Vaughan 2005).[4] It seems

---

[2]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, it 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.

[3]The importance of fairness and within group equity is well known (Johnson and Johnson 1987, Tajfel 1978, Rabin 1993, Fehr and Schmidt 1999 and Brown 2000).

[4]Ruling out correlation between groups does not rule out coordination between groups.

4

therefore important to rule this out. WGA and GI combined mean that we can think of individuals within the same group as behaving in the same way while individuals in different groups may behave in different ways. This seems to take us some way to capturing behavioral conformity.

Our first main result (Theorem 1), similar to Theorem 2 of WCS, provides a family of games for which there exists a correlated equilibrium satisfying WGA and GI and which induces a partition of the player set into a relatively small number of social groups. As in WCS we require continuity with respect to player attributes. We do not however, unlike WCS, require continuity with respect to aggregate actions or, for this first Theorem, require there to be many players. This means that our result holds in contexts where one individual player can have a big influence on others and significantly generalizes the set of games for which our result applies. Behavioral conformity can, therefore, be consistent with Nash equilibrium much more generally when we use a notion of behavioral conformity that allows different actions within social groups.

The correlated equilibrium obtained in Theorem 1 need not be ex-post stable in the sense that once an individual knows the roles of others he may wish to change his action (see, for example, Kalai 2004). There is nothing necessarily wrong with this and it is a property typical of games of incomplete information. If, however, we see a social group as conforming to a norm then we may expect that the aggregate behavior of that group should be 'predictable' in some way. Also, if the ex-post outcome of conformity is to be seen as fair (and consistent with members of a social group not deviating once their assigned roles are known) then we may expect that no individual should wish they had chosen some other action. Ex-post stability gives both of these properties. We therefore introduce a third property called *predictable group behavior property* (PGB) which dictates that the number of players in each social group who will play each action be known ex-ante. This property guarantees ex-post stability. Our second main result (Theorem 2), demonstrates that, with a Lipschitz continuity condition on utility functions, for games with sufficiently many players there is a correlated equilibrium that satisfies PGB, WGA and GI and which induces a partition of the player set into a relatively small number of social groups. In interpretation, this implies that when there are many players it is possible for correlated equilibrium to be consistent with a relative small number of social norms (possibly different in different social groups) satisfying fairness within social groups, no correlation between social groups, and ex-post stability.

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 etc. (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 consider *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. As we have previously noted, stereotyping may be a form of bounded rationality and allow for 'simpler' correlation devices. We demonstrate (Theorem 3) that stereotyping can be consistent with correlated equilibria. 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.[5]

We proceed as follows: Section 2 introduces the model and properties of social groups, Section 3 provides the main results and Section 4 concludes.

2 Model and notation

A *game* Γ is given by a triple $(N, A, \{u_i\}_{i \in N})$ consisting of a finite *player set* $N = \{1, .., n\}$, a finite set of K *actions* $A = \{1, ..., K\}$, and a set of *payoff functions* $\{u_i\}_{i \in N}$. An *action profile* consists of a vector $\overline{a} = (\overline{a}_1, ..., \overline{a}_n)$ where $\overline{a}_i \in A$ 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 $\Gamma(N, \alpha)$ is given by a randomization σ 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 $\overline{\sigma} = (\overline{\sigma}_1, ..., \overline{\sigma}_n)$ where $\overline{\sigma}_i \in \Sigma$ denotes the *strategy of player* i. We assume von-Neumann Morgenstern expected utility functions and with, a slight abuse of notation, denote by $u_i(\overline{\sigma})$ the expected payoff to player i given strategy profile $\overline{\sigma}$. Strategy profile $\overline{\sigma}$ is a *Nash ε-equilibrium* for some real number $\varepsilon \geq 0$ if

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

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

2.1 Pregames

Following WCS we make use of a non-cooperative pregame, which allows us to consider families of games derived from a common underlying structure.[6] Informally, a non-cooperative pregame consists of a set of player attributes or characteristics, a set of actions, and a preference function. Given a finite set N of players, a game is induced by ascribing a point in attribute space to each

---

[5]Earlier versions of our results on stereotyping were presented at the 2004 Stony Brook International Conference on Game Theory and at Hebrew University in March 2006. We thank participants for their comments, especially Francois Forges, Peyton Young and Sergiu Hart.

[6]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 endowments. An economy is then determined by a set of economic agents and a function ascribing a preference relation and an endowment 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.

player. The preference function is used to ascribe utility functions to players in any induced game.

A *non-cooperative pregame* is a triple $\mathcal{G} = (\Omega, A, h)$ consisting of a compact metric space Ω, called an *attribute space*, a set of actions A and a *preference function h*. In order to explain the preference function h we must first define the notion of a weight function. Let W be the set of all mappings from $\Omega \times A$ into $\mathbb{R}_+$ with finite support. A member of W is a *weight function*. A *preference function h* is a mapping from $\Omega \times A \times W$ into the set of non-negative real numbers $\mathbb{R}_+$. As we formalize below, in interpretation $h(\omega, k, w)$ is the payoff to a player of attribute ω if he plays action k and the actions of other players are summarized by weight function w.

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

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

for all $k \in A$ and all $\omega \in \Omega$. Thus, $w_{\alpha, \overline{a}}(\omega, k)$ denotes the number of players with attribute ω who play action k in the strategy $\overline{a}$. An *induced game* $\Gamma(N, \alpha)$ can now be defined:

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

where

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

for all $\omega \in \alpha(N)$. We note that players who are ascribed the same attribute have the same payoff function.

As discussed by WCS a pregame need not imply any assumptions on the induced games. A pregame does provide, however, a useful framework in which to treat a family of games all induced from a common strategic setting, and to be able, relatively simply, to impose assumptions on that family of games through assumptions on the function h.

2.2   The mediator

Given an induced game $\Gamma(N, \alpha)$ we think of a *mediator* who signals a suggested action to each player. The mediator is represented by a *correlating device p*. Once a player observes his own signal, without observing the signals sent to others, he chooses an action. To distinguish suggested action from actual behavior we equate a signal with an *assignment to a role*. Thus, the correlating device p assigns each player a role from set A and the player then chooses an action from set A that may or may be consistent with his assigned role. In interpretation we shall equate the mediator with 'society' that prescribes actions to players according to societal norms.

Given that the correlating device assigns a role to each player we can formally think of the correlating device as being given by a *probability distribution p over action profiles* where $p(\overline{a})$ denotes the probability that players will be assigned roles consistent with action profile $\overline{a}$. We shall denote by $p(\overline{a}_{-i}|\overline{a}_i)$ the probability of role assignments being consistent with $\overline{a}$, conditional on player i having role $\overline{a}_i$. 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.[7] Let P denote the set of possible correlating devices.

The mediator and correlating device p transform game $\Gamma(N, \alpha)$ into what we call a *game with roles,* denoted by $\Gamma^p(N, \alpha)$. In game $\Gamma^p(N, \alpha)$, action choice can be made conditional on assigned role. A *behavioral rule* in game $\Gamma^p(N, \alpha)$ is a function b 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 profile* $\overline{b}^c$ where $\overline{b}_i^c(k) = k$ for all k and i, that is, the behavioral profile where each player plays the action consistent with his assigned role. Note that, although his payoff may indirectly depend on roles through the choice of action that a distribution of roles induce, a player's payoff does not directly depend on his role or the roles of other players.

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 3.2. Given correlating device p, we can now define a payoff function $U_i^\alpha : P \to \mathbb{R}$ for each player $i \in N$, where

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

denotes the expected payoff of player i if roles are assigned according to distribution p and players follow the conformist behavioral profile $\overline{b}^c$.[8] Let

$$U_i^\alpha(p|\overline{a}_i) = \sum_{\overline{a} \in A^N} p(\overline{a}_{-i}|\overline{a}_i)u_i^\alpha(\overline{a})$$

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

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

---

[7] Formally, $p_i(k) = \sum_{\overline{a}:\overline{a}_i = k} p(\overline{a})$

[8] In other words,

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

ε by not conforming. Formally, for any $\varepsilon \geq 0$ we say that *correlating device p is a correlated ε-equilibrium of game* $\Gamma(N, \alpha)$ if and only if

$$U_i^{\alpha}(p|\overline{a}_i) \geq \sum_{\overline{a} \in A^N} p(\overline{a}_{-i}|\overline{a}_i) u_i^{\alpha}(k, \overline{a}_{-i}) - \varepsilon \tag{1}$$

for all $i \in N$, any action k and any $\overline{a}_i$.[9] We refer to a correlated 0-equilibrium as a correlated equilibrium. Note that a correlated equilibrium thus defined is consistent with the standard definition of correlated equilibrium.[10] If $\overline{\sigma}$ is a *Nash ε-equilibrium* of game $\Gamma(N, \alpha)$ then there exists a correlated ε-equilibrium in which the mediator independently assigns player i action (or role) k with probability $\overline{\sigma}_i(k)$.

2.4   Social group structures

Given a game $\Gamma(N, \alpha)$ a *social group structure* is given by a partition $\Pi = \{N_1, ..., N_G\}$ of the player set into G subsets. We refer to each N_g as a *social group*. As discussed in the introduction, to make correlated equilibrium an expression of social conformity it is necessary to impose conditions on the probability distribution over roles. In order to define our first property we need one additional definition. We say that action profile $\overline{a}'$ is a *permutation* of another action profile $\overline{a}$ if the number of players in each social group playing each strategy (or assigned each role) is the same.[11] Let $\mathcal{P}^{\Pi}(\overline{a})$ denote the set of action profiles that are permutations of $\overline{a}$.

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

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

WGA captures two important aspects of group behavior that we wish to model: equity and conformity. A probability distribution satisfying WGA provides equality of opportunity within groups because any two players belonging to the same social group have the same probability of being allocated each role within the group.[12] This can be seen as fair. Indeed, as we shall see below this equality

---

[9]Formally, we only require (1) to hold for actions $\overline{a}_i$ that player i can be assigned with positive probability.

[10]The definition of approximate correlated equilibrium equates to a natural approximation to the standard definition of correlated equilibrium, although 'role' is often termed 'signal' (Fudenberg and Tirole 1998). Note, however, that the use of the term ε-correlated equilibrium by Myerson (1986) has a different meaning to the one here.

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

[12]For instance, if $i, j \in N_g$, then $p_i(k) = p_j(k)$ for all $k \in A$.

of opportunity also results in an equality of expected payoff. WGA also implies that conformity within social groups is observed in a conformist equilibrium satisfying WGA because any two players belonging to the same social group are, ex-ante, expected to behave identically. This is because they have the same probability of being allocated each role and behave in identical ways once allocated a role. Furthermore, it is individually rational for players to conform to the behaviors expected of their assigned roles.

We now turn to the group independence property.

Group independence: Given population (N, α), social group structure Π and correlating device p, let i and j be any two players belonging to different social groups. Also, let $p_i(k|a_j)$ denote the probability that player i has role k condition on player j having role a_j. Correlating device p satisfies *group independence (GI)* if there is no correlation of roles between groups. Formally, it requires that $p_i(k) = p_i(k|a_j)$ for all k and a_j.

If social groups are distinct then correlation of actions between groups may be unlikely. This reflects how correlating actions between groups may be difficult because players in different social groups do 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).

Individuals within the same group are likely to have similar attributes (Akerlof and Kranton 2000, Brown 2000, Hogg and Vaughan 2005, Currarini, Jackson and Pin 2008). This motivates a further property of social group structures.

Homophily: Given population (N, α) a social group structure $\Pi = \{N_1, ..., N_G\}$ satisfies *homophily (H)* if the attribute space Ω can be partitioned into G convex subsets $\Omega_1, ..., \Omega_G$ 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. If a social group structure Π and correlating device p satisfy H and WGA then we can think of p as inducing a partition of the population (N, α) into a set of societies in the sense of WCS.

Besides satisfying WGA, GI, and H, there are a number of motivations to *bound the number of societies required*. These are discussed at some length in Wooders, Cartwright and Selten (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 some groups contain many members – there may be some 'extraordinary' players). 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'. WGA requires that players in the same social group behave identically and WGA therefore suggests that small, homogenous, social groups are advantageous. In our framework, correlation of actions allows, however, as we shall see in Section 3.3, increased payoffs.[13] In order to realize these gains and maintain GI it would seem that larger social groups are advantageous. This creates countervailing gains and losses to larger social groups that suggest an optimal group size determined by the heterogeneity of players and the potential gains from correlating actions.[14]

3 The existence of correlated equilibrium consistent with social group structures

In this Section we present our main results demonstrating the existence of an approximate correlated equilibrium consistent with WGA, GI, H and a bounded number of societies. One would expect that we need to impose some assumptions in order to obtain existence of such equilibria. The following example shows that this is the case.[15] In this example H can be consistent with an equilibrium that satisfies WGA only if the number of social groups is as large as the number of players.

Example 1: Players choose between locations B and C. The attribute space is $[0, 1]$. Consider populations (N, α) 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 $i > 2$ 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 ε small, there exists no correlated ε-equilibrium that satisfies WGA and H 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 and WGA is imposed then either player 1 will not follow the conformist

---

[13]This has already been observed by Aumann (1974,1987) for correlated equilibrium for an arbitrary game (without the restrictions imposed in this paper).

[14]Related is the issue of the number of nations as modeled, for example, by Alesina and Spolaore (1997); larger countries imply benefits from greater internal efficiencies, security and ability to cope with external shocks but also imply 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 Le Breton and Weber (2005).

[15]See also Example 2 of WCS (2006).

behavioral rule because he is not playing B 'often enough' or player 2 will not follow the rule because he is matching player 1's choice of B 'too often'. If we 'leave' player 1 in a social group of his own to play 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 H.$\blacklozenge$

This Example illustrates that some continuity assumption is required on the space of attributes. Following WCS we shall introduce a Lipschitz continuity assumption as follows:

Continuity in attributes: The pregame $\mathcal{G} = (\Omega, S, h)$ satisfies *continuity in attributes* if for any $\varepsilon > 0$ and any two games $\Gamma(N, \alpha)$ and $\Gamma(N, \widetilde{\alpha})$, if for all $i \in N$ it holds that $dist(\alpha(i), \widetilde{\alpha}(i)) < \varepsilon$ then for any $j \in N$ and for any action profile $\overline{a}$, $\left| u_j^{\alpha}(\overline{a}) - u_j^{\widetilde{\alpha}}(\overline{a}) \right| < \varepsilon$.

Continuity in attributes dictates that, given action choices, if the attribute function changes only slightly, then payoffs change only slightly. Note that 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 effect payoffs. Our first result demonstrates that with continuity in attributes an upper bound, independent of population size, can be put on the number of social groups necessary for the existence of an approximate correlated equilibrium satisfying WGA, GI and H.

Theorem 1: Consider a pregame $\mathcal{G} = (\Omega, S, h)$ that satisfies continuity in attributes. For any real number $\varepsilon > 0$ there is an integer $G(\varepsilon)$ such that for any population (N, α) there exists a social group structure Π, with no more than $G(\varepsilon)$ groups, and a correlated ε-equilibrium p of induced game $\Gamma(N, \alpha)$ that satisfy WGA, GI and H. Further, $\left| U_i^{\alpha}(p) - U_j^{\alpha}(p) \right| \leq \varepsilon$ for any $i, j \in N_g$, and any N_g.

This result can be seen as the analogue of Theorem 2 in WCS. It shows that a bound can be placed on the number of social groups required for existence of an approximate correlated equilibrium that satisfies WGA, GI and H. In comparing our result to Theorem 2 of WCS note that we do not require there to be a large number of players nor do we require a 'global interaction' property, ensuring that small changes in strategies of others can have only small changes on the utility of a player (see the next Section for a precise definition).

In Section 3.3 where Theorem 1 is proved we provide a more general result showing that near to any Nash equilibrium is a correlated equilibrium and social group structure satisfying the desired properties. In Section 3.3 we shall also discuss the nature of correlation in more detail. At this stage we note that Theorem 1 only demonstrates the existence of an *approximate* correlated equilibrium. To see why we provide a second example

12

Example 2: Players choose between locations B and C. The attribute space is $[0,1]$. Again, without loss of generality, consider populations (N, α) where players are ordered so that $\alpha(i) < \alpha(i+1)$ for all i. If a player of attribute ω chooses B (or C) then his payoff is $|\omega - \omega'|$ where ω' 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 ω' 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 in the unique Nash equilibrium player 1 chooses B, player 2 chooses C, player 3 chooses 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 WGA and H.

Example 2, however, does satisfy continuity in attributes and so Theorem 1 can be applied to show the existence of an approximate correlated equilibrium satisfying WGA and GI. 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 Ω 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 H 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 WGA and GI. If we let a player in a one-member social group choose an optimal strategy then we have a correlated ε-equilibrium as desired.♦

3.1 Ex-post stability

While Theorem 1 requires WGA, GI and H one may wish to impose even more conditions on the correlating device. One particular issue on which we shall focus on is that of ex-post stability (as formulated in Kalai 2004, for example). A correlated equilibrium requires each player to follow the behavioral conformist rule and therefore follow the role assigned them by the mediator. Ex-post, however, once a player has observed the roles assigned to others he may have an incentive to change his action.[16] The following example illustrates such a situation.

Example 3: 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

---

[16]To use a Rawlsian thought experiment one can see that WGA 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.

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.♦

In many contexts, ex-post instability is an unavoidable property of Nash and correlated equilibrium. If we wish, however, to interpret the mediator as reflecting societal norms then there is something slightly worrying about ex-post instability. First, if the norm is to be followed (over time) then we would prefer that players have an incentive to conform both ex-post as well as ex-ante to the allocation of roles. Second, a notion of a norm suggests that in some sense the behavior of a social group should be predictable. In Example 3, while there is nothing wrong in one player constituting a social group, it is difficult to interpret the actions of the celebrity in terms of behavioral conformity. We feel that this is primarily because behavioral conformity suggests 'predictability in aggregate behavior' which does not hold for Example 3. This motivates a further property in which we do require predictable behavior within groups.

Predictable group behavior (PGB): Let $\Gamma^p(N, \alpha)$ be a game with roles and let Π be a social group structure. Correlative 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 $\overline{a}$ and $\overline{a}'$, if $p(\overline{a}), p(\overline{a}') > 0$ then $\overline{a}' \in \mathcal{P}^\Pi(\overline{a})$.

Note that PGB implies ex-post stability because, ex-ante, ex-post outcomes are predictable. PGB (and Theorem 2 to follow) could be extended by relaxing PGB to require only that aggregate behavior be approximately predictable. To do so would require cumbersome notation and further approximation arguments so we prefer the current PGB condition. Example 3 makes clear that additional restrictions are required to obtain PGB. We impose the following property from WCS and our prior papers.

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

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

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

Global interaction implies that no one individual can have a significant effect on the payoff of any other player in large games. It rules out, for instance,

Example 3. When global interaction is satisfied we obtain PGB in games with sufficiently many players.

Theorem 2: Consider a pregame $\mathcal{G} = (\Omega, S, h)$ 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, α) where $|N| > \eta(\varepsilon)$, there exists a social group structure Π, with no more than $G(\varepsilon)$ social groups, and a correlated ε-equilibrium p of induced game $\Gamma(N, \alpha)$ that satisfy WGA, GI, H and PGB.

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. 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 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 4 illustrates the Theorem.

Example 4: Players have to choose between two locations B and C. There is a unique attribute. Given game $\Gamma(N, \alpha)$ let $\overline{b}$ denote the proportion who choose location B and let $\overline{c}$ denote the proportion who choose C. A player's payoff is given by $-\overline{b}$ if he chooses B and $-\overline{c}$ 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 PGB but not WGA. The second equilibrium satisfies WGA but not PGB. There is no Nash equilibrium that satisfies both WGA and PGB.

Kalai (2004) demonstrates that in games with many players a mixed strategy Nash equilibrium is approximately ex-post stable. This means that aggregate behavior can be predicted ex-ante with some precision. But, because 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). This means that ex-ante expected payoffs are lower with the mixed strategy Nash equilibrium than with any pure strategy Nash equilibria.

There exists, however, a correlated equilibrium that satisfies both WGA and PGB. The device 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 knows 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. This means that payoffs

are the same as in the pure strategy Nash equilibrium (and therefore higher than in the mixed strategy Nash equilibrium).♦

One point to highlight from this example is the necessity of correlation of actions in order to obtain an equilibrium that satisfies both WGA and PGB. While both WGA and PGB can be achieved on their own through uncorrelated actions they are only simultaneously possible with correlation. If players have desires for WGA (because of fairness) and PGB (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.

3.2 Subjective beliefs and stereotyping

The assumption that players know the correlating device p shall now be relaxed. Instead, players are modelled as having subjective beliefs about the device. Specifically, there is a given *set of beliefs* $\{\beta_i\}_{i \in N}$, where β_i denotes the beliefs of player i and is given by a probability distribution over the set of action profiles. Thus, $\beta_i(\overline{a})$ denotes the probability that player i puts on players having being assigned roles according to action profile $\overline{a}$ and $\beta_i(\overline{a}_{-i}|\overline{a}_i)$ denotes the probability that player i puts on roles being assigned according to action profile $\overline{a}$ given that he is assigned role $\overline{a}_i$. 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 ε-equilibrium* if

$$\sum_{\overline{a} \in A^N} \beta_i(\overline{a}_{-i}|\overline{a}_i)u_i(\overline{a}) \geq \sum_{\overline{a} \in A^N} \beta_i(\overline{a}_{-i}|\overline{a}_i)u_i(k, \overline{a}_{-i}) - \varepsilon$$

for each $i \in N$ and $\overline{a}_i, k \in A$. 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* β_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. That is, the player expects WGA to hold. We do assume, however, that a player does not 'stereotype' himself and this requires a slight reformulation of WGA.[17]

---

[17] In earlier versions of this work we referred to this as "other-stereotyping."

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

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

It is a simple extension of Theorem 1 to show that a subjective correlated ε-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 WGA but suppose, however, that a player has stereotyped beliefs. The following result shows that in games with sufficiently many players 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 an equilibrium there is no reason to expect that stereotyping would be consistent with equilibrium.

Theorem 3: Consider a pregame $\mathcal{G} = (\Omega, S, h)$ that satisfies continuity in attributes and global interaction. For any real number $\varepsilon > 0$ there are integers $\eta(\varepsilon)$ and $G(\varepsilon)$ with the properties that for any population (N, α) with $|N| > \eta(\varepsilon)$ there is a social group structure Π, of no more that $G(\varepsilon)$ groups, such that, for any Nash equilibrium p, there exists a subjective ε-correlated equilibrium $\{\beta_i\}_{i \in N}$ where each β_i is stereotyped and $|U_i^\alpha(p|\overline{a}_i) - U_i^\alpha(\beta_i|\overline{a}_i)| \leq \varepsilon$ for all $i \in N$.

Stereotyping can therefore be consistent with equilibrium and not change the expected payoff of the player who is stereotyping. It is worth noting, however, that stereotyping can influence behavior and thus influence the payoffs of other players. An example illustrates.

Example 5: 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 $\overline{\omega}_B$ and $\overline{\omega}_C$ denote the average charmingness of players in locations B and C. The celebrity likes to have charming neighbors

17

but has a slight preference for C over B; his payoff is $\overline{\omega}_B$ if he chooses location B and $\overline{\omega}_C + \delta$ for some small δ 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 the example, for instance, 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.

3.3 Permutation and payoffs

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 $\mathcal{G} = (\Omega, S, h)$ that satisfies continuity in attributes. Until otherwise stated we shall also take as given a population (N, α) and social group structure $\Pi = \{N_1, .., N_G\}$.

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

Lemma 1: Consider any action profile $\overline{a}$. If $\overline{a}' \in P^\Pi(\overline{a})$ then there exists a (not necessarily unique) permutation of players γ such that $\overline{a}' = \overline{a}^\gamma$.[18] Furthermore, if γ is a permutation of players then $\overline{a}^\gamma \in P^\Pi(\overline{a})$.

Thus, if action profile $\overline{a}'$ is a permutation of $\overline{a}$ 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 $\overline{a}'$, plays the same action that $\gamma(i)$ plays, according to action profile $\overline{a}$.

---

[18] That is, $\overline{a}_i' = \overline{a}_i^\gamma$ for all $i \in N$.

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. Let

$$D := \max_{i,j \in N_g} \{dist(\alpha(i), \alpha(j))\}.$$

A simple example is provided after the proof.

Lemma 2: Consider any action profile $\overline{a}$ and permutation of players γ. If $D < \delta$ then

$$\left| u_i^\alpha(k, \overline{a}_{-i}^\gamma) - u_{\gamma(i)}^\alpha(k, \overline{a}_{-\gamma(i)}) \right| < \delta \tag{2}$$

for any $k \in A$.

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

$$\left| u_{\gamma(i)}^{\widetilde{\alpha}}(k, \overline{a}_{-\gamma(i)}) - u_{\gamma(i)}^\alpha(k, \overline{a}_{-\gamma(i)}) \right| < \delta \tag{3}$$

for all $i \in N$ and any $\overline{a} \in A^N$. We know that $\overline{a}_i^\gamma = \overline{a}_{\gamma(i)}$ for all $i \in N$. The inequality (2) now follows.∎

We illustrate Lemma 2 with an example: Consider a population (N, α) with four players, $N = \{1, 2, 3, 4\}$ and a social group structure Π 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 $\overline{a} = (\overline{a}_1, \overline{a}_2, \overline{a}_3, \overline{a}_4)$ we obtain that $\overline{a}^\gamma = (\overline{a}_2, \overline{a}_3, \overline{a}_1, \overline{a}_4)$. Observe that $\overline{a}_{-1}^\gamma = (\overline{a}_3, \overline{a}_1, \overline{a}_4)$ and $\overline{a}_{-\gamma(1)} = (\overline{a}_1, \overline{a}_3, \overline{a}_4)$. Lemma 2 implies that

$$\left| u_1^\alpha(k, \overline{a}_{-1}^\gamma) - u_2^\alpha(k, \overline{a}_{-2}) \right| < \max_{i,j \in \{1,2,3\}} \{dist(\alpha(i), \alpha(j))\}.$$

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 γ we denote by p^γ the correlating device where $p^\gamma(\overline{a}^\gamma) = p(\overline{a})$ for all $\overline{a} \in A^N$. One way to interpret the device p^γ is that it randomly determines allocated roles $\overline{a}$ according to the correlating device p but then allocates player i the role of player $\gamma(i)$ – that is, it gives player i role $\overline{a}_{\gamma(i)}$ 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.[19]

Lemma 3: Let p be any correlating device, γ any permutation of players and i any player. If $D < \delta$ then

$$\left| \sum_{\overline{a} \in A^N} p^\gamma(\overline{a}_{-i}|\overline{a}_i) u_i^\alpha(k, \overline{a}_{-i}) - \sum_{\overline{a} \in A^N} p(\overline{a}_{-i}|\overline{a}_i) u_{\gamma(i)}^\alpha(k, \overline{a}_{-\gamma(i)}) \right| < \delta \tag{4}$$

---

[19] For example, Lemma 3 implies that $\left| U_i^\alpha(p^\gamma|\overline{a}_i^\gamma) - U_{\gamma(i)}^\alpha(p|\overline{a}_{\gamma(i)}) \right| < \delta$.

for any action $k \in A$.

Proof: Instead of $\sum_{\overline{a}} p^\gamma(\overline{a}_{-i}|\overline{a}_i) u_i^\alpha(k, \overline{a}_{-i})$ we can write $\sum_{\overline{a}^\gamma} p^\gamma(\overline{a}_{-i}^\gamma|\overline{a}_i^\gamma) u_i^\alpha(k, \overline{a}_{-i}^\gamma)$ which equals $\sum_{\overline{a}} p(\overline{a}_{-i}|\overline{a}_i) u_i^\alpha(k, \overline{a}_{-i}^\gamma)$. Equation (4) can therefore be restated

$$\left| \sum_{\overline{a} \in A^N} p(\overline{a}_{-i}|\overline{a}_i) \left[u_i^\alpha(k, \overline{a}_{-i}^\gamma) - u_{\gamma(i)}^\alpha(k, \overline{a}_{-\gamma(i)}) \right] \right| < \delta$$

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 γ 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 allocated role given correlating device p^* then player i could gain at most 2δ from deviating from his allocated role given correlating device $p^{*\gamma}$. This leads to the following result.

Lemma 4: Let p^* be any correlated ε-equilibrium of game $\Gamma(N, \alpha)$ and let γ be any permutation of players. If $D \leq \delta$ then correlating device $p^{*\gamma}$ is a correlated $2\delta + \varepsilon$-equilibrium of game $\Gamma(N, \alpha)$.

Proof: If p^* is a correlated ε-equilibrium then

$$\sum_{\overline{a} \in A^N} p^*(\overline{a}_{-i}|\overline{a}_i) u_{\gamma(i)}^\alpha(\overline{a}) \geq \sum_{\overline{a} \in A^N} p^*(\overline{a}_{-i}|\overline{a}_i) u_{\gamma(i)}^\alpha(k, \overline{a}_{-\gamma(i)}) - \varepsilon$$

for all $i \in N$ and any $k \in A$. Applying Lemma 3 implies that

$$\sum_{\overline{a} \in A^N} p^{*\gamma}(\overline{a}_{-i}|\overline{a}_i) u_i^\alpha(\overline{a}) \geq \sum_{\overline{a} \in A^N} p^{*\gamma}(\overline{a}_{-i}|\overline{a}_i) u_i^\alpha(k, \overline{a}_{-i}) - 2\delta - \varepsilon$$

for all $i \in N$ and $k \in A$ 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.

Proof of Theorem 1: First, from standard theorems, a Nash equilibrium exists. This implies the existence of a correlated equilibrium p^* in which roles are distributed independently across players. Given $\varepsilon > 0$ partition Ω into G convex sets $\Omega_1, ..., \Omega_G$ where $\max_{\omega, \omega' \in N_g} dist(\omega, \omega') < \frac{\varepsilon}{2}$. Let $\Pi = \{N_1, ..., N_G\}$ denote the social group structure where $i \in N_g$ if $\alpha(i) \in \Omega_g$. Let Λ denote the set of permutations of players (consistent with Π). By Lemma 4 we know that each $p^{*\gamma}$ is a correlated ε-equilibrium. Consider correlating device p' given by

$$p'(\overline{a}) = \frac{1}{|\Lambda|} \sum_{\gamma \in \Lambda} p^{*\gamma}(\overline{a}). \tag{5}$$

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 ε-equilibrium. Also, p' satisfies WGA by construction. Finally, roles are distributed independently across players by device p^* and in every p^{*^γ} which means that p' must satisfy GI.■

Proof of Theorem 2: For any $\varepsilon > 0$, Theorem 1 of WCS demonstrates that in games with sufficiently many players there exists a Nash $\frac{\varepsilon}{2}$-equilibrium *in pure strategies* p^*. That p^* is an equilibrium in pure strategies implies $p(\overline{a}^*) = 1$ for some action profile $\overline{a}^*$. Clearly p^* satisfies PGB. Partition Ω into G convex sets $\Omega_1, ..., \Omega_G$ where $\max_{\omega, \omega' \in N_g} dist(\omega, \omega') < \frac{\varepsilon}{4}$. Let $\Pi = \{N_1, ..., N_G\}$ denote the social group structure where $i \in N_g$ if $\alpha(i) \in \Omega_g$. As in the proof of Theorem 1 let Λ denote the set of permutations of players (consistent with Π) and let p' be defined as in (5). We can use the same arguments as in the proof of Theorem 1 to see that p' is a correlated ε-equilibrium satisfying WGA, GI and H. It should be clear that the device p' also satisfies PGB.■

Proof of Theorem 3: Given any $\varepsilon > 0$ in games with sufficiently many players, for any Nash equilibrium p there exists a Nash $\frac{\varepsilon}{2}$-equilibrium in pure strategies p^* where $|U_i^\alpha(p|\overline{a}_i) - U_i^\alpha(p^*|\overline{a}_i)| < \frac{\varepsilon}{2}$ (Kalai 2004 and WCS). Partition Ω into G convex sets $\Omega_1, ..., \Omega_G$ where $\max_{\omega, \omega' \in N_g} dist(\omega, \omega') < \frac{\varepsilon}{4}$. Let $\Pi = \{N_1, ..., N_G\}$ denote the social group structure where $i \in N_g$ then $\alpha(i) \in \Omega_g$. Fix a player $i \in N$. Let Λ_i denote the set of permutations of players in which i is not permuted, that is, $\gamma(i) = i$. Define beliefs β_i where

$$\beta_i(\overline{a}) = \frac{1}{|\Lambda_i|} \sum_{\gamma \in \Lambda_i} p^{*^\gamma}(\overline{a}). \tag{6}$$

Beliefs β_i are stereotyped and following the logic of the proofs of Theorems 1 and 2 $|U_i^\alpha(p^*|\overline{a}_i) - U_i^\alpha(\beta_i|\overline{a}_i)| < \frac{\varepsilon}{2}$. Further, constructing β_i for each $i \in N$ we obtain a subjective ε-correlated equilibrium $\{\beta_i\}_{i \in N}$.■

4   Concluding remarks

This paper models conformity and social norms in settings where different people 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 such as WGA, GI and PGB and have demonstrated the existence of a correlated equilibria satisfying these properties. One way to interpret this is to argue that social interaction acts as a form of equilibrium-selection device that selects correlated equilibria satisfying certain properties.

One obvious question is 'where does the correlation come from if we recognize that there is no formal device telling people what to do?' 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 'perfect' correlation required of PGB or WGA may be asking too much it may be possible for social groups to obtain correlated equilibrium that approximates PGB and WGA. Example 4, 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 those that are 'similar'. This raises the question of how a player would form his beliefs about the actions expected of others.

One way to address some of these issues would be to make the role-allocation 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-allocation device would enable one to determine from the model whether correlation and WGA and GI 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 endogenizing the allocating 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).

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 WCS 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 Lipshitz 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.

[4] Aumann R.J. (1974) "Subjectivity and correlation in randomized strategies," *Journal of Mathematical Economics* 1, 67-96.

[5] Aumann, R.J. (1987) "Correlated equilibrium as an expression of Bayesian rationality," *Econometrica* 55: 1-18.

[6] Bernheim, B. D. (1994) "A theory of conformity," *The Journal of Political Economy*, Vol. 102, No. 5. (Oct., 1994), pp. 841-877.

[7] Binmore, K., (1989) "Social contract I: Harsanyi and Rawls," *The Economic Journal* C: 84-102.

[8] Brandenburger, A. and E. Dekel (1987) "Rationalizability and correlated equilibria," *Econometrica* 55:1391-1402.

[9] Brown, R. (2000) "Social identity theory: past achievements, current problems and future challenges," *European Journal of Social Psychology* 30: 745-778.

[10] Cartwright, E. (2003) "Bounded rationality, social conformity and large games," University of Warwick Ph. D. thesis.

[11] Cartwright, E. and M. Wooders (2003a) "On equilibrium in pure strategies in games with many players," Warwick Economics Research Paper No. 686.

[12] Cartwright, E. and M. Wooders (2003b) "Conformity and bounded rationality in games with many players," Warwick Economics Research Paper No. 687.

[13] Cartwright, E. and M. Wooders (2004) "Behavioral conformity in games with many players," Presented at ISER University of Osaka, June 2005; invited plenary session at the 15th International Game Theory Conference, Stony Brook, New York, July 2004, under the title of "Behavioural conformity and purification in games with many players'; Birmingham Workshop on Economic Theory, July 2004.

[14] Cartwright, E. and M. Wooders (2005) "Correlated equilibrium and behavioural conformity," The Warwick Economics Research Paper Series (TWERPS) 732.

[15] Cartwright, E. and M. Wooders (2006) "Conformity, correlation and equity," typescript, presented at: PET 07 Hanoi, July 2006, Stony Brook Game Theory International Conference July (2006) under the title of "Correlated subjective equilibrium with stereotyping"; SING (Spanish, Italian, Netherlands Game Theory) Meeting, invited plenary lecture, June (2006); Center for Rationality, Hebrew University, Jerusalem, March 2006, the University of Indiana, Bloomington, March 2006.

[16] 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.

[17] Conley, J. and M.H. Wooders (2001) "Tiebout economics with differential genetic types and endogenously chosen crowding characteristics," *Journal of Economic Theory* 98: 261-294.

[18] Currarini S., M. Jackson and P. Pin (2008) "An economic model of friendship: homophily, minorities and segregation," working paper.

24

[19] 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.

[20] Dhillon A., and J. F. Mertens, (1996) "Perfect correlated equilibria," *Journal of Economic Theory* 68: 279-302.

[21] Elster, J., (1989) "Social norms and economic theory," *Journal of Economic Perspectives* 3: 99-117.

[22] Fehr, E. and K. Schmidt (1999) "A theory of fairness, competition and cooperation," *Quarterly Journal of Economics* 114: 817-868.

[23] Forges, F., (1986) "An approach to communication equilibria," *Econometrica* 54: 1375-1385.

[24] Fudenberg, D. and J. Tirole (1998) *Game Theory,* MIT Press.

[25] Friedman, D. (1996) "Equilibrium in evolutionary games: Some experimental results," *Economic Journal* 106: 1-25.

[26] Gravel, N. and S. Thoron (2007) "Does endogenous formation of jurisdictions lead to wealth-stratification?" *Journal of Economic Theory* 132: 569-583

[27] 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.

[28] Hayek, F.A., (1960) *The Constitution of Liberty* Routledge and Keegan Paul Ltd. London.

[29] Hayek, F.A., (1982) *Law, Legislation and Liberty*, T.J. Press, Padstow.

[30] 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.

[31] 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.

[32] Hart, S., (2005) "Adaptive heuristics," *Econometrica* 73: 1401-1430.

[33] 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.

[34] Hogg, M.D., and G.M. Vaughan, (2005) *Social Psychology*, Pearson/Prentice Hall, Marlow.

[35] 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.

[36] Johnson, D.W., and F. Johnson, (1987) *Joining Together: Group Theory and Group Skills*, Prentice Hall, London.

[37] Khan, A. and Y. Sun (2002) "Noncooperative games with many players," *Handbook of Game Theory*, R. Aumann and S. Hart, eds. North Holland.

[38] Konishi, H. and M. Utku Unver (2006), "Credible group-stability in many-to-may matching problems," *Journal of Economic Theory* 129: 57-80.

[39] Kalai, E. (2004) "Large robust games," *Econometrica* 72: 1631-1666.

[40] Myerson, R.B. (1986) "Acceptable and predominant correlated equilibria," *International Journal of Game Theory* 15: 133-154.

[41] Rabin, M. (1993) "Incorporating fairness into game theory and economics," *American Economic Review* 83: 1281-1302.

[42] Rapoport, A, D. Seale and E. Winter (2001) "Coordination and learning behavior in large groups with asymmetric players," *Games and Economic Behavior* 39, # 1, 111-136

[43] Rawls, J., (1972) *A Theory of Justice* Oxford University Press.

[44] Robson, A. and M. Wooders (1997) "The growth-maximizing distribution of income," *International Economic Review* 38, 511-526.

[45] Sherif, M. (1966) *In Common Predicament: Social Psychology of Intergroup Conflict and Cooperation* Houghton Mifflin: Boston.

[46] Schelling, T. C. (1960) *The Strategy of Conflict.* Cambridge, Harvard University Press, 1960.

[47] Selten, R. (1980) "A note on evolutionary stable strategies in asymmetric animal conflicts", *Journal of Theoretical Biology* 84: 93-101.

[48] Sugden, R. (1989) "Spontaneous order," *Journal of Economic Perspectives* 3, # 4, 85-97.

[49] Tajfel, H., (1978) *Differentiation Between Social Groups: Studies in the Social Psychology of Intergroup Relations*, Academic Press, London.

[50] 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.

[51] Wooders, M., E. Cartwright and R. Selten (2006) "Behavioral conformity in games with many players," *Games and Economic Behavior*, 57: 347-360 .