Modelling Fairness and Reciprocity

References

- Rabin, M.: Incorporating Fairness into Game Theory and Economics, *American Economic Review*, 83 (1993), 1281-1302.
- Fehr, E. and K. M. Schmidt: A Theory of Fairness, Competition, and Cooperation, *Quarterly Journal of Economics*, 114 (1999), 817-868.
- Bolton, G. E. and A. Ockenfels: ERC: A Theory of Equity, Reciprocity, and Competition, *American Economic Review*, 90 (2000), 166-193.
- Charness, G. and M. Rabin: Understanding Social Preferences with Simple Tests, *Quarterly Journal of Economics*, 117 (2002), 817-869.
- Binmore, K. and A. Shaked: Experimental Economics: Where Next?, *Journal of Economic Behavior & Organization*, 73 (2010), 87-100.

Background

- The standard model, using standard game theory, fails to predict many empirical outcomes.
- This is not a failure of game theory.
- This could be due to:
  - either people not engaging in strategic thinking in practice (meaning that game theory is an inappropriate analytical technique), or
  - the situations are not modelled realistically, in particular, by failing to incorporate social preferences into utility functions.

Objectives of Modelling

1. A good model must be able to explain the results of a wide variety of different games in terms of endogenous parameters. Thus, a good model is parsimonious and does not resort to *ad hoc* complications.

2. A good model must rely on known psychological mechanisms. It does not necessarily mean that the model must explicitly incorporate psychological or neurological processes in its mathematical formulation, but it must, at least, be compatible with such processes.

Inequality Versus Inequity

- Inequality is a neutral term and implies no value judgement.

- Inequity is a value-laden or normative term involving the subjective notion of fairness/justice.

The Fehr-Schmidt Model

- Fehr and Schmidt (1999) propose a model, which assumes that, in addition, to purely selfish subjects, some subjects will also dislike inequitable outcomes.

- Such inequity could arise from being either worse off or better off than others. An individual's utility function therefore depends not only on their own monetary payoff, but on differences between this payoff and those of others.

The Fehr-Schmidt Model (Cont.)

- Given an allocation $(x_1, x_2, \ldots, x_n)$ a person's utility is

  $U_i(X) =$
  $x_i - \frac{\alpha_i}{n-1} \sum_{k \neq i} max(x_k - x_i, 0) - \frac{\beta_i}{n-1} \sum_{k \neq i} max(x_i - x_k, 0)$
  where $0 \leq \beta_i < 1$ and $\beta_i \leq \alpha_i$.

- Thus, α_i is a measure of player i's aversion to disadvantageous inequality (i.e. envy), and β_i is a measure of player i's aversion to advantageous inequality (i.e. guilt).

The Fehr-Schmidt Model (Cont.)

Three further assumptions are involved.

1. $0 \leq \beta_i < 1$: The lower boundary is questionable and assumes that players do actually suffer from advantageous inequality, rather than feeling a benefit from it. Fehr and Schmidt admit the possibility of status-seeking players with negative values of β_i. The upper boundary $\beta_i = 1$ can be interpreted as meaning that a player is willing to throw away a dollar in order to reduce his relative advantage over the other player, which again seems unlikely.
2. $\beta_i \leq \alpha_i$: This means that players suffer less from advantageous than disadvantageous inequality.
3. The population of players is heterogeneous, meaning that different players have different values of α and β.

The Fehr-Schmidt Model (Cont.)

- The Fehr-Schmidt model derives a number of conclusions regarding, for instance, different variations of the Public Goods games. While the standard model predicts complete free riding with 0 contributions, and no willingness to engage in costly punishments, the model predicts that both contributions and punishments will occur if players are sufficiently envious and guilty.
- The model derives specific conditions regarding
  1. the conditions under which people free ride (i.e. $\beta_i \leq 1 - m$, where m is the marginal return to the public good),
  2. how many free riders k it takes to cause everyone to free ride (i.e. $k > \frac{m(n-1)}{2}$), and
  3. how much guilt and envy are necessary for an equilibrium to emerge.

Strengths of the Fehr-Schmidt Model

1. Simplicity: The model is simple in form, in terms of the number of parameters involved, and the linearity of the function.

2. Robustness: The model explains the wide variations in results that are observed in different games.

- However, there are also criticisms. For instance, Binmore and Shaked (2010) claim that Fehr and Schmidt focus on data and predictions that support their model and ignore data and predictions that do not, thus engaging in cherry-picking of results.

The Bolton-Ockenfels Model

- Bolton and Ockenfels (2000) proposed the ERC model, referring to equity, reciprocity and competition. It is similar to the Fehr and Schmidt model in many respects, since players care about their own payoffs and their relative share.
- The situation can be modelled mathematically as follows.
- $U_i(x_i) = U(x_i, \frac{x_i}{\sum x_j})$
- The Bolton-Ockenfels model, like the one by Fehr and Schmidt, uses game-theoretic analysis to derive specific conclusions regarding equilibria in various games.
- For instance, in the Ultimatum game, the model predicts that respondents will reject zero offers all the time, and that the rejection rate will fall with increasing percentage offers. These predictions are empirically confirmed.

Bolton-Ockenfels Versus Fehr-Schmidt

1. The Bolton-Ockenfels model is concerned with relative shares, whereas the Fehr-Schmidt model is concerned with absolute differences.
2. The Bolton-Ockenfels model only makes a comparison between an individual's payoffs with the average payoff of all the other players. It does not compare each player's payoffs with the maximum and minimum of the other payoffs, like the Fehr-Schmidt model.
3. The Bolton-Ockenfels model proposes a symmetrical attitude towards inequality, where guilt and envy are equal in force (i.e., $\alpha_i = \beta_i$), whereas the Fehr-Schmidt model proposes that envy is stronger than guilt.

Bolton-Ockenfels Versus Fehr-Schmidt (Cont.)

- In all three respects, the Fehr-Schmidt model is superior to the Bolton-Ockenfels one.
- Assume the payoff allocation is given by $(x, x - \epsilon, x + \epsilon)$. According to the Bolton-Ockenfels model the preferences of the first player should be independent of ϵ, since the sum of the payoffs will be constant and therefore the share of the first player is not affected. However, the Fehr-Schmidt model indicates that as ϵ increases, envy of the third player's payoff and guilt regarding the second player's payoff both increase, causing the first player's utility to fall.
- A study by Charness and Rabin (2002) has confirmed this prediction.

The Rabin Model

- Rabin (1993) proposed a reciprocity model that rests on the following statement:

  If somebody is being nice to you, fairness dictates that you be nice to him. If somebody is being mean to you, fairness allows - and vindictiveness dictates - that you be mean to him (p. 1281).

- The model is a two-player one, in which utilities depend upon beliefs.
- Player i's strategy a_i depends on his belief about the strategy of the other player b_j, and his belief about player j's belief regarding player i's strategy c_i. A similar description can be applied to player j's strategy.
- Two important constructs can then be determined.

THE RABIN MODEL (CONT.)

1. Player i's kindness to player j, which is expressed as follows: $f_i(a_i, b_j) = \frac{\pi_j(b_j,a_i)-\pi_j^{fair}(b_j)}{\pi_j^{max}(b_j)-\pi_j^{min}(b_j)}$.
2. Player i's perception of player j's kindness, which is written as follows: $f_j(b_j, c_i) = \frac{\pi_i(c_i,b_j)-\pi_i^{fair}(c_i)}{\pi_i^{max}(c_i)-\pi_i^{min}(c_i)}$.

- Rabin then assumes that player i's social preferences are given by a three-component utility function: $U_i(a_i, b_j, c_i) =$ $\pi_i(a_i, b_j) + \alpha f_j(b_j, c_i) + \alpha f_j(b_j, c_i) \times f_i(a_i, b_j)$
- An equilibrium for the model can then be derived on the basis that players maximize social utilities, assuming rational expectations (i.e., $a_i = b_j = c_i$). Thus, beliefs about the other player's strategy are correct, and that beliefs about the other player's beliefs are also correct.

Example - The Prisoner's Dilemma

	Cooperate	Defect
Cooperate	4, 4	0, 6
Defect	6, 0	1, 1

- Cooperate/Cooperate: $U_i = 4 + \alpha\frac{4-2}{4-0} + \alpha\frac{4-2}{4-0} \times \frac{4-2}{4-0} = 4 + 0.5\alpha + 0.25\alpha = 4 + 0.75\alpha.$
- Cooperate/Defect: $U_i = 0 + \alpha\frac{0-2}{4-0} + \alpha\frac{0-2}{4-0} \times \frac{6-3.5}{6-1} = 0 - 0.5\alpha - 0.25\alpha = 0 - 0.75\alpha.$
- Defect/Cooperate: $U_i = 6 + \alpha\frac{6-3.5}{6-1} + \alpha\frac{6-3.5}{6-1} \times \frac{0-2}{4-0} = 6 + 0.5\alpha - 0.25\alpha = 6 + 0.25\alpha.$
- Defect/Defect: $U_i = 1 + \alpha\frac{1-3.5}{6-1} + \alpha\frac{1-3.5}{6-1} \times \frac{1-3.5}{6-1} = 1 - 0.5\alpha + 0.25\alpha = 1 - 0.25\alpha.$

Example - The Prisoner's Dilemma (Cont.)

	Cooperate	Defect
Cooperate	$4 + 0.75\alpha$, $4 + 0.75\alpha$	$0 - 0.75\alpha$, $6 + 0.25\alpha$
Defect	$6 + 0.25\alpha$, $0 - 0.75\alpha$	$1 - 0.25\alpha$, $1 - 0.25\alpha$

- The payoff for cooperation given that the other player cooperates is greater than the payoff for defection if $4 + 0.75\alpha > 6 + 0.25\alpha$, i.e., if $\alpha > 4$.
- The implication here is that if people have sufficiently strong social preferences related to fairness there will be two Nash equilibria in pure strategies: cooperate/cooperate and defect/defect.