Abstract

In this paper I extend Matthew Rabin's model of fairness equilibria (1993) to groups of individuals. This allow me to introduce three aspects from reality that are absent in game theory: i) individuals discriminate in favor of members of their own groups, ii) individuals like individuals that not only are kind to them, but are kind to other individuals, specially individuals of their own groups, and iii) individuals discrimate in favor of members of groups they like. I define a new equilibrium that takes in consideration this emotions, what I call group fairness equilibrium. Rabin defines the mutual-max outcomes for a single game as outcomes where each player maximize the other player's material mayoffs and the mutual-min outcomes as outcomes where each player minimize the other player's material payoffs. Some basic results of my model are that a combination of strict Nash equilibrium in several games, will always be a group fairness equibrium for large values of the material payoffs, and that any outcome that is either strictly mutual-max for both games or strictly mutual-min for both games is a group fairness equilibrium for large values of the material payoffs.

Group Fairness and Game Theory

Alejandro T. Moreno

1 Introduction

Most economic models assume that individuals focus exclusively in the material gains and that individuals do not care about the groups of the players they are interacting with. However, when we interact with other individuals, we care about the groups they belong, generally treating better those individuals that belong to our own group.[1] We treat better somebody if he or she is our relative, countryman, if he cheers for the same team and even if he is assigned to a group with us randomly.[2] For example, if an individual plays the prisoners dilemma with a player that do not have any relation to him or if he plays it with somebody that belongs to a close group, as his family, he would treat them differently. In most situations we would expect a player to be kinder to a relative than somebody that does not have any relation with him.

Individuals also care if the individuals they are interacting with are kind or unkind. In his seminal paper, Rabin (1993) introduces fairness to game theory by modelling how individuals want to be kind to other individuals that are kind to them, and be unkind to other individuals that are unkind to them. Rabin shows how in the Prisoners Dilemma the desire to be kind to an individual that has been kind to us makes possible an equilibrium (Rabin defines it as fairness equilibrium) where both players play cooperate.

However, when individuals assess how kind other individuals are, they not only care about how kind they are with themselves, but also how kind they are with other individuals, specially with other individual they care. For example, if two members of the same family, a father and a son, play a prisoners' dilemma in two games with another player, it is reasonable to think that the father would be kinder to the player if the player is kind to his son, and the father would be unkinder to the player if the player is unkind to his son.

[1] See Stereotyping, Prejudice, and Discrimination, Chapter 25 from the Handbook of Social Psychology by Susan T. Fiske.

[2] See George Akerlof and Rachel Kranton in Economics and Identity. Quarterly Journal of Economics, 2000.

And individuals sometimes are kind to members of a group they like and individuals sometime are unkind to members of a group they dislike. Individuals sometimes like a group if somebody from that group has helped them or other members of their own group and individuals sometimes dislike a group if somebody from that group has treated them, or to other member of their own group, badly.

For example, if members from one family are playing against members from another family, let's say the sons play against each other in a game and the fathers play against each other in another game, it is reasonable to think that the father not only cares about how the other father treats him, but he also cares how the son of the other family treats his own son. If one son is kind to the other son, the father of the other son may be want to be kinder to the other member of that family in return.

We can see from the examples above that individuals form emotions of fairness between groups, and this is not an insignificant phenomena. In interethnic conflicts, individuals from each group are targeted in order to retaliate for the attacks perpetrated by members from their groups, even if the individuals that are targeted are not related to previous attacks. And individuals sometimes buy products at a higher price or of subpar quality if the owners of the firms that produce the products are from the same country as them.

In this paper I formulate a model that introduces to game theory what I call group fairness, that is, the emotions of fairness and reciprocity over the treatment of members of the same group. I do this by extending the model developed by Matthew Rabin (1993) of fairness equilibria to groups of individuals.

My model incorporates three observations of individuals' interaction:

1) Individuals discriminate in favor of individuals of the same group.

2) Individuals are willing to sacrifice their own material well-being to help those that not only help them but help others and punish those who not only are unkind to them but that are unkind to others.

3) Individuals are willing to sacrifice thier own material well-being to help members of a group they like and punish those that belong to a group they dislike. I assume that individuals like a group if somebody from that group has helped them or other members of their own group and individuals dislike

a group if somebody from that group has treated them, or to other member of their own group, badly.

In section 2 I introduce my model. I start by reviewing Matthew Rabin's model of fairness equilibrium and then extending it to groups of individuals. Rabin's model is defined over a single game of two players, however, in order to analyze interaction between groups of individuals, I have to work with more games. I model the easiest case: two games of two individuals each game.

Some basic results of my model are the following: a) a combination of strict Nash equilibrium in both games will always be a group fairness equibrium for large values of the material payoffs; b) any outcome that is either strictly mutual-max for both games or strictly mutual-min for both games, is a group fairness equilibrium for large values of the material payoffs, and c) if one of the games has a strict Nash equilibrium that is a mutual-max outcome, then, when the material payoffs of the game grow arbitrarily large, the fairness equilibrium of the single game is neutral and the group fairness of the whole game will be defined by the other game; if one game does not have mutual-max outcome, then, as the material payoffs of that game grow arbitrarily large, the whole game has a weakly negative group fairness equilibrium.

In section 3 I analyze the sequential case where one game is played first and then the other. This allows me to model how players in the first game may try to influence the emotions of players in the second game by being kind or unkind themselves.

In section 4 I introduce another emotion into my model: individuals dislike to being discriminated against. Although for some groups as family, it is seen naturally that individuals treat better those members of their own group, for some groups as race and gender, individuals dislike to be discriminated.

In section 5 I apply my model to the example of a monopoly that gives a product for free to an individual in need to improve its image to its consumers. By improving its image, the monopoly is able to charge consumers a higher price, as individuals are kinder to the monopoly.

I concludeand discuss possible extensions in section 6.

2 Model

2.1 Review of Matthew Rabin's Fairness Equilibrium

Matthew Rabin (1993) introduces fairness to game theory by modeling how if one player, let's say player 1, believes that another player, let's say player 2, is sacrificing his own material payoffs to help him, then player 1 may want to sacrifice his own material payoffs as well in order to help player 2; and if player 1 believes that player 2 is treating him badly, he may sacrifice his own material payoff to treat him badly in return.

Rabin models a two players game, where S_i is set of possible actions for individual i, a_i are individual $i's$ actions, b_j are individual $i's$ beliefs of the actions of individual j and c_i is what individual i believes are individual $j's$ beliefs of the actions of individual i. $\pi_i(a_i, b_j)$ are individual $i's$ material payoffs given that he takes action a_i and he believes individual $j's$ actions are b_j. $\pi_j^e(b_j)$ is what individual i think is the "equitable payoff" for individual j and is defined as $\pi_j^e(b_j) = \left[\pi_j^h(b_j) + \pi_j^l(b_j)\right]/2$, where $\pi_i^h(b_i)$ is individual $i's$ highest possible payoff and $\pi_i^l(b_j)$ is individual $i's$ lowest possible payoff from all possible Pareto outcomes. $\pi_i^{\min}(c_i)$ is the lowest possible outcome.

Rabin models fairness between two individuals by defining two functions: one that represents how kind an individual is to the other individual and a second that represent an individual's beliefs about how kind the other individual is to him. Individual's kindness function is given by:

$$f_i(a_i, b_j) \equiv \frac{\pi_j(b_j, a_i) - \pi_j^e(b_j)}{\pi_j^h(b_j) - \pi_j^{\min}(b_j)}$$

while individual's belief in how kind the other individual is to him is given by:

$$\widetilde{f}_j(b_j, c_i) \equiv \frac{\pi_i(c_i, b_j) - \pi_i^e(c_i)}{\pi_i^h(c_i) - \pi_i^{\min}(c_i)}$$

Rabin defines a utility function that incorporates these kindness functions to represent how an individual wants to be kind to an individual he believes is being kind to him and wants to be mean to an individual he believes is mean to him. The utility function is defined as:

$$U_i(a_i, b_j, c_i) = \pi_i(a_i, b_j) + \widetilde{f}_j(b_j, c_i)\left[1 + f_i(a_i, b_j)\right]$$

This model captures individuals' desire to be kind to somebody that has been kind to them and individuals' desire to be unkind to somebody that has been unkind to them. If individual i believes that individual j is kind to him (the function $\widetilde{f}_j(b_j, c_i)$) is positive, then he would increase his utility by being kind in return (the funcion $f_i(a_i, b_j)$ would be positive). If individual i believes that individual j is unkind to him (the function $\widetilde{f}_j(b_j, c_i)$) is negative, then he would increase his utility by being unkind in return (the funcion $f_i(a_i, b_j)$ would be negative).

I objetive is to extend Rabin's model to introduce some aspects from reality that are absent from his analysis. First, it is easier for individuals to cooperate if they belong to the same group, for example if they are relatives. Evidence from social psychology shows that individuals tend to treat better those individuals that belong to their own group, even if the group was formed randomly. Second, individuals think that a person is kind not only if he is nice to them, but if he is nice to other individuals. And third, individuals see themselves as part of groups and they care about the animosity of one group toward the other. Rabin models fairness only for two players. However, in order to analyze emotion of group fairness, I extend Rabin's concept of fairness to include more than two players.

Although I model games where individuals play directly only in pairs, I assume that individuals observe and take in consideration the interaction of players in other games when they form their beliefs of kindness. While most economists have assumed that players only care about what's happen in the games they play, my objective is to model how the outcome in one game may affect the outcome in other games.

2.1.1 Some single game propositions

Before extending Rabin's model to more players, I give some single game propositions that complement those of Rabin and that will help me with the propositions for the case of group fairness equilibrium for the next section. All proofs are in the appendix.

Rabin defines a mutual-max strategy as a strategy where both players mutually maximize each other's material payoffs and a mutual-min strategy as a strategy where both players mutually minimize each other's utility. Rabin also defines the sign of the outcome of a game in funcion of the sign of the kindness function of each player. I write his definitions to use them in my propositions.

Definition 1: A strategy pair $(a_1, a_2) \in (S_1, S_2)$ is a mutual-max outcome if, for $i = 1, 2$, $j \neq i$, $a_i \in \arg\max_{a \in S_i} \pi_j(a, a_j)$.

Definition 2: A strategy pair $(a_1, a_2) \in (S_1, S_2)$ is a mutual-min outcome if, for $i = 1, 2$, $j \neq i$, $a_i \in \arg\min \pi_j(a, a_j)$.

Definition 3: a) An outcome is strictly positive if for $i = 1, 2$, $f_i > 0$. b) An outcome is weakly positive if for $i = 1, 2$, $f_i \geq 0$. c) An outcome is strictly negative if for $i = 1, 2$, $f_i < 0$. d) An outcome is weakly negative if for $i = 1, 2$, $f_i \leq 0$. e) An outcome is neutral if for $i = 1, 2$, $f_i = 0$. f) An outcome is mixed if for $i = 1, 2$, $i \neq j$, $f_i f_j < 0$.

Proposition 1: For a single game, there is an $\overline{X}$ for which for all $X > \overline{X}$ all fairness equilibria that remain in a game have to be Nash equilibria (not necessarily strict).

Proposition 1 tell us that as the material payoffs increase arbitrarily, the fairness equilibria of the game have to be also Nash equilibria. If individuals are not playing a Nash equilibrium then there is at least one deviation for one player that improves his material payoffs. As the material payoffs increase arbitrarily large individuals care more about them and less about the fairness payoffs, until the point that material considerations dominate the fairness consideration and individual deviate.

Proposition 2: For a single game, there is a value $\overline{X}$ for which for all $X > \overline{X}$, there is not a positive fairness equilibrium. If a single game does not have a mutual-max outcome, there is a value $\overline{X}$ for which for all $X > \overline{X}$ only exists weakly negative fairness equilibrium.

Proposition 2 tell us that as the material payoffs grow arbitrarily large, the positive fairness equilibria are eliminated and only the weakly negative fairness equilibria are left. As the income increases, the material payoffs dominate the fairness considerations. Because individuals are maximizing their own material payoffs, other individuals would not think they are been kind and the positive fairness is eliminated. The only possible equilibria are neutral or negative. The second part of proposition 2 refers to the fact that if the Nash equilibrium is not a mutual-max outcome, then at least one individual is playing a strategy that is not maximizing the other's utility and therefore he is been strictly unkind to him. The other individual would also be unkind to him (at least weakly), making the equilibria weakly negative.

2.2 My model: 4 individuals and 2 games

I extend Rabin's fairness to groups of individuals by analyzing the case of two games with two players each game. The games are game 1 and game 2, and each game has two players: player 1 and player 2. I will call player i that plays in game m as player im, where $i = 1, 2$ and $m = 1, 2$. For example, player 1 from game 1 will be player 11. The players are members of groups and some players can belong to the same group. Let me give an assumption that will simplify the notation: if there are two individuals from the same group, one in each game, then both players will be players 1 from game 1 and 2 (player 11 and player 12) or both players will be players 2 from game 1 and game 2 (21 and player 22).

S_{im} is the set of possible actions of individual im. $a_{im} \in S_{im}$ are the actions of the individual im, $b_{im} \in S_{im}$ are the beliefs of the individual jm about the actions of the individual im and $c_{im} \in S_{im}$ are the beliefs of individual im about the beliefs of the individual jm about his own actions (im's actions). The last two variables refer to the beliefs that individuals have about the actions and beliefs of the individuals they are playing with. I introduce two new variables that represent individual's beliefs about the actions and beliefs of the individuals that play in the other game. $d_{im} \in S_{im}$ are the beliefs of the individual jn of the actions of individual im and $e_{im} \in S_{im}$ are the beliefs of individual in about the beliefs of individual jm about the actions of player im.

Individuals sometimes belong to groups whose members are be very close to each other, like members of the same family, and individuals sometimes belong to groups whose members are not so close to each other. I define a variable v_1 that represent if both individuals 1 (in game 1 and 2) belong to the same group and if they do, how close the members of that group are. v_2 represents if both individuals 2 belong to the same group and if they do, how close the members of that group are. v_i is defined from zero to one, where small values of v_i, mean that $i1$ and $i2$ belong to a group whose members are not very close while high values of v_i mean that player $i1$ and player $i2$ belong to a group whose members are close. At the extremes, if $v_i = 0$, then player $i1$ and player $i2$ do not belong to any common group and if $v_i = 1$, then $i1$ and $i2$ are the same player. The variable σ_1 represents if both players in game 1 belong to the same group, and if they do, how close they are, and the variable σ_2 represents ib both players in game 2 belong to the same group and if they do how close they are.

I define a function that represents how kind an individual is.

Definition 4: The kindness of the player im is given by:

$$f_{im}(a_{im}, b_{jm}) \equiv \frac{\pi_{jm}(b_{jm}, a_{im}) - \pi^e_{jm}(b_{jm})}{\pi^h_{jm}(b_{jm}) - \pi^{\min}_{jm}(b_{jm})}$$

where $\pi^e_{jm}(b_{jm}) = \left[\pi^h_{jm}(b_{jm}) + \pi^l_{jm}(b_{jm})\right]/2$. This function is exactly the same as Rabin's kindness function, but the notation changes to take in consideration that there are two games with two players each game. Now I define a funtion that represents how an individual judges other individuals. I modify Rabin's function to take in consideration that individuals not only care about how kind is the individual they are playing with, but how kind are other members of the same group.

Definition 5: Individual im beliefs of how kind is individual jm and his peer is given by:

$$\widetilde{f_{jm}} \equiv \frac{\pi_{im}(c_{im}, b_{jm}) - \pi^e_{im}(c_{im}) + (v_j)(1/2 + v_i/2)(\pi_{in}(e_{in}, d_{jn}) - \pi^e_{in}(e_{in}))}{\pi^h_{im}(c_{im}) - \pi^{\min}_{im}(c_{im}) + (v_j)(1/2 + v_i/2)\left(\pi^h_{in}(e_{in}) - \pi^{\min}_{in}(e_{in})\right)} + \sigma_m \tag{1}$$

The function $\widetilde{f_{jm}}$ represents how individual im judges individual jm for his actions and intentions with himself, with his actions and intentions with other individuals, and for the actions and intentions of other members of his group.

The first two terms of the numerator and the denominator represent the beliefs of an indivudal of what's happen in his own game. These terms are the equivalent of Rabin's definition for how kind an individual believes is another individual. The last two terms of the numerator and the denominator represent the beliefs of an individual of what's happen in the other game.

By choosing to define $\widetilde{f_{jm}}$ as only one fraction I am representing that an individual cares about the magnitud of kindness for each player. If the stakes of one game are higher than the other, then a player would be giving a higher material payoff to the other player when kind, and therefore he would be thought as much kinder person. In this case, then the two terms from that game will grow with respect to the terms for the other game, and $\widetilde{f_{jm}}$ would represent that an individual thinks much better of a person that is very kind than to somebody that is only a shligtly kind.

I could have defined $\widetilde{f_{jm}}$ as the sum of two fractions, one that represents the beliefs of a player about how kind is the player he is playing with and other that represents his beliefs about how kind is the other member of the group of the player he is playing with. However, by normalyzing both terms before adding them this function would represent that an individual do not care about the magnitud of kindness or unkindness for each player, but only its sign.

The choice for $\widetilde{f_{jm}}$ is important, as some of my results depend on its form. However I think my definition of $\widetilde{f_{jm}}$ is more realistic this way.

The importance player im gives to what happen in game n depends on the term $(v_j)(1/2 + v_i/2)$. I include the term v_j to represent that as the affiliation between two player grow large, so it grows how other individuals relate their actions and intentions. The term $1/2 + v_i\ /2$ represents that the person that is making the judgement, in this case player im, cares more of the other game if somebody close to him plays in that game. I add 1/2 because I want to represent that even if an individual does not have anybody related to him in the other game, he may still care on that game.

As the term $(v_j)(1/2 + v_i/2)$ becomes smaller, player im pays less attention to what's happen in the other game when he makes his judgement about player jm. When $(v_j)(1/2 + v_i/2) = 0$, then player jm is not related to any player in game n and individual im cannot use any information from game n to judge him. In this case the equation $\widetilde{f_{jm}}$ reduces to the same equation used by Rabin.

Once I have completed the definition of kindness and the belief of kindness I can define an individual's utility function:

$$U_{im} = \pi_{im} + \widetilde{f_{jm}}(1 + f_{im})$$

Definition 6: The strategies $a_{im} \in S_{im}$ for all $i, j \in [1, 2]$, and $m, n \in [1, 2]$, where $i \neq j$ and $m \neq n$ are a Group Fairness Equilibrium if:

1) $a_{im} \in \arg\max_{a_{im} \in S_{im}} U_{im}$
2) $a_{im} = b_{im} = c_{im} = d_{im} = e_{im}$

The model captures the observation that individuals treat better those individuals that belong to their own groups, by including a variable σ_m, where σ_m is positive when player im and player jm belong to the same group, zero otherwise. In this case, individuals would have a greater utility if they are kinder to somebody from their same group, specially if they belong to a group whose members are close.

		Father of Player 1	
		Cooperate	Defect
Player 1	Cooperate	$4x, 4x$	$0, 6x$
	Defect	$6x, 0$	x, x

Example 1

Rabin shows that in the Prisoners Dilemma the cooperative outcome exists for low values of x ($x \leq 1/4$). In my model cooperation can be sustained for higher values of x if both players belong to the same group, as in example 1, where father and son are faced each other. In the case that $v_m > 0$, the equilibrium where both players cooperate exists if $x \leq 1/4 + v_m/2$, that is, individuals of the same group can cooperate for higher values of material payoffs. Additionally, if $v_m > 1/2$, the equilibrium where both players play defect does not exists for low values of x ($x < v_m - 1/2$), that is, individuals that belong to the same group will always cooperate for small material payoffs.

The model also captures the idea that individuals care not only about how kind other individuals are to them, but how kind they are to other individuals. As result, the outcomes of different games for the same individual could be related. If an individual is unkind to a second individual, a third individual may think of him as unkind and may be unkind to him in response.

		Player 2	
		Cooperate	Defect
Father	Cooperate	$4x, 4x$	$0, 6x$
	Defect	$6x, 0$	x, x

Example 2a

Player 2

		Cooperate	Defect
Son	Cooperate	$4x, 4x$	$0, 6x$
	Defect	$6x, 0$	x, x

Example 2b

Example 2

In example 2, the outcome in game 1 and game 2 are related. In the case that v_i (the relation between father and son) is close to one, the father would think that player 2 is not very kind if player 2 plays defect with his son, even if he plays cooperate with himself. In this case the outcomes would often be (cooperate, cooperate) for both games or (defect, defect) for both games and the equilibrium where player 2 and the father play cooperate and player 2 and the son play defect does not exist but for small values of x.

My model also captures the idea that individuals may be kind to individuals that belong to groups they like and treat badly individuals that belong to groups they do not like. I assume that individuals form emotions of like or dislike for a group depending how member of those groups have treated them or to other members of their own groups. Also in this case, when members of two groups play against each other, the outcomes may be related.

		Player 2	
		Cooperate	Defect
Player 1	Cooperate	$4x, 4x$	$0, 6x$
	Defect	$6x, 0$	x, x

Example 3a

		Father of Player 2	
		Cooperate	Defect
Father of Player1	Cooperate	$4x, 4x$	$0, 6x$
	Defect	$6x, 0$	x, x

Example 3b

Example 3

In example 3, when v_i and v_j are close to one, if the father of player 1 plays deviate with the father of player 2, then player 1 would not be very happy with family 1 and would play deviate with player 2, for all but small values of x, even if player 2 is kind and plays cooperate with him. In this example the equilibrium where in one of the games they play cooperate and in the other deviate does not exists, but for small values of x.

Now I analyze if group fairness equilibrium makes possible outcomes that are not possible with fairness equilibrium or other type of equilibrium. For example, is it posible that a group fairness equilibrium exists where in the Prisonners Dilemma one player plays cooperate while the other plays defect?

		Player 2	
		Cooperate	Defect
Player 1	Cooperate	$16x, 4x$	$0, 6x$
	Defect	$24x, 0$	$4x, x$
	Game 1		

		Father of Player 2	
		Cooperate	Defect
Father of Player1	Cooperate	$4x, 16x$	$0, 24x$
	Defect	$6x, 0$	$x, 4x$
	Game 2		

Example 4

In example 4, for values of v_i and v_j close to one, a groop fairness equilibrium exists where in game 1 player 1 plays defect and player 2 plays cooperate and where in game 2 the Father of player 1 plays cooperate and the Father of player 2 plays defect. Even if player 1 treats badly player 2, player 2 would still think well overall of family 1, given that the father of player 1 is much more kind than his son is unkind.

2.3 Basic Results

In this section I give some general propositions for the case of four players that play two games, but before, I extend the definition of positive and negative outcomes for the case of two games.

Definition 5: a) An outcome is strictly positive for the case of two games if for $i = 1, 2$ and $m = 1, 2$, $f_{im} > 0$. b) An outcome is weakly positive if for $i = 1, 2$ and $m = 1, 2$, $f_{im} \geq 0$. c) An outcome is strictly negative if for $i = 1, 2$, $m = 1, 2$, $f_i < 0$. d) An outcome is weakly negative if for $i = 1, 2$, $m = 1, 2$, $f_i \leq 0$. e) An outcome is neutral if for $i = 1, 2$, $m = 1, 2$ $f_i = 0$. f) An outcome is mixed if for any $i = 1, 2$, $m = 1, 2$, where $i \neq j$ or $m \neq n$, $f_{im} f_{jn} < 0$.

Proposition 3: If an outcome A is a combination of strict Nash equilibrium in games 1 and 2, there is an $\overline{X}$ for which for all $X > \overline{X}$ A is a group fairness equilibrium. If A is not a combination of Nash equilibrium of games 1 and 2, there is an $\overline{X}$ for which for all $X > \overline{X}$ A is not a group fairness equilibrium.

Proposition 3 is a direct translation of Rabin's proposition 5 to group fairness. As the material payoffs increase, the importance of fairness considerations becomes smaller. As the material payoffs increase arbitrarily, eventually the material payoffs dominate fairness considerations and the group fairness equilibrium are the combination of Nash equilibria for both games.

Proposition 4: There is a value $\overline{X}$ for which for all $X > \overline{X}$, any game does not have a positive group-fairness equilibria.

Proposition 4 tell us that as the material payoffs grow large, the positive group fairness equilibria are eliminated and only the weakly negative and neutral group fairness equilibria are left. As the income increases, the material payoffs dominate the fairness considerations. Because individuals are maximizing their own material payoffs, other individuals would not think their are been kind and the positive fairness is eliminated.

Proposition 5: For any outcome that is either strictly mutual-max for both games or strictly mutual-min for both games, there exists an $\overline{X}$ for which for all $X < \overline{X}$ A is a group fairness equilibrium.

Proposition 5 is a direct translation of Rabin's proposition 3. As material payoffs approach to zero, the game is dominated by the fairness considerations. In the case that an outcome that is strictly mutual-max for both games, every player is playing a strategy that maximize the material payoffs of the other players and therefore they are being kind to each other. In this case nobody wants to change strategy since they want to be kind to each other in response. In the case that an outcome that is strictly mutual-min for both games, every player is playing a strategy that minimize the material payoffs of the other players and therefore they are being unkind to each other. In this case nobody wants to change strategy since they want to be unkind to each other in response.

Now I analyze the case where the material payoffs of one of the games change while the other is left constant. I define the playoffs of game one as function of

x and the payoffs of game two as a function of y as in figure 6. I analyze the case where y changes, but x keeps constant. I will assume in these propositions that v_i or v_j are positive, since if both were zero, it would be equivalent to two single separate games.

		Player 2	
		Cooperate	Defect
Player 1	Cooperate	$4x, 4x$	$0, 6x$
	Defect	$6x, 0$	x, x

Game 1

		Player 2	
		Cooperate	Defect
Player 1	Cooperate	$4y, 4y$	$0, 6y$
	Defect	$6y, 0$	y, y

Game 2

Example 6

Proposition 6: If game two has a strict Nash equilibrium that is a mutual-max outcome, then there is a $\overline{Y}$ for which for all $Y > \overline{Y}$ the sign of the group fairness equilibrium is the sign of the fairness equilibrium of game 1. If game two does not have any mutual-max outcome, there is a value $\overline{Y}$ for which for all $Y > \overline{Y}$ the whole game has a weakly negative group fairness equilibrium.

Proposition 6 tell us that in the case that game two has a strict Nash equilibria that are mutual-max, then the fairness equilibrium of game two is neutral, because each individual is playing the action that maximizes his own material payoffs, so the other player sees his action as neutral, and the group fairness of the whole game will be defined by the other game. In the case that the Nash equilibria are not mutual-max, as the material payoffs grow large the equilibrium of the game becomes weakly negative. Because the material payoffs of that game become large its fairness considerations tend to dominate those of the whole game. In example 6, as Y grows large, the only fairness equilibrium that exists is defect, defect that is strictly negative. As Y grows arbitrarily large, the material payoffs of game two are going to dominate the material payoffs of game one and eliminate any positive or neutral equilibrium.

Proposition 7: If game 1 does not have a mutual-max outcome, then as $Y \rightarrow 0$ individuals in game 2 are kind to each other only if individuals in game

1 are kind to each other and individuals in game 2 are unkind to each other if player in game 1 are unkind to each other.

Proposition 7 tell us that as the material payoffs of one game become arbitrarily small, their importance on group fairness is going to be reduced until the fairness considerations of the other game dominate the group fairness. In example 6, as X becomes small, individuals in game one cooperate only if individuals in game 2 also cooperate and they defect if individuals in game 2 defect.

3 Two period games

In this section I analyze the case where both games are played sequentially: one game is played first and then the other. I assume that players in the second game observe the outcome of the first game before they play. If players in the first game know that their actions can affect the outcome in game two, they may play differently in order to change the actions of the players of the second game. For example, if two sons and two fathers from two different families are playing with each other, we can think that both fathers will be nice in order to have good relations between both families and help their sons to be nice to each other.

Other than assuming that game 1 is played first and game 2 is played second I assume that there are no differences with respect to the case where both games are played simultaneosly and individuals utilities are the same. By keeping the same utility functions I am implicitly assuming that individuals in the second game do not take in consideration that individuals in the first period may be kind or unkind in order to influence their decisions in the second period. The difference of this case with the sequential games analyzed by Dufwenberg and Kirchsteiger (2004) is that in my model the simple structure of the game allow me to solve these games by backward induction.

Let h be a non terminal history that takes us to a subgame, where $h \in H \backslash Z$, H is the set of possible histories and Z is the set of terminal histories, $a_{im}(h)$ be a strategy for player im at history h.

Definition 6: The strategies $a_{im}(h) \in S_{im}(h)$ for all $i, j \in [1, 2]$, and $m, n \in [1, 2]$, where $i \neq j$ and $m \neq n$, are a Sequential Group Fairness Equilibrium if, for every non terminal history $h \in H \backslash Z$ we have:

1) $a_{im}(h) \in \arg\max_{a_{im}(h) \in S_{im}(h)} U_{im}$

2) $a_{im} = b_{im} = c_{im} = d_{im} = e_{im}$

4 Discrimination

Individuals see themselves as part of groups and they care if they are discriminated for belonging to those groups. In some groups, as it is the case of families, it is seen naturally that individuals treat better to the members of their own group. However, individuals do not like to be discriminated based on some groups, as race and ethnicity. In this section I include the emotion of disliking being discriminated in my model.

		Player 2	
		Opera	Boxing
Player 1	Opera	$2x, x$	$0, 0$
	Boxing	$0, 0$	$x, 2x$

Example 5

In example 4, if player one goes to the Opera, he would be angry if he believes that player two is going to Boxing in order to be unkind to him, however he would be even angrier if he believes that player two is unkind to him because of the group he belongs. That is, the belief of discrimination increases individuals sense of unfairness.

I assume that individuals have beliefs about what would other players would have played if they had belonged to their same group. If individual 1 believes that individual 2 would have treated him better if he were from the same group, he would feel discriminated and this would increase his sense of unfairness.

I define g_{jm} as individual $i's$ beliefs about individual j's hypothetical actions if i have been of the same group as j. For simplicity I define individual's dislike for being discriminated for the case of a single game between two players (from different groups), although the model can easily be extended for several games.

Definition 7: Player $i's$ belief about how kind is player j:

$$\widetilde{f}_j(b_j, c_i, g_j) \equiv \frac{\pi_i(c_i, b_j) - \pi_i^e(c_i) + \pi_i(c_i, b_j) - \pi_i(c_i, g_j)}{\pi_i^h(c_i) - \pi_i^{\min}(c_i)} \tag{2}$$

The last two terms of the numerator of equation 2 represent how much individuals detest to be discriminated.

Definition 8: The strategies a_{ij} for all $i \in N$, where $i \neq j$ are a Group Fairness Equilibrium if:

The strategies $a_i \in S_i$ for all $i, j \in [1, 2]$, where $i \neq j$ are a Discrimination Fairness Equilibrium if:

1) $a_i \in \arg\max_{a_i \in S_i} U_i$

2) $a_i = b_i = c_i = g_j$

In example 5, for values of $x \leq 2$, there is an equilibrium where both players play the same action if they are from the same group, but they play different outcomes if they are from different groups. In this case $\widetilde{f_j} \equiv -2$, given that individuals resent being treated different because the group they belong. In the case of Rabin's fairness equilibria, the maximum value of x for which individuals are unkind to each other is one. This means that an emotion of dislike for being discriminated increases the range for which a negative outcome is possible.

5 Application: Firms giving to Charity

Rabin shows that when individuals care about fairness, a monopoly cannot extract all consumer's surplus, given that individuals see this as an unfair practice and retaliate by not buying its product. However, if consumers care not only about how the monopoly treats them, but how it treats other individuals, a monopoly could improve its public image by being kind to a group of individuals in need or a charity.

Rabin solves an example wher a consumer wants to buy one unit of a product from a monopoly. The consumer's valuation of the product is given by β, while the marginal cost for the monopoly is given by c. Simultaneosly, the monopoly chooses the price and the consumer chooses a reservation price r, above which he is not willing to pay. If the monopoly prices at $p = r = z$ (charging the highest price the consumer is willing to pay), the consumer's belief in the fairness of the monopoly would be given by:

$$\widetilde{f_M} = \frac{c - z}{2(\beta - c)}$$

As long as the monopoly is pricing above its marginal cost, this function is always negative. This is because the monopolist is choosing the price that extracts as much surplus as possible from the consumer, given the consumer's refusal to buy at a price higher than z. This means that the consumer will always see a price higher than the cost of the product as an unfair practice from the monopoly and would prefer to retaliate (by not buying) if his material gains from buying the product are too low. This forces the monopoly to reduce its price below v.

If the monopoly can improve how kind the consumer think it is, then the monopoly would be able to charge a higher price to the consumer without being retaliated. In my model the monopoly can acomplish this by being kind to another player, let's say an individual that is in need of its product, but does not have the resources to pay for it.

I extend Rabin's example by adding a another game in which the monopoly can improve its image by giving its product for free to this individual in need. Let's say that the player in need values the product at x and cannot pay any price for it. With respect to this consumer, the monopoly has two options, give him the product for free or not. If the monopoly gives the product for free to the individual in need, it incurs in a cost of c, but the consumer would think better of it, and if the monopoly does not give the product for free to the individual in need, the consumer would think worst of it. For example, people thought that it was unfair that pharmaceutical companies did not provide cheap drugs to people with aids in Africa. After a public backlash, the pharmaceutical companies gave the drugs for free, improving their image.

In the first period the monopoly decides if it gives the product for free to the individual in need and in the second period the monopoly sells its product to the consumer. I solve the problem by backward induction. In the second period, the consumer values the product at β and chooses a reservation price above which he is not willing to pay and the monopoly simultaneosly chooses the price. If $p \geqslant r$, the consumer buys the product.

If the monopoly gives the product for free to the individual in need in the first period and if the monopoly prices at $p = r = z$ in the second period, the consumer believes that the kindness of the monopoly is the following:

$$\widetilde{f_{Mk}} = \frac{(c-z)/2 + (1/2 + v_i/2)(x - x/2)}{\beta - c + (1/2 + v_i/2)(x - 0)} \tag{3}$$

where v_i is the relation between the consumer and the individual in need. I assume that the consumer does not have any special reason to be kind to the monopoly, like if the owner of the monopoly and the consumer belong to the same group, and therefore I assume that $v_j = 0$. If the monopoly does not give the product for free to the individual in need in the first period the consumer believes that the kindness of the monopoly is the following:

$$\widetilde{f_{MNk}} = \frac{(c-z)/2 + (1/2 + v_i/2)(0 - x/2)}{\beta - c + (1/2 + v_i/2)(x - 0)}$$

If the consumer buys the product from the monopoly, at a price lower than his valuation, he would not been kind to the monopoly, given that he is doing an action that improves his own material payoffs. However, if he does not buy the product (by choosing a reservation price higher than the price of the monopoly) he would been unkind, given that he is sacrificing his material payoffs in order to punish the monopoly. The consumer's utility from consuming a product from the monopoly if it gives to the individual in need is given by:

$$U_c = \beta - z + \widetilde{f_{Mk}}(1 + 0)$$

and the consumer's utility from consuming a product from the monopoly if it does not give to the individual in need is given by:

$$U_c = \beta - z + \widetilde{f_{MNk}}\,(1 + 0)$$

If the consumer do not buy the product of a monopoly that has given to the individual in need then their utility would be given by:

$$U_c = 0 + \widetilde{f_{MNk}}\,(1 - 1) = 0$$

In the second period, the monopoly charges a price that makes indifferent the consumer between consuming and not consuming. The maximum price the monopoly is able to charge without the consumers be willing to retaliate is:

$$p = \frac{2\beta^2 - 2\beta c + c + (1/2 + v_i/2)x(2\beta + 1)}{1 + 2\beta - 2c + 2(1/2 + v_i/2)x}$$

and if the monopoly does not give to the individual in need, the maximum price that the monopoly is able to charge is:

$$p = \frac{2\beta^2 - 2\beta c + c - (1/2 + v_i/2)x(2\beta + 1)}{1 + 2\beta - 2c + 2(1/2 + v_i/2)x}$$

and therefore the monopoly is able to increase its price if it gives the product for free to the individual in need. If the cost of donating the product to the consumer in need is lower than the extra revenue it brings, that is, if $c < \frac{(1/2+v_i/2)x(2\beta+1)}{1+2\beta-2c+2(1/2+v_i/2)x}$, the monopoly will give the product for free.

We have to note that the price the consumer is willing to pay can be higher than his valuation of the product. If $(1/2+v_i/2)(x-x/2) > (c-z)/2$, equation 1 is higher than zero, and therefore the consumer thinks that the monopoly not only is kind to the individual in need, but overall is kind. In this case, individuals would be willing to pay a higher price than their valuation of the product in order to be kind in response to the monopoly. For example, many people buy the cookies that are sold by the girl scouts at a higher price than their valuation of the cookies because they want to help the organization as much as they want to eat the cookies.

6 Conclusions

In this paper I introduced to game theory the emotions of fairness between groups by extending Matthew Rabin's model of fairness equilibrium to groups of individuals. There is a number of possible extensions to this work. First, in the real world the majority of interactions is repeated and therefore, a repeated game version of group fairness would bring new and more realistic results. In international relations, countries construct their relations little by little, increasing their trust with kind actions over time. It is reasonable to think that if an individual or a group of individuals are kind or unkind once and again and again, the feeling of kindness or unkindness would grow larger over time. It will be interesting to extend my model by defining a function of kindness that can increase or decrease over time. I believe that by doing this, group fairness would help reduce the large set of possible equilibria that exists in repeated games.

Second, I believe that group fairness can be very useful to help explain other phenomena of group interactions. For example: a) hatred between ethnic groups due to a conflict, b) nationalism, where individuals treat better firms or individuals from their own country, and c) charity, where it is observed that individuals donate more money to the groups they belong.

Third, in this paper I assumed the value of v_i (the closeness of the members of a group) to be fixed. However, the closenest to the members of a group depends on the actions and intentions of the members toward each other and the actions and intentions of other individuals toward the members of the group. For example, if the members of a family are unkind toward each other, we should not expect it to be as close as a family whose member are kind toward each other. And it has been observed that unkindness toward the members of a group tend to bring them toghether. Extending my model by endogeneizing the closeness of groups would help explainning many phenomena like the increase of religion fervor or nationalism after wars or intherethnic conflicts.

7 Bibliography

Akerlof, George and Kranton R. “Economics and Identity,” *Quarterly Journal of Economics,* 2000.

Dufwenberg, Martin and Kirchsteiger, George. “A Theory of Sequential Reciprocity,” *Games and Economic Behavior.* Vol. 47, 2004, pp. 268-298.

Dufwenberg, Martin and Kirchsteiger Georg. “Reciprocity and Wage Undercutting,” *European Economic Review* 44 (2000), 1069-78.

Fehr, Ernst and Schmidt, Klaus M. “A Theory of Fairness, Competition, and Cooperation,” *Quarterly Journal of Economics.* 1999.

Fiske, Susan T. “Stereotyping, Prejudice, and Discrimination,” *Handbook of Social Psychology*, Chapter 25.

Geanakoplos, John, Pearce, David and Stacchetti, Ennio. “Psychological Games Sequential Rationaliry,” *Games and Economic Behavior.* March 1989, 1, 60-79.

Rabin, Matthew. “Incorporating Fairness into Game Theory and Economics,” *The American Economic Review*, Vol. 83, No. 5. (Dec. 1993), pp. 1281-1302.

Rabin, Matthew, “Fairness in Repeated Games,” Berkeley Department of Economics Working Paper No. 97-252. January 1997.

8 Appendix

Proof of proposition 1

For contradiction: If (a_i, a_j) is a fairness equilibria that is not a Nash equilibria, then there is another strategy that gives higher material payoffs to at least one player. If X grows arbitrarily large, then these material difference would grow arbitrarily large and would dominate any material payoffs. Therefore, at least one player would deviate and (a_i, a_j) cannot be a fairness equilibrium.

Proof of proposition 2

As X increases arbitrarily, the material payoffs increase. However, the fairness payoffs are independent of X and therefore eventually the material payoffs dominate the fairness payoffs and the fairness equilibrium becomes the Nash equilibrium. Because players are maximizing their own material payoffs and the other individual would not think that they are being kind. Therefore there is not a positive fairness equilibrium.

As X becomes large, individuals would play a Nash equilibrium. If the Nash equilibrium is not mutual-max outcome, then at least one individual is playing a strategy that is not maximizing the other utility and therefore he is unkind to him. By Rabin's proposition 1, the other individual also would be unkind to him (at least weakly) and the equilibrium would be weakly negative fairness equilibrium.

Proof of proposition 3

This is the same proof of Rabin's proposition 5, but extended for group fairness. Group-fairness gains or losses are independent of X. However, material payoffs are proportional to X and as X becomes large, the difference between the equilibrium strategies and the non equilibrium strategies becomes large. Therefore, as X grows arbitrarily large, group fairness gains or losses become unimportant with respect to the material payoffs and the strategies A becomes a strict best reply.

If A is not a Nash equilibrium then there is at least one other strategy that improves the material payoffs for at least one player. As X becomes arbitrarily large the material payoffs eventually dominate the group fairness payoffs and another strategy eventually improves for at least one player with respect to A.

Proof of proposition 4

As X becomes large, the group-equilibrium for each game are Nash equilibrium and therefore each individual is maximizing their own payoff. This eliminates that other individuals think that their are kind to them and their group

Proof of proposition 5

Proposition 5 is a direct translation of Rabin's proposition 3 As $X \rightarrow 0$, the material payoffs goes to zero and it is dominated by the group fairness payoffs. If the outcome is strictly mutual-max for both games then both players are being kind to the players of the other group and therefore they are maximizing the group fairness payoffs and it is a group-fairness equilibrium. If the outcome is strictly mutual-min for both games then both players are being unkind to the players of the other group and therefore they are maximizing the group fairness payoffs and therefore it is a group-fairness equilibrium.

Proof of proposition 6

If the game has a strict Nash equilibrium it would become part of the group-fairness as X grows arbitrarily large. Because it is a mutual-max outcome, then it is maximizing each other outcomes and therefore it is not being unkind to them. But because they are maximizing each other payoffs they are neither been kind to each other. Therefore the fairness of game two is zero and the group-fairness of the whole game is defined by game one.

As Y becomes large, the group fairness equilibrium of the game is a Nash equilibrium for game 2 by proposition 1. Because the Nash equilibrium is not a mutual-max outcome, then at least one of the players is not maximizing the other players material payoffs and then he is being unfair to him. By Rabin's proposition 2 we know that fairness equilibria are symmetric and that the other player will be unfair in response (at least weekly unfair). As Y becomes large with respect to X, the unfairness of game 2 dominates over any result in game 1 and the group-fairness becomes weakly negative.

Proof of proposition 7

As Y becomes small, the material payoffs of game 1 dominate equation 1. As the material payoffs of game 2 approach zero, the group fairness of the game is proportional to the fairness equilibria for game 1.

8.1 Appendix B

In this appendix I extend the definition of group fairness to more than two games and four players.

I assume that individuals think that the kindness of other individuals is simply the average of how kind these individuals are with other individuals they play with (including himself). For this, I define a variable $Kindness_j^i$ that represent the overall perception of kindness of individual j from the point of view of i :

$$Kindness_j^i = \sum_{k=1}^{N} \phi_{ik} \widetilde{f_{jk}^i}(b_{jk}^i, c_{jk}^i)$$

Additionally, individuals tend to form affective emotions from groups which members have been kind or unkind. I model this by assuming that individuals think that the kindness of a group is the average of the kindness they observe in the individuals that belong to each group.

$$KindnessGroup_j^i = \sum_{x \in N} \sum_{y \in A_j} \phi_{xy} \widetilde{f_{xy}^i}(b_{xy}^i, c_{xy}^i)$$

I normalize this by dividing over the difference between the average of the maximum and the minimum possible material well-being:

$$MaxKindness_j^i = \sum_{k=1}^{N} \phi_{ik} \left(\pi_{kj}^h(c_{jk}^i) - \pi_{kj}^{\min}(c_{jk}^i) \right)$$

$$MaxKindnessGroup_j^i = \sum_{x \in N} \sum_{y \in A_j} \phi_{xy} \left(\pi_{xy}^h(c_{xy}^i) - \pi_{xy}^{\min}(c_{xy}^i) \right)$$

Definition 3: Individual $i's$ belief of how kind is player j is given by:

$$\theta_j^i \equiv \frac{Kindness_j^i + v_i KindnessGroup_j^i}{MaxKindness_j^i + v_i MaxKindnessGroup_j^i} + h_{ij} \tag{4}$$

where h_{ij} is a positive constant when i and j belong to the same group, zero otherwise. h_{ij} allows me to represent individuals' preference to help other individuals of their own group. v_i is a parameter that represents how important is the behavior of the members other group for individual i (likely depends on factors such as the level of education). I limit the size of h_{ij} to be no greater than one.

Once I have completed the definition of the kindness functions I can define an individual's utility function when he plays with another individual:

$$U_{ij} = \pi_{ij} + \theta_j^i(1 + f_{ij})$$

Given that an individual can have interaction with more than one player, I define individual $i's$ total utility as the addition of his utility functions with those individuals:

$$U_i = \sum_{j=1}^{N} U_{ij}$$

Definition 4: The strategies $a_{ij} \in S_{ij}$ for all $i, j \in N$, where $i \neq j$ are a Group Fairness Equilibrium if:

1) $a_{ij} \in \arg\max_{a_{ij} \in S_{ij}} U_i$

2) $a_{ij} = b_{ij}^k = c_{ij}^k$