Theories of Social Preferences

- Other-regarding behavior in the field
- Other-regarding behavior in experiments
  - Trust game
  - Third party punishment game
- Facts consistent with self-regarding behavior
  - Responder Competition
- Objections to social preferences
  - Reputation and Retaliation
- Rabin's reciprocity model
  - Applications to the PD and the UG
- Levine's model of type-based reciprocity
- The Fehr-Schmidt model
  - Applications to the DG, UG, proposer competition, responder competition, three-person UGs, PG with and without punishment
  - Combining Fehr-Schmidt with the quantal response equilibrium approach
- The Bolton-Ockenfels model
  - The punishment motive in the Bolton-Ockenfels model
- Equity versus efficiency in social preferences (Quasi-Maximin preferences)
- The limits of inequity aversion
  - Punishment when income differences cannot be changed
  - The role of available, yet unchosen, alternatives
  - The role of intentions
- Inequity averion and reciprocity - problems of a consequentialist reciprocity model?
- Where do we stand – summing up
- Issues in field applications

Other-Regarding Behavior in the Field

- Mass demonstrations to overturn dictatorships (China, Eastern Europe)
- Tax morale – perceived fairness of the tax system may affect amount of tax evasion (Alm, Sanchez, de Juan in Kyklos 1995).
- Political support for the welfare state (strongly shaped by the perception of the recipients deservingness; Fong, Bowles & Gintis)
- Law enforcement depends on the perceived fairness of the law (Lind and Tyler 1988)
- Labor markets and organizations are riddled with social comparison processes
    - Firms rarely employ underbidders (Agell and Lundborg SJE 1995)
    - Firms rarely cut wages in case of an excess supply of labor (Agell & Lundborg 1995, Bewley 1999; Campbell & Kamlani 1997)
    - Krueger shows that when Firestone proposed wage cuts for new entrants the quality of Firestone tyres dropped significantly leading to several hundred accidents.

Why Firms don‘t employ underbidders (Agell & Lundborg SJE 1995)

Cited reason	Frequency	Percent
No vacancies, lack of meaningful work assignments	5	10.6
Not possible because of the collective wage agreement	5	10.6
Job seeker is considered to have inferior skills	16	34
Hiring an underbidder violates the firm´s wage policy, creates internal inequities	18	38.3
Other reasons	3	6.5
Total	47	100

Most important factors shaping wage policy (Agell & Lundborg 1995, 174 firms, multiple answers possible)

Cited factor	Frequency	Percent
Ability to pay	115	66
Relative wages	102	59
Centrally negotiated wage agreement	66	38
Labour market situation	36	21
Corporate policy	13	7
Other factors	15	9

Other-Regarding Behavior in Experiments

- Bargaining games
    - Offers above the prediction of the self-interest hypothesis
    - Disadvantageous counteroffers
    - Rejections
    - Positive transfers in the dictator game
- Public good games
    - Contributions higher than predicted by self-interest hypothesis
    - Conditional cooperation
    - Punishment of free riders
- Gift Exchange and Trust Games
    - Positive relation between wages and effort
    - Wages far above the competitive level
    - Many transfers in the trust game
- Third Party Punishment Game
    - Uninvolved third parties punish unfair dictators and defectors

The Trust Game

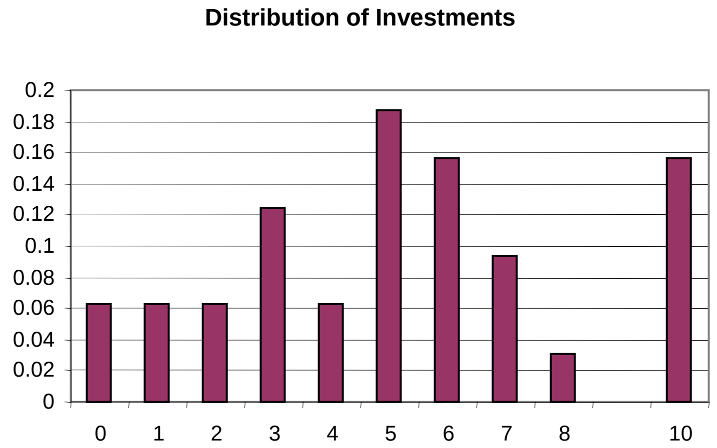
Berg, Dickhaut, McCabe (GEB 1995)

- Player 1 and 2 are endowed with $10.
- Player 1 decides how much of her $10 to transfer to player 2.
- Experimenter triples any amount sent.
- Player 2 is informed about 1's transfer and decides how much of the tripled transfer to send back.

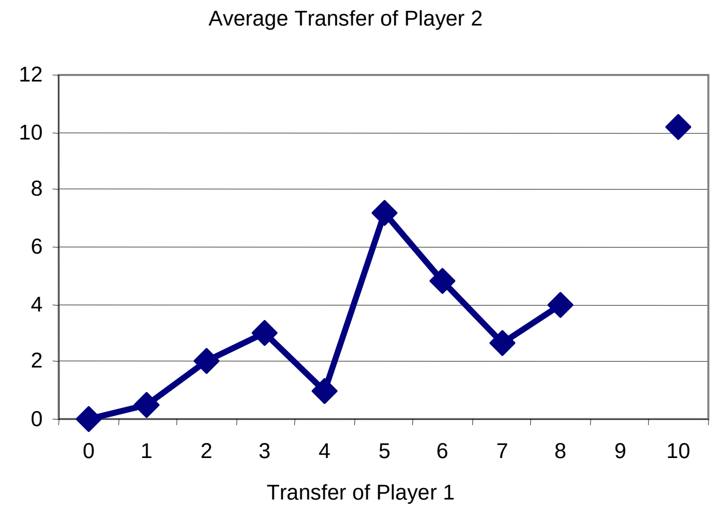
- Standard prediction
  - Player 2 sends back nothing.
  - Player 1 sends nothing.

Results
Berg, Dickhaut, McCabe (GEB 1995)

- Vast majority of player 1 make a transfer (investment).
- Reflects unconditional kindness (altruism) or trust



- Player 2 sends back money.
- Reflects unconditional kindness or trustworthiness.
- Investments of 5 and 10 are profitable.
- On average, player 1 gets back the amount that is sent.

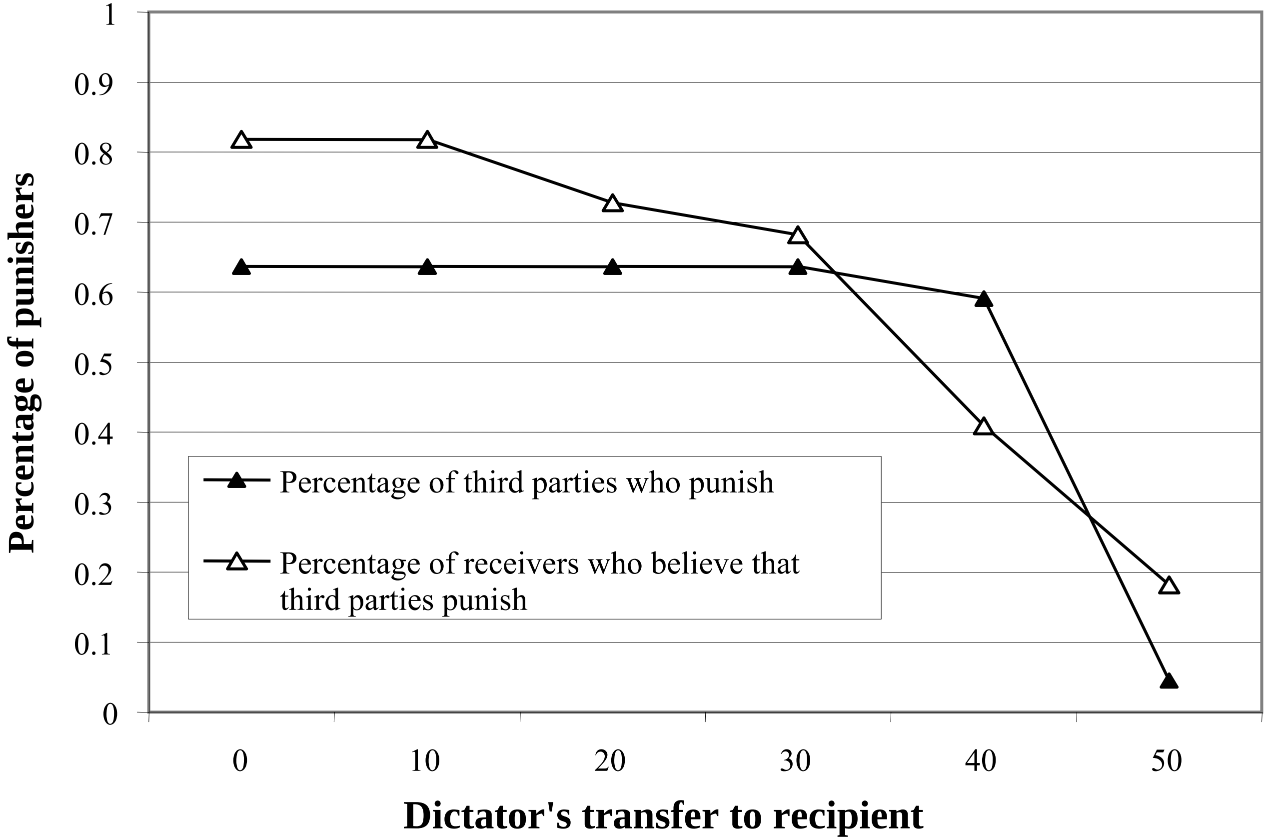


The Third Party Punishment Game

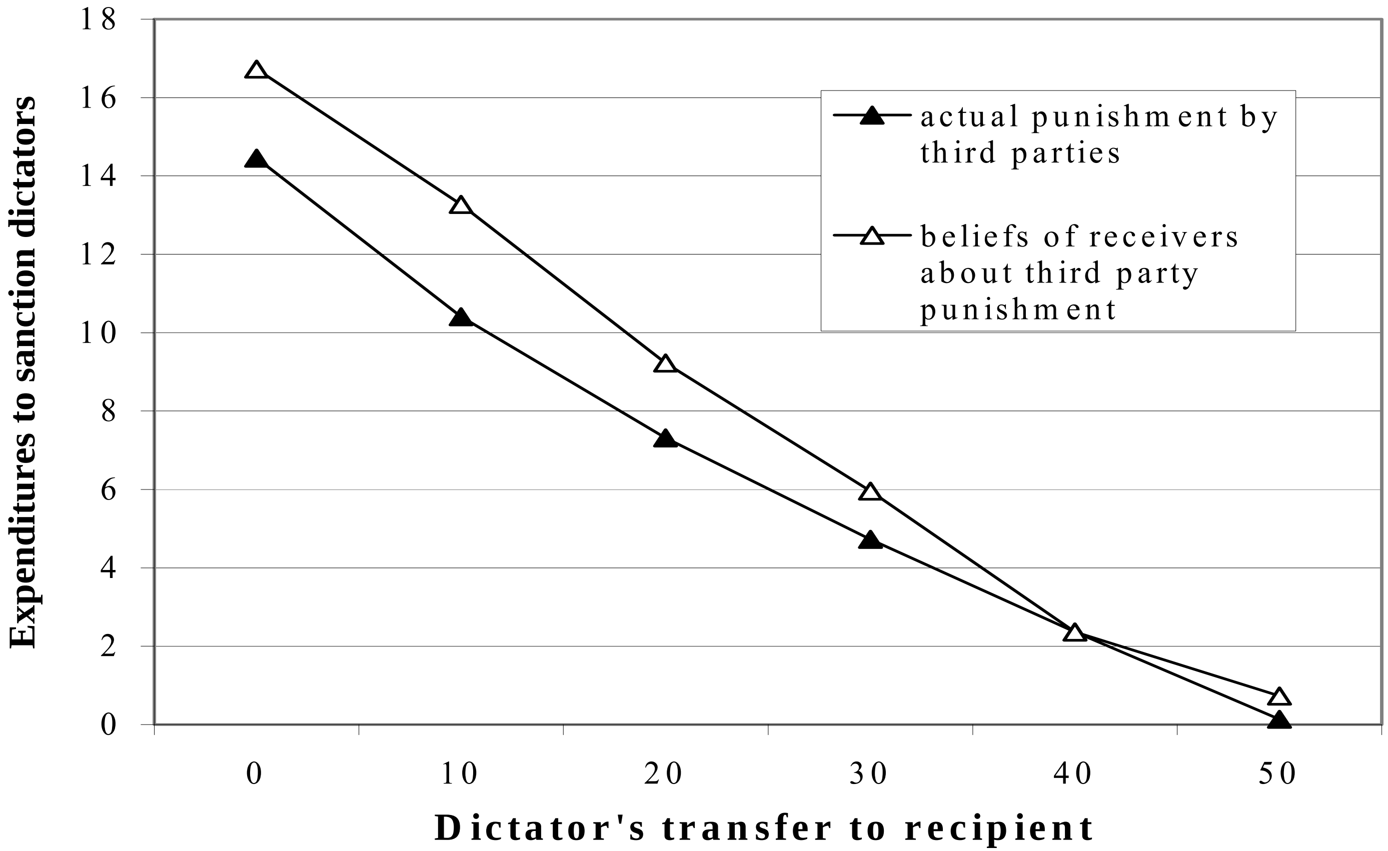
Fehr and Fischbacher (Evolution & Human Behavior, forthcoming)

- Third party punishment of dictators
  - Three players, A the dictator, B the recipient, C the third party.
  - A has endowment of 100, B of zero, and C of 50..
  - A can transfer money to B.
  - C observes what A did and can punish A (strategy method).
  - Every punishment point assigned to A costs C 1 point and A 3 points.
  - B's beliefs about the punishment of A by C are also elicited (strategy method).
- Results
  - Transfers below 50 are punished. Punishment is the higher the lower the transfer.
  - B's belief that transfers below 50 are punished and that punishment is higher the lower the transfer.
  - Third parties punish less than second parties.

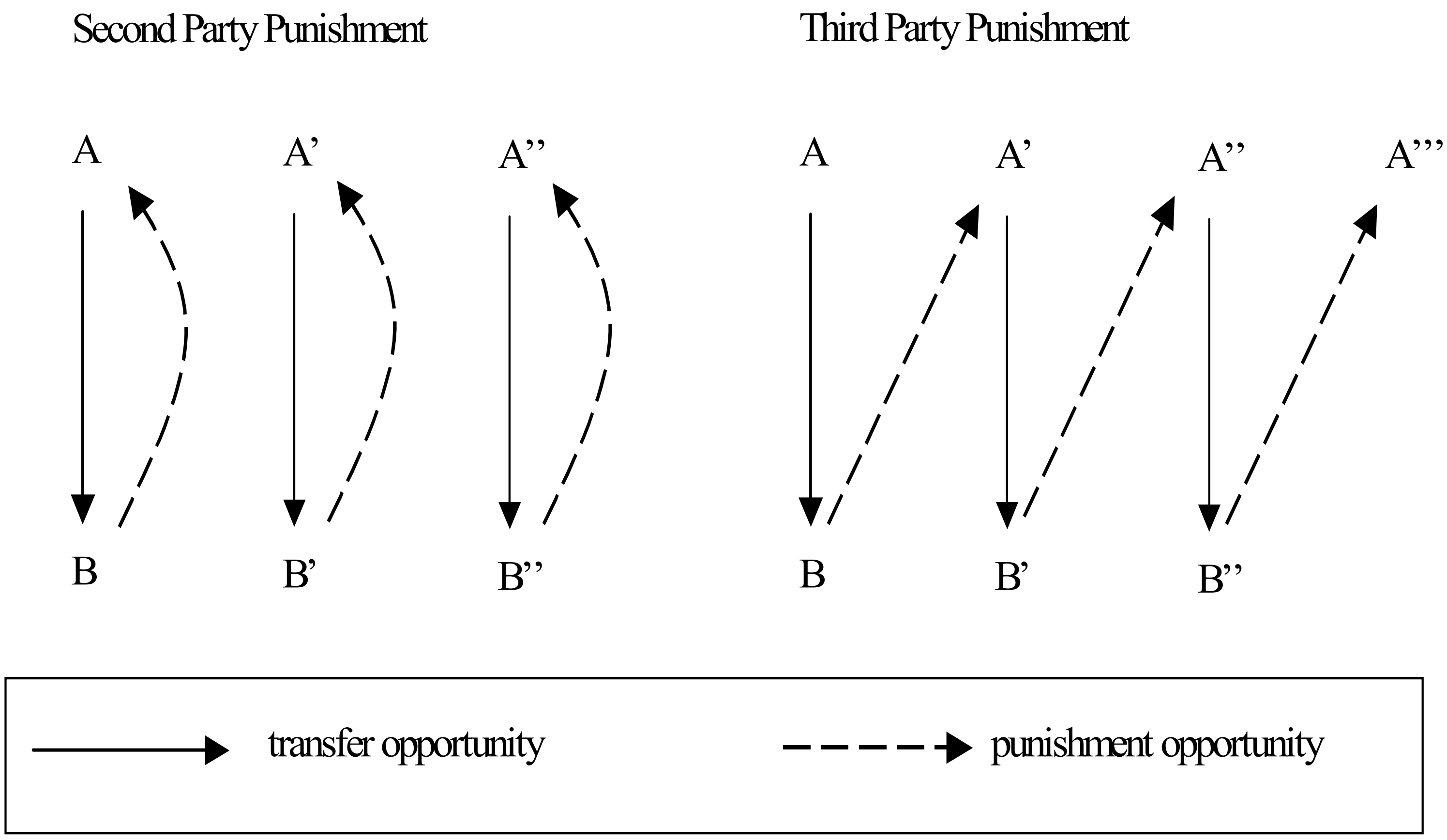
Percentage of third parties punishing the dictators
Source: Fehr&Fischbacher in EHB



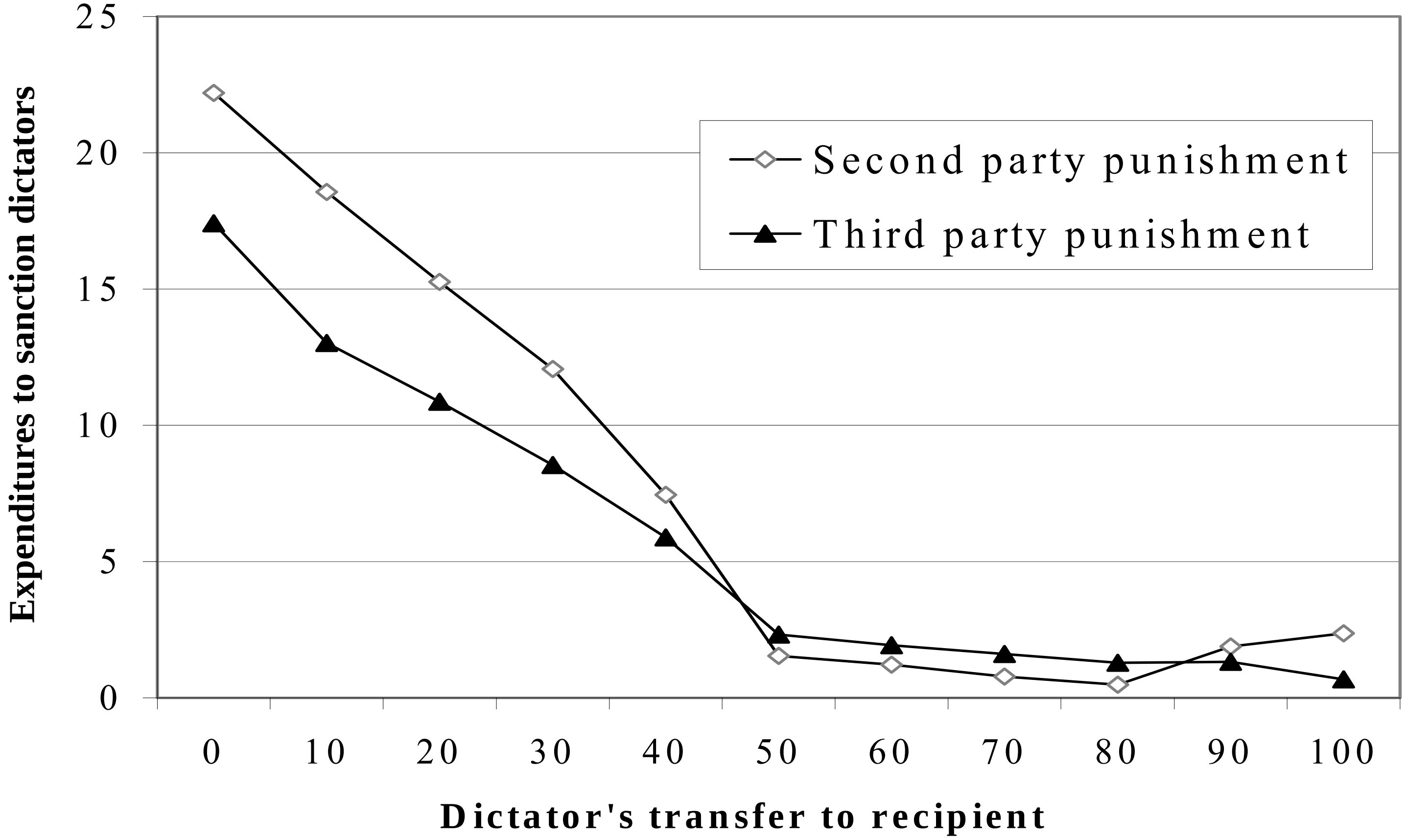
Pattern of third party punishment in dictator game
Source: Fehr & Fischbacher EHB



Comparison of 2nd and 3rd Party Punishment



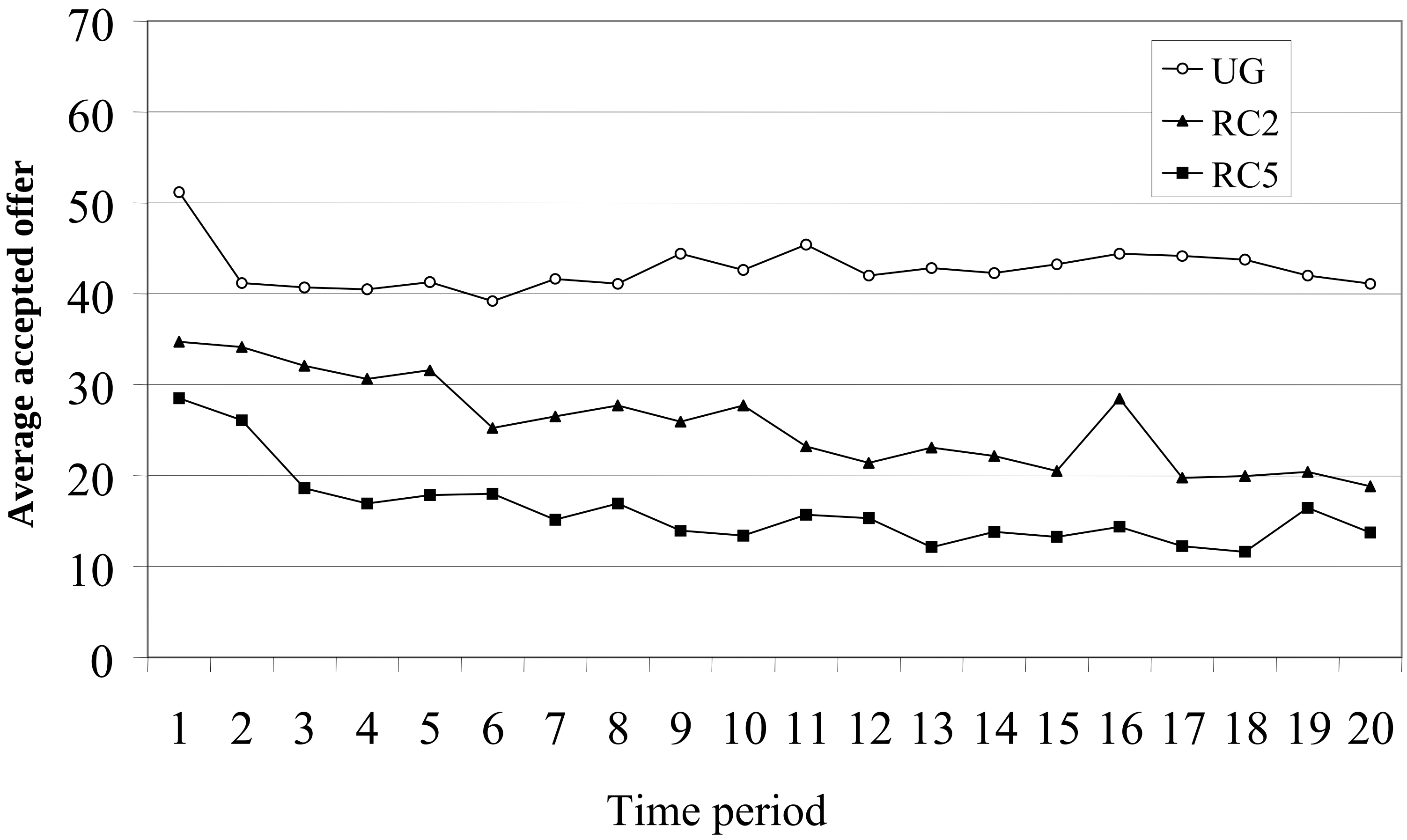
Comparison of 2nd and 3rd Party Punishment



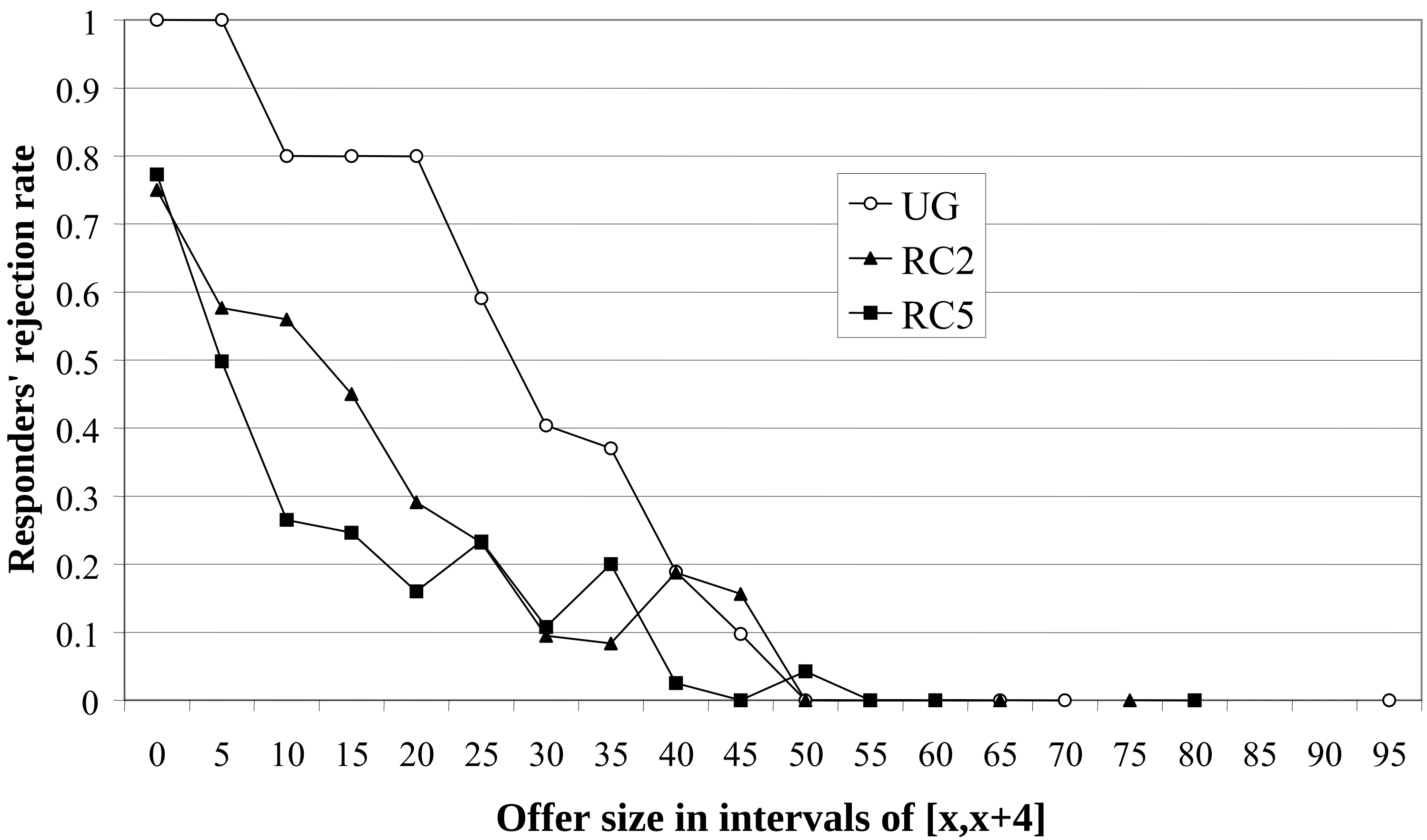
Facts (largely) consistent with the Standard Prediction

- Quick convergence to the CE in double auctions.
- Convergence to CE in posted offer markets.
- First movers modal choice in the UG and the best shot game maximizes their expected monetary payoff.
- Convergence to very low cooperation levels in repeated public good games.
- Very uneven (unfair) outcomes in markets with proposer competition.
- Very uneven (unfair) outcomes in markets with responder competition.
- In the three-player ultimatum game of Güth and van Damme (1998) the proposer and the responder with veto-power do not care about the recipient without veto-power.

Responder Competition – Average Accepted Offers
Source: Fischbacher-Fong-Fehr 2002



Responder Competition – Rejection Behavior
Source: Fischbacher-Fong-Fehr 2002



Modeling other-regarding behavior

- **Models of boundedly rational behavior**
  - Learning models (Roth and Erev GEB 1995, Camerer and Ho 1998)
  - Models of noisy best reply behavior – QRE-models (McKelvey and Palfrey 1995, Goeree and Holt)
- **Models of social preferences**
  - ➢ Reciprocity
    - Rabin (AER 1993)
    - Falk and Fischbacher (DP 1999) (FF)
    - Dufwenberg and Kirchsteiger (DP 1998) (DK)
    - Levine (RED 1998)
  - ➢ Inequity aversion
    - Fehr and Schmidt (QJE 1999) (FS)
    - Bolton and Ockenfels (AER 2000) (BO)
  - ➢ Quasi Maximin
    - Charness and Rabin (QJE 2002) (CR)

Objections to social preferences I

- One can explain everything by changing preferences!!
- True, but if people have other-regarding preferences this objection is simply irrelevant for positive economics.
- My view
  - ➢ The convention against changing preference assumptions made sense in a situation where economists had no tools to discipline the choice of preference assumptions.
  - ➢ Experimental tools provide this discipline.
  - ➢ Useful models have to predict well across many different situations with the same parameters.
  - ➢ Thus, the convention no longer makes sense.

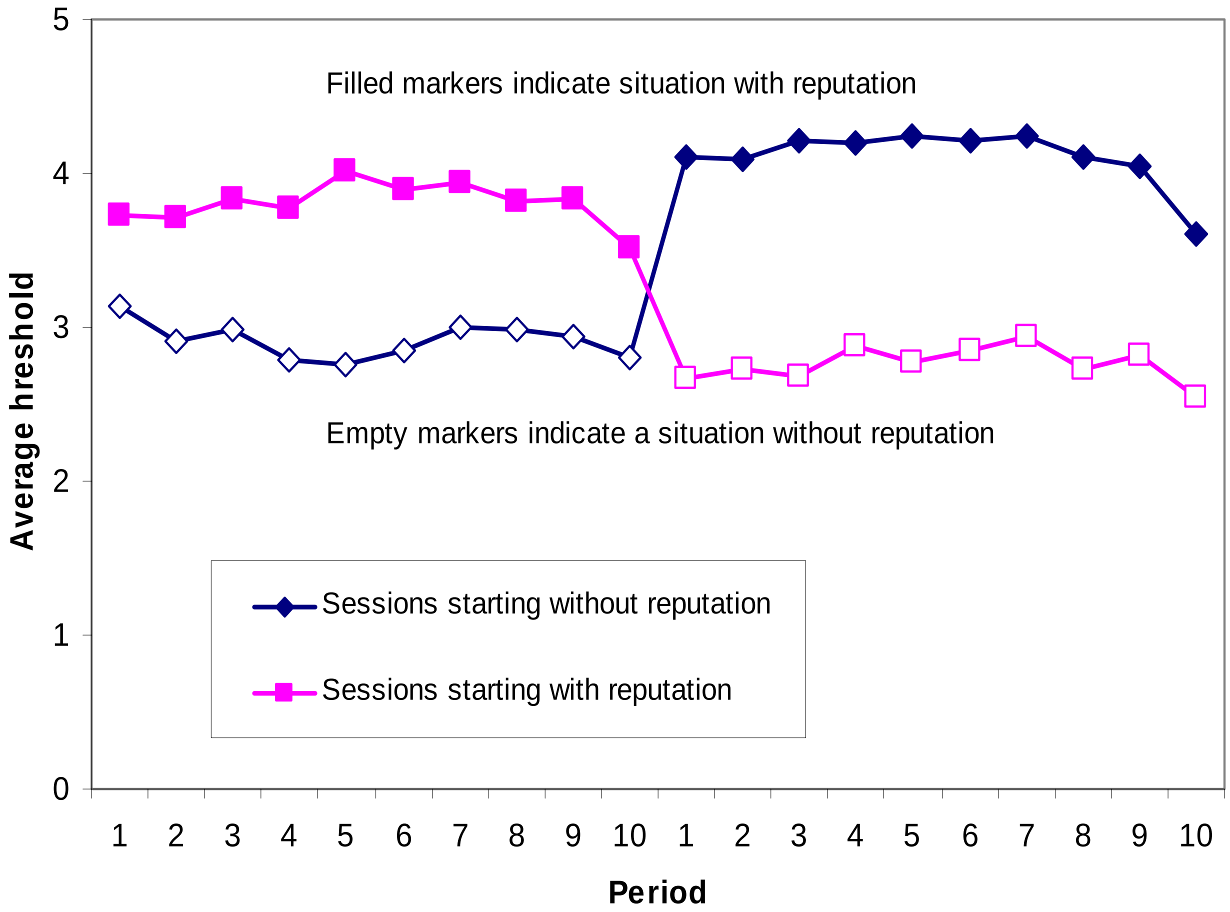
Objections to social preferences II

- Subjects who behave as if they have social preferences don't understand the one-shot nature of the game. They apply repeated game heuristics to one-shot situations.
- My view
  - So far this criticism has been completely unproductive because it has led neither to new models nor to any new insights or predictions.
  - If people really don't understand that they are in a one-shot then one can trash 90 percent of (rational choice) economics. Is it easier to understand that one loses $2 if one rejects an 8:2 offer in a one-shot UG than to play a particular equilibrium in a repeated game?
  - Subjects do respond "in the right way" to the introduction of repeated play opportunities.
  - Even if the argument were correct, the actual behavior of subjects in the one-shot or finitely repeated game must be taken into account.
  - The critics often confuse the evolutionary question of why people behave in other-regarding ways with the question of how this behavior is best modelled.

Reputation and Retaliation
Source: Fehr & Fischbacher (Nature, forthcoming)

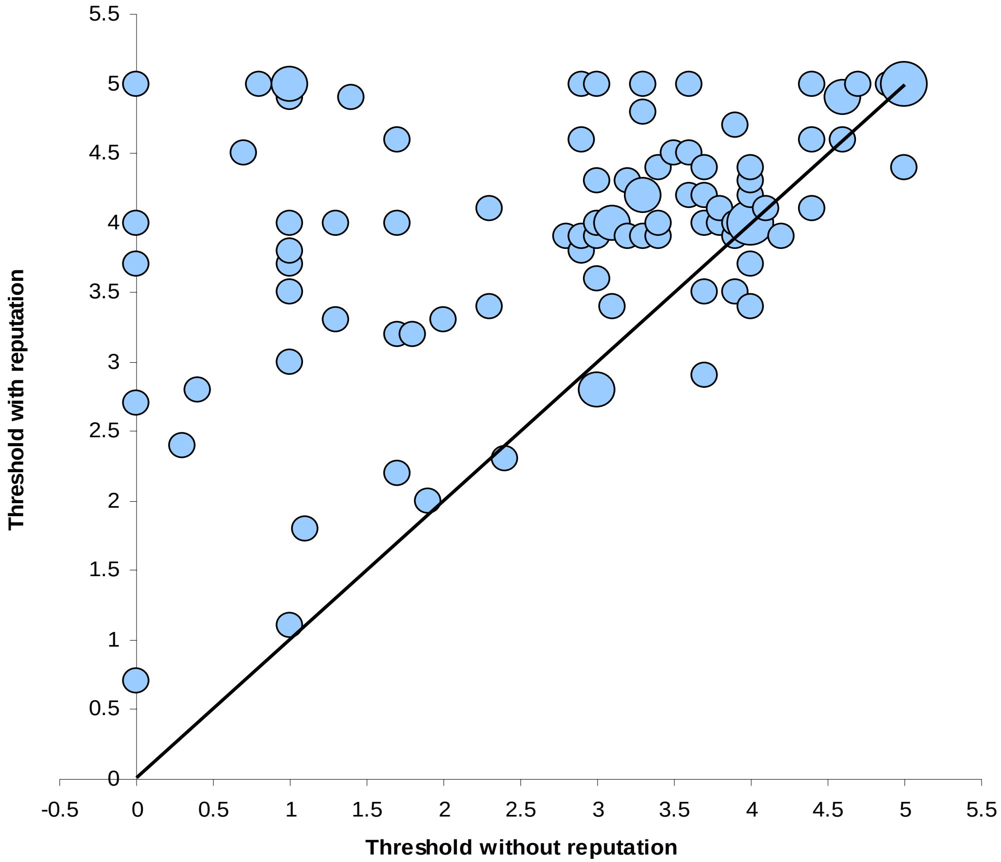
- 10 proposers and 10 responders play 10 ultimatum games with a different opponent in each period. Stake size: 10 money units.
- Baseline treatment: No information about past behavior of opponent.
- Reputation treatment: Responders' past rejection behavior is known to the current proposer. Responders know this and can build a reputation for being a tough responder by rejecting relatively high offers.
- If subjects confuse repeated with one-shot games, there should be no systematic differences in responder behavior across treatments.
- If subjects value their own payoff but have also a preference for fairness they should increase their rejection thresholds in the reputation condition if their thresholds in the baseline was below the equal split.

Reputation and Retaliation – Results
Source: Fehr & Fischbacher (Nature, forthcoming)



Reputation and Retaliation – Results
Source: Fehr & Fischbacher (Nature, forthcoming)

- 82.5 percent of the subjects (n = 95) increase their thresholds; the rest keeps them roughly constant.



Rabin's Reciprocity Model (AER 1993)

- Players reward kind and punish unkind intentions. Beliefs about other players actions, and beliefs about other players' beliefs about the own action, enter **directly** into the utility function; restricted to two person normal form games.
- A_1 and A_2 denote the (mixed) strategy sets for players 1 and 2, and x_i: $A_1 \times A_2 \rightarrow IR$ is player *i*'s material payoff function.
- $a_i \in A_i$ is a strategy of player *i*.
- b_j is player *i*'s belief about player *j*'s strategy ($i \in \{1,2\}$ and $j=3-i$).
- c_i is player *i*'s belief about what player *j* believes that i will choose. The first two levels of beliefs are sufficient to define reciprocal preferences.

Rabin's model - continued

- f_i measures how kind player i is to player j. If $f_i > 0$ player i is kind if $f_i < 0$ she is hostile.
- f_j' is the perceived kindness of player j. $f_j' > 0$ is perceived kindness, $f_j' < 0$ is perceived hostility.
- i's utility function is given by

$$U_i = x_i + f_j'[1+f_i]$$

- Player i wants to be kind to j (i.e., render $f_i > 0$) if she believes that j is kind to her (i.e., if $f_j' > 0$) and she wants to be nasty to j (i.e., render $f_i < 0$) if she believes j to be nasty to her (i.e., if $f_j' < 0$).
- A "fairness equilibrium" is an equilibrium in a psychological game with these payoff functions, i.e., a pair of strategies (a_1,a_2) that are mutually best responses to each other and a set of rational expectations $b=(b_1,b_2)$ and $c=(c_1,c_2)$ that are consistent with equilibrium play.

Measuring Kindness

- $f_i(a_i,b_j) \equiv [x_j(b_j,a_i) - x_j^f(b_j)]/[x_j^h(b_j) - x_j^{min}(b_j)]$
- $x_j(b_j,a_i)$ is the actual payoff *i* gives to *j* given that *i* believes that *j* chooses b_j. $x_j^{min}(b_j)$ is the worst possible payoff for j, given b_j.
- $x_j^f(b_j)$ is the **fair or equitable payoff** for *j* for given b_j. The fair payoff is the average of the lowest payoff *i* can give to *j*, $x_j^l(b_j)$, and the highest payoff that *i* can give to *j*, $x_j^h(b_j)$, excluding Pareto-dominated payoffs, however. *Note the "fair" payoff is independent of the payoff of player i*, i.e., there are no inter-personal fairness considerations.
- If $x_j^h(b_j) - x_j^{min}(b_j) = 0$, $f_i(a_i,b_j) = 0$. Player *i* is kind to *j* if she gives *j* more than *j*'s fair payoff, given b_j. The kindness of *i*, $f_i(a_i,b_j)$, is thus measured by the difference between the actual payoff *i* gives to player *j* and the "fair" payoff, relative to the whole range of feasible payoffs.

Measuring Perceived Kindness

- $f_j'(b_j,c_i) \equiv [x_i(c_i,b_j) - x_i^f(c_i)]/[x_i^h(c_i) - x_i^{min}(c_i)]$ with $j=3-i$.
- $x_i(c_i,b_j)$ is the actual payoff that player i believes player j wants to give to her, given the second-order belief about i's action, c_i. $x_i^{min}(c_i)$ is the worst possible payoff player j can cause for i at a given c_i.
- $x_i^f(c_i)$ is player i's belief about what is her fair payoff, given that i believes that j believes that i plays c_i. As before the fair payoff is the average of the highest payoff, $x_i^h(c_i)$, that j can give to i according to i's belief and the lowest payoff, $x_i^l(c_i)$, that j can give to i according to i's belief (again excluding Pareto-dominated payoffs).
- The perceived kindness of *j*, $f_j(b_j, c_i)$, is zero if $x_i^h(c_i) - x_i^l(c_i) = 0$, i.e, if *j* is believed to be unable to affect *i*'s payoff.
- The kindness terms f and f' are in the interval interval $[-1,0.5]$, i.e., higher stakes render fairness less important.

Application to the simultaneous PD

- $t > c > d > s$ are the PD-payoffs
- Assume that the strength of the nonpecuniary motive is measured by α.
- **Can (C,C) be a fairness equilibrium?**
  - If 1 plays C, then $x_1^f = (c+s)/2$.
  - Thus, if 2 plays C she gives 1 a payoff of c which is larger than x_1^f, i.e. 2 is kind.
  - Thus, if α is big enough 1 has an incentive to be kind to 2, i.e., to play C.
- **Is (D,D) a fairness equilibrium?**
  - If 1 plays C, then $x_1^f = (t+d)/2$.
  - Thus, if 2 plays D she gives 1 a payoff of d which is smaller than x_1^f, i.e. 2 is unkind. Since the situation is symmetric 1 is unkind to 2 if she chooses D.
  - Thus, regardless of the value of α, 1 is always better of by D because D gives a higher material payoff than C and a higher nonpecuniary payoff.
- With strong enough preferences for reciprocity the PD is a coordination game.

	C	D
C	c, c	s, t
D	t,s	d,d

Application to the sequential PD

- CD is the prototypical reciprocal strategy
- One would like that (C, CD) is an equilibrium. Is it?
  - If 1plays C, $x_1^f = (c+s)/2$. Thus if 2 plays CD she gives 1 more than x_1^f, i.e., she is kind. Thus, 1 has an incentive to be kind.
  - If 2 plays CD, $x_2^f = c$ because only the pareto-optimal payoff combinations, given the strategy of 2, enter into x_2^f. Thus, by playing C, 1 is neither kind nor unkind because she gives 2 exactly the fair payoff.
  - Therefore, 1 cares only for the material payoff, which means preferring (c,c) over (d,d).
  - However, since 1 is neither kind nor unkind to player 2, 2 cares only for the material payoff, implying that DC or DD is a better reply to C.
- **Thus, (C,CD) is <u>never</u> a fairness equilibrium (regardless of α)!**

	CC	CD	DC	DD
C	c,c	c,c	s,t	s,t
D	t,s	d,d	t,s	d,d

	CC	CD	DC	DD
C	c,c	c,c	s,t	s,t
D	t,s	d,d	t,s	d,d

- **Is (C,CC) a fairness equilibrium?**
- If 1 plays C, $x_1^f = (c+s)/2$. Thus, if 2 plays CC, she gives 1 c and is, hence, kind. Therefore, 1 has a nonpecuniary incentive to be kind.
- Can 1 be kind by responding to CC with C?
  If 2 plays CC, $x_2^f = (c+s)/2$. Thus by playing C, 1 gives player 2 c which is kind.
- (C, CC) implies mutual kindness and is therefore a fairness equilibrium <u>for sufficiently high α</u>.

- **Is (D, DD) a fairness equilibrium?**
- If 1 plays D, $x_1^f = (t+d)/2$. By choosing DD, 2 gives only d to 1, which is unkind.
- If 2 plays DD, $x_2^f = (t+d)/2$. By choosing D, 1 gives only d to 2, which is unkind.
- Therefore, (D, DD) implies mutual hostility. Given the strategy of the other player, each player has not only a pecuniary incentive to defect but also a nonpecuniary incentive to defect.
- (D,DD) is a fairness equilibrium <u>regardless of the value of α</u>.

Application to a mini-ultimatum game

- 1 proposes a Fair or Unfair division of the pie.
- 2 accepts or rejects.
- If F is accepted: m/2, m/2
  If U is accepted: a,d
  $a > m/2 > d$

	YY	YN	NY	NN
F	m/2, m/2	m/2, m/2	0,0	0,0
U	a,d	0,0	a,d	0,0

- **Is (F, YN) a fairness equilibrium**?
- If 1 plays F, $x_1^f = m/2$. Thus, if 2 plays YN she is neither kind nor unkind because 1 gets exactly m/2. This implies that 1 maximizes her economic payoff and, therefore, F is a best reply to YN.
- If 2 plays YN, $x_2^f = m/2$. Thus, if 1 plays F she is neither kind nor unkind because 2 gets exactly m/2. This means that 2 maximizes her economic payoff and, therefore, YN is a best reply to F.
- This equilibrium already exists without preferences for reciprocity. Note, that with preferences for reciprocity no mutual kindness is involved here.

- **Is (U, YN) a fairness equilibrium?**

	YY	YN	NY	NN
F	m/2, m/2	m/2, m/2	0,0	0,0
U	a,d	0,0	a,d	0,0

- If 1plays U, $x_1^f = a$. If 2 plays YN she gives 0 to 1 and is, hence, unkind. Thus, 1 has a nonpecuniary incentive to be unkind, too.
- If 2 plays YN, $x_2^f = m/2$. If 1 plays U, she gives 0 to 2 and is, hence, unkind. Therefore, 2 has a nonpecuniary incentive to be unkind.
- The strategy pair (U,YN) involves mutual hostility so that for sufficiently high α the strategy pair constitutes a fairness equilibrium.
- Intuitively, the strategy of 2 makes sense. If 2 plays YN, the choice of U by 1 is hostile because is implies sure rejection. In response to hostility 2 is hostile by rejecting (i.e. YN makes sense).
- Intuitively, it is less obvious why, given the choice of U, the strategy YN should be viewed as unkind. Isn't a rejection, after all, a legitimate "self-defense" against being treated unfairly?

Type-based Reciprocity (Levine 1998)

- Each player has altruistic (α_i>0) or spiteful (α_i<0) preferences and α_i obeys $-1 < \alpha_i < 1$.
- If i faces an altruistic player j she becomes more altruistic or less spiteful.
- If i faces a spiteful player j she becomes less altruistic or more spiteful.

- $U_i = x_i + \Sigma_j \kappa_i^j x_j = x_i + \Sigma_j[(\alpha_i + \lambda\alpha_j)/(1+\lambda)]x_j$
- The parameter λ ($0 \leq \lambda \leq 1$) measures a player's inclination to reciprocate to altruistic or spiteful types. The parameter restrictions imply that $-1 < \kappa_i^j < 1$.
- Selfish players exhibit $\alpha_i = \lambda = 0$. Pure altruists (spiteful types) exhibit $\alpha_i > 0$ ($\alpha_i < 0$) and $\lambda = 0$.
- The theory circumvents the problem of what is a fair outcome and assumes instead that we respond to the other players' types.
- In contrast to Rabin's model Levine's theory can explain the presence of third party punishment. However, third parties should punish as strong as second parties, which is refuted by the data.

Altruism & Spitefulness - continued

- Another advantage of the theory is that it is more tractable than models based on psychological game theory.
- The theory is based on the assumption that the cumulative distribution of types $F(\alpha_i)$ is common knowledge whereas α_i is private information.
- This turns the game into a Bayesian game in which agents signal their type by their actions.
- Such signalling games are known to almost always have multiple equilibria.
- Levine shows that he can explain behavior in the UG, market games with proposer competition and public good games.
- He estimates that 20% exhibit $\alpha_i = -0.9$, 52% exhibit $\alpha_i = -0.2$ and only 28% exhibit $\alpha_i > 0$.
- In DG the theory predicts zero transfers, and in the trust game zero back-transfers and, therefore, zero transfers since $\kappa_i^j < 1$. More generally, since the marginal utility of transferring $1 to another player is less than 1, the transfer never takes place.
- Since $\kappa_i^j > -1$, the marginal utility of reducing another players' payoff by $1 is less than $1. Thus, the reduction does not take place if it costs $1 or more.

Outcome-based models of social preferences

- $U_i = U_i(x_i, x_{-i})$
- How does U_i depend on x_{-i}?
  - Share: $x_i/\Sigma x_j$ (BO)
  - Payoff differences to all other players:= x_i-x_j (FS)
  - Total surplus: Σx_j; payoff of worst off player: min(x_j) (CR)
  - Only total surplus: Σx_j

The Fehr-Schmidt Model (QJE 1999)

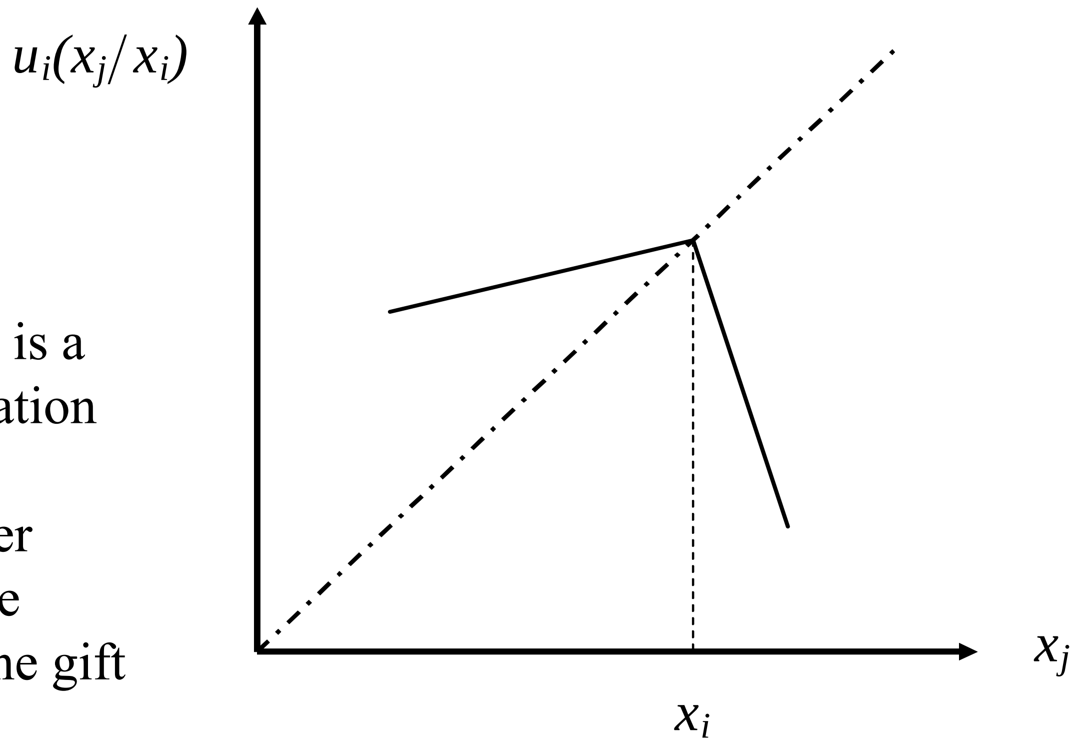
- A substantial percentage of people are inquity averse. In anonymous experimental games payoff equality is often a salient reference point.
- Apply with care to field situations!!!
- Other players' payoffs are negatively valued if others are better off, and positively valued if others are worse off.
- $U_i = x_i - (\alpha_i/(n-1))\Sigma\max(x_j-x_i,0) - (\beta_i/(n-1))\Sigma\max(x_i-x_j,0)$

  α_i measures aversion against disadvantageous inequity (envy).
  β_i measures aversion against advantageous inequity (guilt).

- Assumptions: $\alpha_i \geq \beta_i \geq 0$; $\beta_i < 1$, implies that people don't burn money to reduce inequality.
- Non-pecuniary payoffs normalized by (n-1) to keep the relative importance of non-pecuniary payoff constant

A graphical representation

Two-Player Case



Piece-wise linearity is a deliberate simplification and leads to counterfactual corner solutions in, e.g., the dictator game and the gift exchange game.

Dictator Game

- No inequity averse dictator will give more than 50% to the recipient.
- The dictator's utility for $s \leq 0.5$ is $U_1 = 1 - s - \beta[(1-s)-s)]$.
- The derivative $\delta U_1/\delta s = -1 + 2\beta$ so that for $\beta < ½$ the proposer chooses $s = 0$ and for $\beta \geq ½$ she prefers $s = ½$.
- Similar bang-bang solution occurs for the workers' effort choice in the gift exchange game.

Behavior in the Ultimatum Game

- Offers above 50% are always accepted.
- An offer s below 50% will be accepted if

$$s - \alpha_i(1-2s) > 0, \text{ i.e. } \alpha_i < s/(1-2s) \text{ or } s > \alpha_i/(1+2\alpha_i)$$

- The rejection threshold $\alpha_i/(1+2\alpha_i)$ is increasing and strictly concave in α_i and approaches ½ if α_i approaches infinity.
- Proposers never offer more than 50%. In case of complete information about α_i
  - Offer 50% if $\beta_i > ½$.
  - Offer exactly the rejection threshold if $\beta_i < ½$.

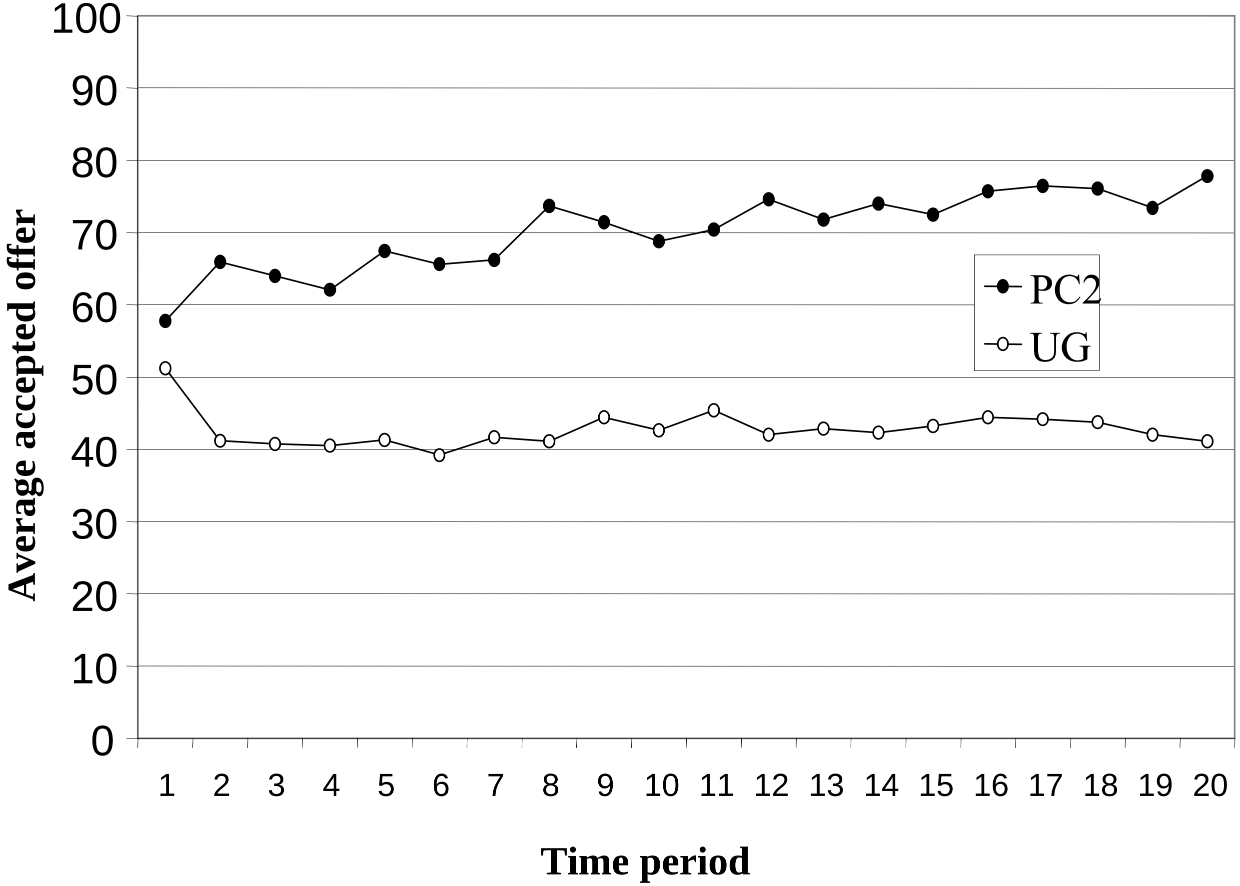
➢ In case of incomplete information
  - Offer 50% if $\beta_i > ½$.
  - For $\beta_i < ½$ the proposers in general accept some risk of being rejected.
  - The higher β_i the higher the offer.

Market Game with Proposer Competition

- For any distribution of parameters α_i and β_i there is a unique subgame perfect equilibrium outcome in which at least two sellers offers $s = 1$ which is accepted by the buyer.
- Responder accepts any $s > 0.5$ because rejection is too costly
- *A kind of Bertrand competition among proposers:*
- Assume that proposer i makes the highest offer
- $0.5 < s^h < 1$. If j offers slightly more she has three advantages:
    - Less disadvantageous inequality relative to the responder
    - Less disadvantageous inequality relative to proposer i
    - Monetary gain

- **Equilibrium outcome "preference-free" or "culture-free"**
- Compare with the results of Roth et al. (1991).

A Stress Test: Proposer Competition with only two Proposers (Fischbacher-Fong-Fehr 2002)



Market Game with Responder Competition – Proposer Behavior

- If $\beta_1 < (n - 1)/n$ there exists an equilibrium in which all responders accept any $s \geq 0$ and the proposer offers $s = 0$.
- Transferring $1 to a responder reduces inequality relative to the responder who receives the $ by 2 units and relative to the other n-2 responders by 1 unit.
  - Average reduction in inequality is $(2 + n\text{-}2)/(n\text{-}1) = n/(n\text{-}1)$.
  - If $\beta[n/(n\text{-}1)] < 1$ or $\beta < (n\text{-}1)/n$ the sellers prefers to make inegalitarian offers.
  - For $n = 3$ we have $\beta < 2/3$; for $n = 6$ we have $\beta < 5/6$
- Under responder competition even very fair-minded proposers are willing to make very low offers.   .

Responder Competition – Responder Behavior

- Suppose responder i accepts s = ε, i.e. there is inequality anyway. If responder j accepts too she has three advantages:
    - she may win ε
    - she may turn the disadvantageous inequality relative to i into advantageous inequality
    - she may reduce the inequality relative to the proposer.
- Thus every positive offer is accepted, no matter how fair-minded the responder is, if she believes that one competing responder accepts.
- **A single selfish responder triggers this equilibrium**

Responder Competition - continued

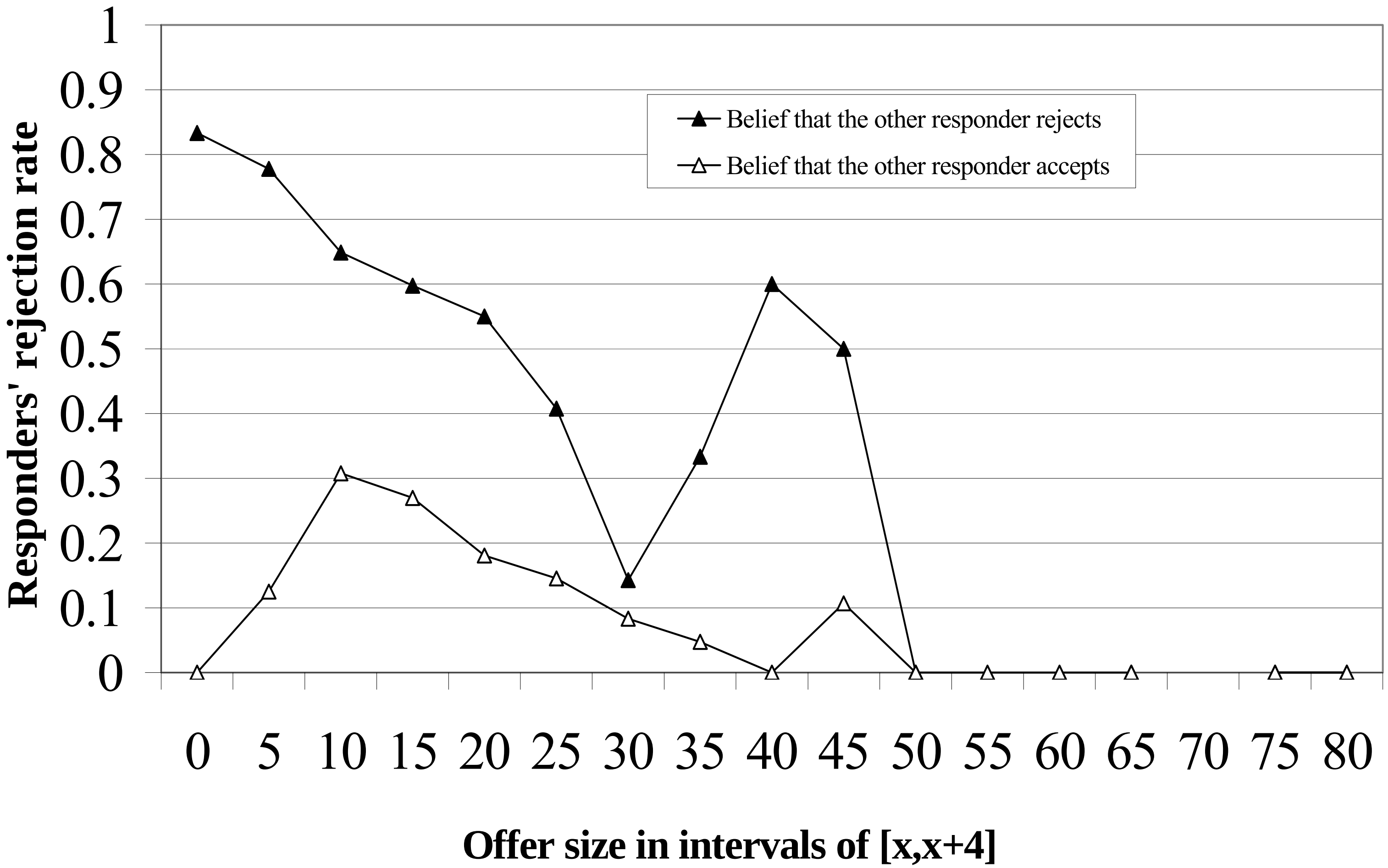
- If all responders are inequality averse there are equilibria with positive offers. The least inequality averse responder determines the size of the equilibrium offer.

$$\min\left\{\frac{\alpha_i}{(1-\beta_i)(n-1)+2\alpha_i+\beta_i}\right\}$$

- Below the threshold of any player in the ultimatum game $\alpha_i/(1+2\alpha_i)$.
- Increasing the number of competitors increases the chances that the least inequality averse players is now less inequality averse. Thus, the highest sustainable equilibrium offer decreases.

- The model explains why very uneven outcomes are obtained although many players have a preference for fairness.

Responders' rejection rate in RC2 conditional on offer size and beliefs about the other responder's behavior



Responder Beliefs and Rejection Behavior
Source: Fischbacher-Fong-Fehr 2002

Probit model for rejection behavior	coeff.	marginal effect
Constant	0.087	
Period	-0.021**	0.007
Offer	-0.055***	-0.017
Belief that all others reject (dummy)	1.089***	0.374
Dummy for RC5	-0.272	0.088
Dummy for RC2	-0.236	0.072

- Interactions between treatment and period also included but both for RC2 and for RC5 the std. error is larger than the coefficient.
- If players believe that they can punish or hurt the proposer by a rejection they are much more likely to reject.
- Direct effect of treatment dummies is insignificant.

Three Person Ultimatum Game of Güth and van Damme (1998)

- There is a proposer X, a responder Y, and a recipient Z. The proposer proposes an allocation of material payoffs (x, y, z) which can be accepted or rejected by Y. The recipient Z can do nothing.
- Facts: Proposers give almost nothing to Z and the Responder does only reject if y is low but not if z is low.
- **Why do responders not reject low offers to Z?**
- Assume x = y and z = 0. Responder Y will never reject such an offer even if he is very fair-minded, i.e., has a high β-coefficient!! Her utility is $U_2 = y - (\beta/2)(y - z)$, which is positive for all $\beta \leq 1$, and thus higher than the rejection payoff of zero.

Güth & van Damme & Third Party Punishment

- **Why do proposers make low transfers to Z?**
- The Proposers utility is given by $U_x = x - (\beta/2)[(x-y)+(x-z)]$.

  Let $x + y + z = 1$, then $U_x = (\beta/2) + (3/2)x[(2/3) - \beta]$.

  Unless $\beta > 2/3$ the proposer will always propose z = 0 and x just high enough to prevent rejection.

Third party punishment

- The third party has a non-pecuniary incentive to punish if the dictator gives less than 50 to the recipient.

Public Goods Game

- $\pi_i = y_i - c_i + a\Sigma c_j$ with $a < 1$ and $na > 1$.
- If all n players obey $1\text{-}a < \beta_i$, every symmetric contribution vector is an equilibrium. Suppose all players contribute c and i considers to reduce c_i by one unit.
    - i **gains** $1\text{-}a$
    - There is disadvantageous inequality relative to each of the remaining $n\text{-}1$ players with a total **nonpecuniary loss** of $\beta_i(n\text{-}1)/(n\text{-}1)$
    - $1\text{-}a < \beta_i$ ensures cooperation
- **Inequality averse players are conditionally cooperative which turns the public goods game into a coordination game.**

Heterogeneity and Institutions

- Consider the PG with one selfish and one inequity averse player who obeys $1-a < \beta_i$.
- Assume that the players' types are common knowledge.
- In the sequential PG where the selfish player moves first, the unique equilibrium outcome is that both players cooperate.
- In the simultaneous PG full defection by both players is the unique equilibrium outcome.
- Prediction: in the sequential two-player PG there is more cooperation than in the simultaneous PG.

Public Goods Game - continued

- If $1\text{-}a > \beta_i$, player i is an **unconditional defector**, i. e., it is a dominant strategy for that player to choose $c_i = 0$.
- Let k denote the number of unconditional defectors ($0 \leq k \leq n$). If there are **sufficiently many unconditional defectors** ($k/(n\text{-}1) > a/2$) then there is a **unique equilibrium with $c_i = 0$** for all $i \in \{1, ..., n\}$.
- Examples: n = 5, a = 0.4, then k = 1 unconditional defector suffices to render $c_i = 0$ the unique equilibrium.
- If the **number of unconditional defectors is sufficiently low** so that for all players $j \in \{1, ..., n\}$ with $1\text{-}a < \beta_j$ the condition $[k/(n\text{-}1) < (a+\beta_j\text{-}1)/(\alpha_j+\beta_j)]$ holds, then it is an equilibrium if all unconditional defectors choose $c_i=0$ while all other players contribute $c_j = c \in [0,y]$. Furthermore, $(a+\beta_j\text{-}1)/(\alpha_j+\beta_j) < a/2$.

Public Good with Punishment

- Suppose there is a group of m, $1 \leq m \leq n$ "conditionally cooperative enforcers" while all other players are completely selfish. Conditionally cooperative enforcers obey $a + \beta_i > 1$ and exhibit sufficiently large inequity aversion relative to the cost γ ($\gamma < \alpha_i/[(n-1)(\alpha_i+1)-(m-1)(\alpha_i+\beta_i)]$. Then full co-operation by everybody is part of a subgame perfect equilibrium.
- Sufficiently large inequity aversion makes the punishment threat credible so that selfish players cooperate
- $a + \beta_i > 1$ prevents defection by the enforcers.
- **A minority of conditionally cooperative enforcers can "trigger" the full cooperation outcome**
- **Note, there are multiple equilibria but with efficiency as a refinement criterion one captures the data quite well.**

Summary

- The model of Fehr and Schmidt predicts the **qualitative** outcomes of many experimental games correctly – and provides insights into the mechanisms that generate fair and unfair results.
    - Ultimatum game
    - Dictator game
    - Responder competition
    - Proposer competition (partially)
    - Güth-van Damme three player ultimatum game
    - Public goods game
    - Public goods with punishment
    - Trust game
    - Third party punishment game

Combining Fehr&Schmidt with Quantal Response to make precise quantitative predictions

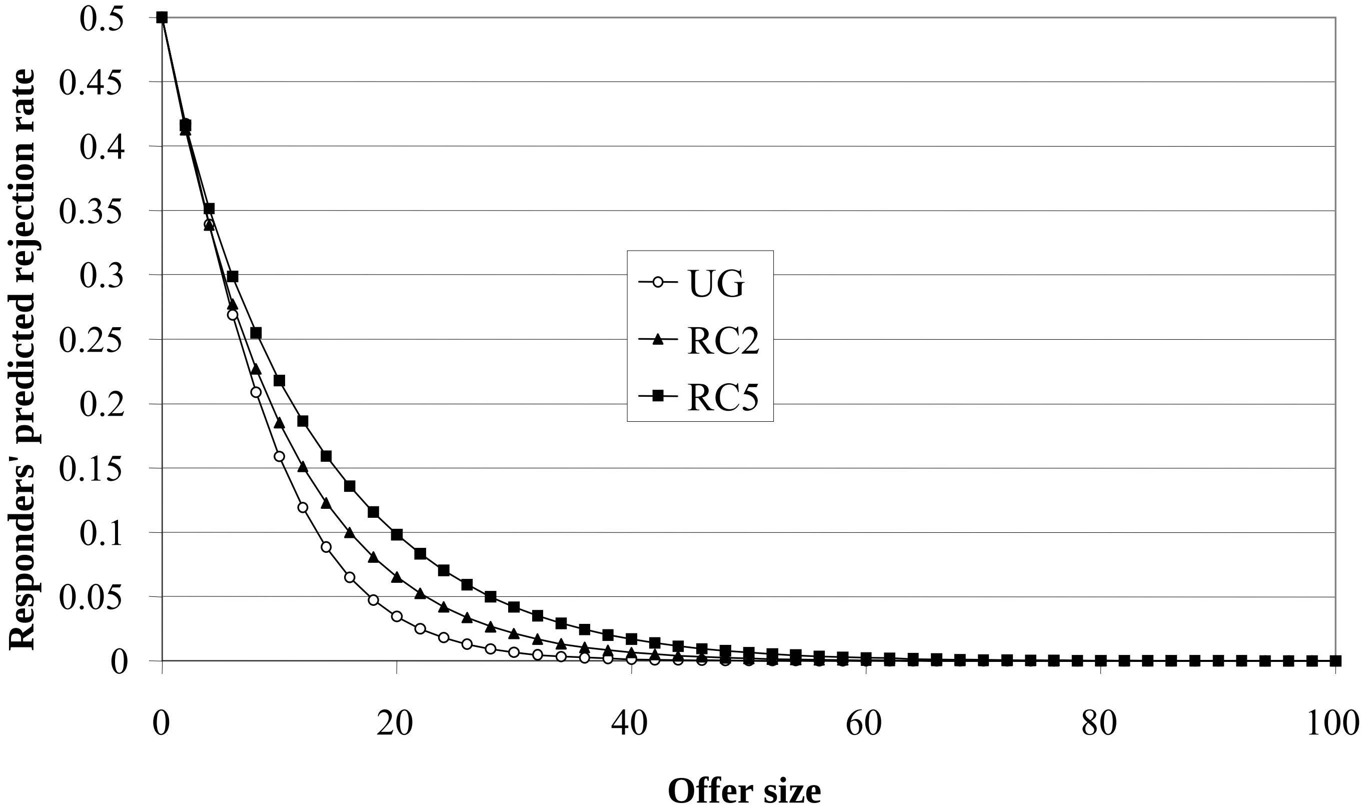
- QRE approach assumes that subjects play noisy best replies. The exact best reply is played with the highest probability but due to errors there are deviations. Errors are extreme-value distributed. Leads to logit rules: The probability of action i is given by

$$Pr(i) = \frac{exp(\pi^e(i)/\mu)}{\sum_{j=1}^{N} exp(\pi^e(j)/\mu)}, \qquad i = 1,...,N$$

- In QRE the subjective probabilities that enter into the calculation of the expected payoff of action i coincide with the objective probabilities.
- We applied a combination of Fehr-Schmidt and QRE to explain the results in Fischbacher-Fong-Fehr 2002.

QRE-prediction of responder behavior

- QRE alone predicts the wrong comparative statics of buyers’ rejection behavior. If other buyers are more likely to accept an offer, it is cheaper to reject. Therefore, rejection rate goes up if others are believed to accept!!!!

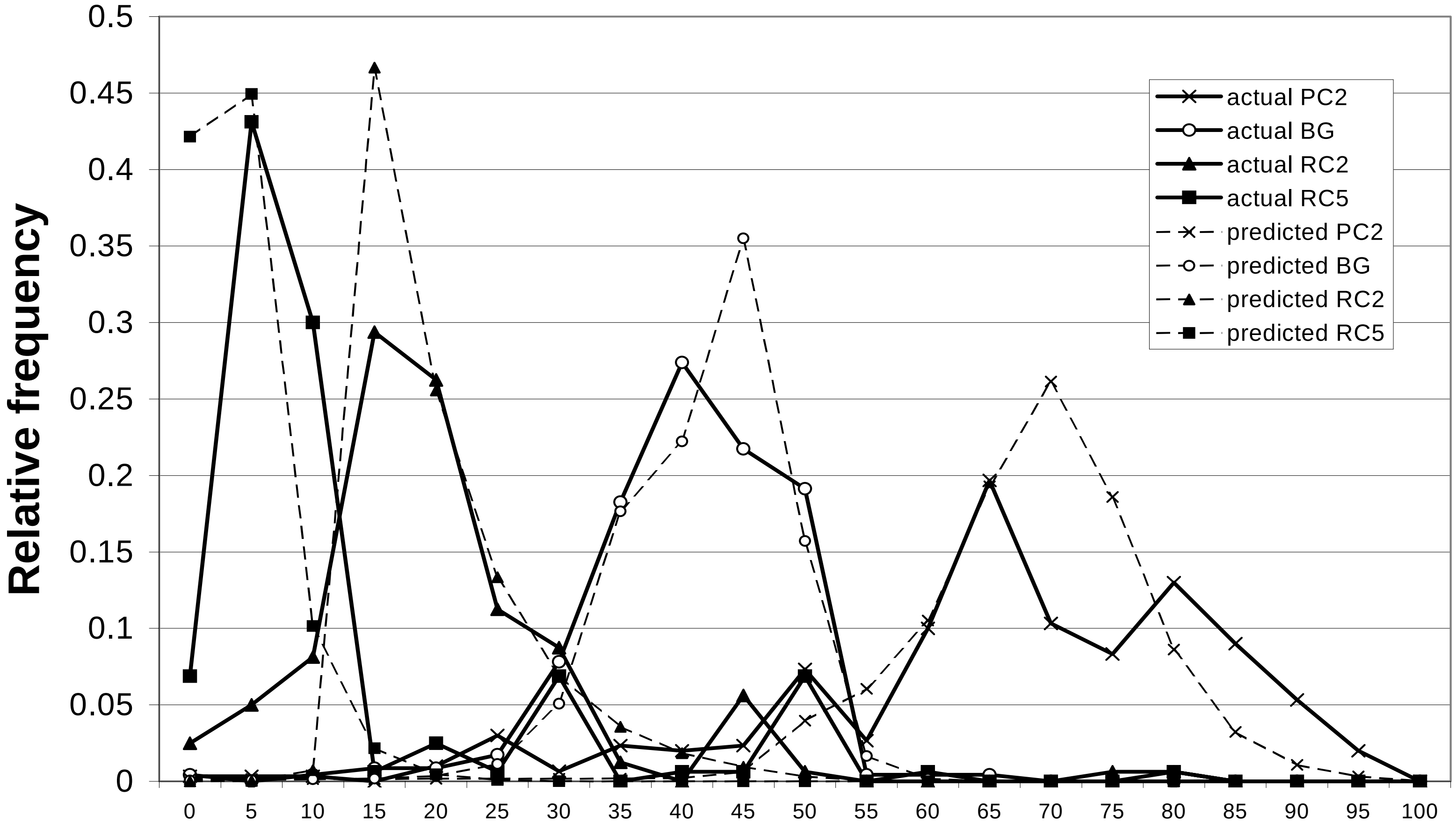


Combining QRE and Fehr-Schmidt
The FS-parameters (exactly the same as in QJE 1999)

α_i	Population share	Rejection threshold in UG	β_i	Population share	Offer in the UG
0	30%	0	0	30%	1/3
0.5	30%	1/4	0.25	30%	4/9
1	30%	1/3	0.6	40%	1/2
4	10%	4/9			

- These parameters predict the average behavior in the ultimatum game quite well.
- Question: Do these parameters also predict well in other games?

Predicting the Distribution of Accepted Offers (only data from periods 11-20 where stable average behavior prevailed)



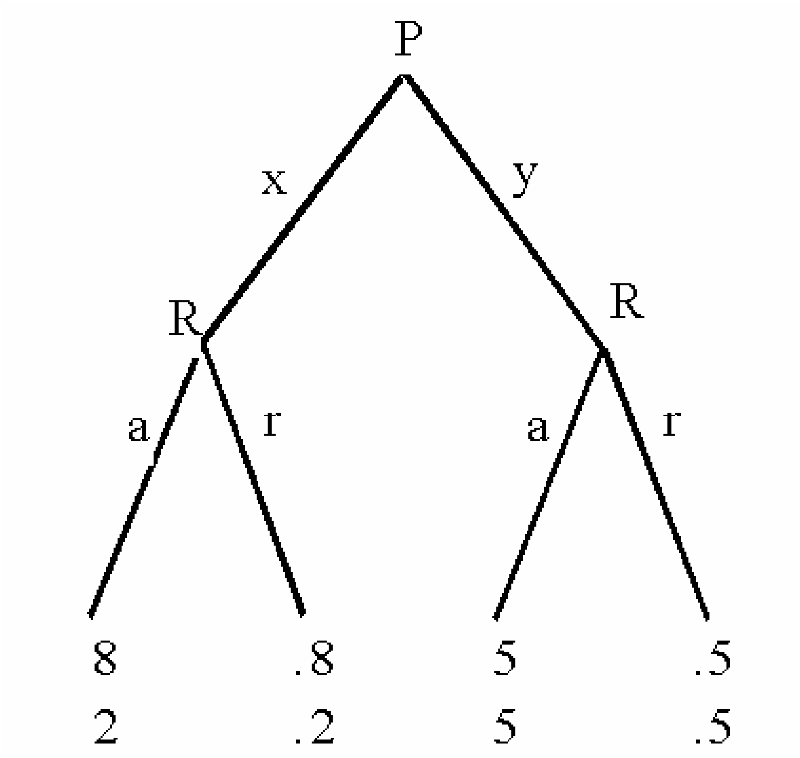
The Bolton-Ockenfels model (AER 2000)

- Subjects care for getting a fair share: $s_i = x_i / \Sigma x_j$ ($s_i = 1/n$, in case of $\Sigma x_j = 0$)
- $U_i = U_i(x_i , s_i)$
- U_i is (weakly) increasing in x_i and concave in s_i, becoming maximal at $s_i = 1/n$
- U_i is differentiable over the whole domain. No kinks.
- The BO-model is consistent with the UG data, DG data, Güth & van Damme data, proposer competition data in the Roth et al. cross cultural experiments (BO also predict the "selfish" outcome), and with reciprocal behavior of 2^{nd} movers in trust and gift exchange games.
- BO fails to capture the patterns of punishment in cooperation and third party punishment experiments and makes counter-intuitive predictions in variations of the UG.

Punishment in the BO-model

- Suppose that a rejection in the UG gives the proposer $\varepsilon > 0$ and the responder zero. Subjects with BO preferences will never reject a positive offer because rejection reduces their share to zero.
- The same holds true in an UG in which a rejection implies that, say, 90% of each player's proposed payoff is destroyed. Here a rejection cannot change the share and is, therefore, futile (see next page).
- In the cooperation and punishment game BO-preferences cannot explain why the bulk of the sanctions is imposed on the defectors. A player who wants to change his share of the payoff, $s_i = x_i / \Sigma x_j$, does not care whether he punishes a cooperator or a defector. Thus, with BO-preferences players should randomly target their punishment on cooperators and defectors.
- If the punishment of defectors is more expensive than the punishment of cooperators, then BO-preferences imply that only the cooperators will be punished. The data in Falk, Fehr, Fischbacher (Driving Forces of Informal Sanctions, 2001) refute this prediction.
- In the third party punishment game where the dictator has an endowment of 100, the recipient has zero endowment and the 3rd party has an endowment of 50, the 3rd party has a fair share and will never punish!!
- **Mispredicting punishment is a systematic feature of the BO-model.**

Punishment if the income share remains unchanged? (Source: Falk, Fischbacher, Fehr 2001)



- In the baseline UG 57% of the responders rejected the 8/2 offer.
- A BO-subject never punishes in the modified UG.
- Yet, still 38% of the responders rejected the 8/2 offer.

The punishment motive in the BO-model

- *p* denotes the punishment cost for the punisher, *cp* are the costs for the punished, $x_i / \Sigma x_j$ is the pre-punishment share and x_i the pre-punishment income of the punisher. The post-punishment share is then $s(p) = [x_i - p]/[\Sigma x_j - (1+c)p]$.
- Differentiation with respect to *p* shows that the derivative s_p is positive (negative) if and only if *s* exceeds (is below) $1/(1+c)$. There are thus 4 cases, depending on the position of the subject's relative share before the punishment, *s*(0):
- Case 1: *s*(0) < 1/n and s(0) < 1/(1+*c*)
- Case 2: s(0) > 1/n and s(0) < 1/(1+c)
- Case 3: s(0) < 1/n and s(0) > 1/(1+*c*)
- Case 4: s(0) > 1/n and s(0) > 1/(1+*c*)

The punishment motive in the BO-model

- **Case 1: $s(0) < 1/n$ and $s(0) < 1/(1+c)$ so that s_p is negative.**
  The subject would like to increase *s*, because *s*(0) is below the fair share 1/n. Yet, since s_p is negative, sanctioning another subject would only decrease the relative share of the sanctioning subject. Thus, in case 1 a subject will never sanction. ***Thus highly cooperative subjects ($s(0) < 1/n$), who can punish free-riders at a low relative cost ($s(0) < 1/(1+c)$), will never punish.***
- **Case 2: $s(0) > 1/n$ and $s(0) < 1/(1+c)$ so that s_p is negative.**
  Since the subject has an unfairly high share she wants to decrease her relative share and she can do so by sanctioning another subject because $s_p < 0$. ***Thus, strong free-riders ($s(0) > 1/n$), who can punish others at a low relative cost ($s(0) < 1/(1+c)$), are predicted to punish.*** This is inconsistent with any reasonable notion of fairness and equity.
- **Case 3: $s(0) < 1/n$ and $s(0) > 1/(1+c)$ so that s_p is positive.**
  The subject wants to increase the share and can do so by punishing others because s_p is positive.
- **Case 4: $s(0) > 1/n$ and $s(0) > 1/(1+c)$ so that s_p is positive.**
  The subject wants to decrease his share, yet sanctioning another subject would only increase his share (because $s_p > 0$).
- **The UG-case:** for low offers $s(0) = x_2/(x_1+x_2) < ½$ so that the responder wants to increase s. The subject can do so because by assumption zero payoff for everybody is coded as a share of ½.

Fehr-Schmidt versus Bolton-Ockenfels in simple allocation games (Engelmann & Stobl 2002)

Treatment	F			E			Fx			Ex		
Allocation	A	B	C	A	B	C	A	B	C	A	B	C
Person 1	8.2	8.8	9.4	9.4	8.4	7.4	17	18	19	21	17	13
Person 2	5.6	5.6	5.6	6.4	6.4	6.4	10	10	10	12	12	12
Person 3	4.6	3.6	2.6	2.6	3.2	3.8	9	5	1	3	4	5
Total	18.4	18	17.6	18.4	18	17.6	36	33	30	36	33	30
Average 1.3	6.4	6.2	6	6	5.8	5.6	13	11.5	10	12	10.5	9
Relative	.304	.311	.318	.348	.356	.364	.278	.303	.333	.333	.364	0.4
Efficient	A			A			A			A		
B&O			C	A					C	A		
F&S	A					C	A					C
Maximin	A					C	A					C
Choices (absolut)	57	7	4	27	16	25	26	2	2	12	5	13
Choices (%)	83.8	10.3	5.9	39.7	23.5	36.7	86.7	6.7	6.7	40	16.7	43.3

Equity versus Efficiency

- If FS coincides with the efficiency motive, then 83.8 % choose the allocation that is consistent with FS and only 5.9% choose the BO allocation.
- If BO coincides with the efficiency motive, then 39.3 % choose the allocation that is consistent with BO. Yet, still 36.7% choose the FS-allocation.
- Controlling for the efficiency motive FS outperforms BO
  - ➢ but FS-success could be driven by maximin motive.
  - ➢ Efficiency motive is obviously important.
- Charness & Rabin (2002) also report that subjects often prefer efficient over egalitarian allocations.
  - ➢ 69% of subjects B prefer (750, 400) over (400, 400) and 48% prefer (750, 375) over (400, 400)
- The E&S subject pool consisted exclusively of economics students and the C&R subject pool consisted primarily of economics and business administration students.

Quasi-Maximin Preferences (Charness & Rabin 2002)

- Charness & Rabin assume that players are driven by reciprocity and outcome-oriented social preferences.
- Their formalization of reciprocity is, however so cumbersome and complicated that they do not even apply it to explain their own data.
- The outcome-oriented part is defined as follows:
  - $W(x_1, x_2,..., x_n) = \delta\min(x_1, x_2,..., x_n) + (1-\delta)\Sigma x_i$
  - $U_i = (1-\lambda)\, x_i + \lambda W$

Equity versus Efficiency among Economists and Noneconomists (Source: Fehr & Schmidt 2003)

	Ey			P		
Allocation	**A**	**B**	**C**	**A**	**B**	**C**
Person 1 Payoff	**21**	**17**	**13**	**14**	**11**	**8**
Person 2 Payoff	**9**	**9**	**9**	**4**	**4**	**4**
Person 3 Payoff	**3**	**4**	**5**	**5**	**6**	**7**
Total Payoff	33	30	27	23	21	19
Average Payoff of 1 and 3	12	10.5	9	9.5	8.5	7.5
Efficiency	A			A		
Inequity Aversion (F&S and B&O)			C			C
Rawlsian maximin rule			C	A	B	or C
Engelmann & Strobel results						
Choices (abs. number)	12	7	11	18	2	10
Choices (%)	**40**	**23.3**	**36.7**	**60**	**6.7**	**33.3**
Our results: non-economists Institute for Advanced Study in Berlin						
Choices (abs. number)	9	4	30	9	9	27
Choices (%)	**21**	**9**	**70**	**20**	**20**	**60**
Our results: economists University of Munich						
Choices (abs. number)	72	12	25	63	16	30
Choices (%)	**66.1**	**11**	**22.9**	**57.8**	**14.7**	**27.5**
Our results: non-economists University of Munich						
Choices (abs. number)	22	13	48	21	17	45
Choices (%)	**26.5**	**15.7**	**57.8**	**25.3**	**20.5**	**54.2**
Our results: non-economists Zürich, Switzerland						
Choices (abs. number)				8	8	20
Choices (%)				**22.2**	**22.2**	**55.6**

Social Preferences in Strategic and Non-Strategic Games (Source: Fehr & Schmidt 2003)

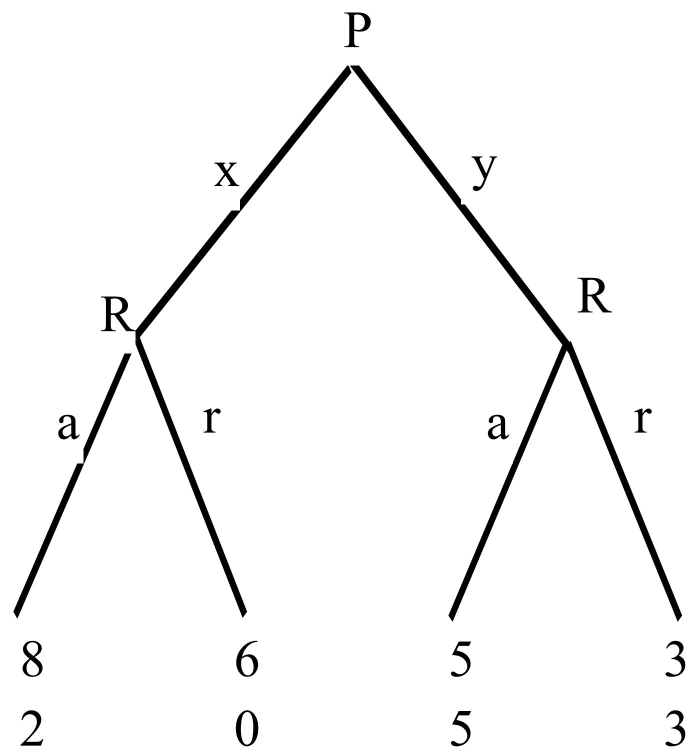
Conjecture: Equity more important in strategic games

	Ey			P		
Allocation	**A**	**B**	**C**	**A**	**B**	**C**
Person 1 Payoff	**21**	**17**	**13**	**14**	**11**	**8**
Person 2 Payoff	**9**	**9**	**9**	**4**	**4**	**4**
Person 3 Payoff	**3**	**4**	**5**	**5**	**6**	**7**
Total Payoff	33	30	27	23	21	19
Average Payoff of 1 and 3	12	10.5	9	9.5	8.5	7.5
Efficiency	A			A		
Inequity Aversion (F&S and B&O)			C			C
Rawlsian maximin rule			C	A	B	or C
Subject pool 1: economists						
without prior ultimatum game						
choices (abs.)	72	12	25	63	16	30
choices (%)	**66.1**	**11**	**22.9**	**57.8**	**14.7**	**27.5**
with prior ultimatum game						
choices (abs.)	69	28	73	73	36	61
choices (%)	**40.6**	**17**	**42.9**	**42.9**	**21.2**	**35.9**
Subject pool 2: non-economists						
without prior ultimatum game						
choices (abs.)	22	13	48	21	17	45
choices (%)	**26.5**	**15.7**	**57.8**	**25.3**	**20.5**	**54.2**
with prior ultimatum game						
choices (abs.)	16	17	44	18	11	48
choices (%)	**20.8**	**22.1**	**57.1**	**23.4**	**14.3**	**62.3**

How important is maximin in a strategic context?

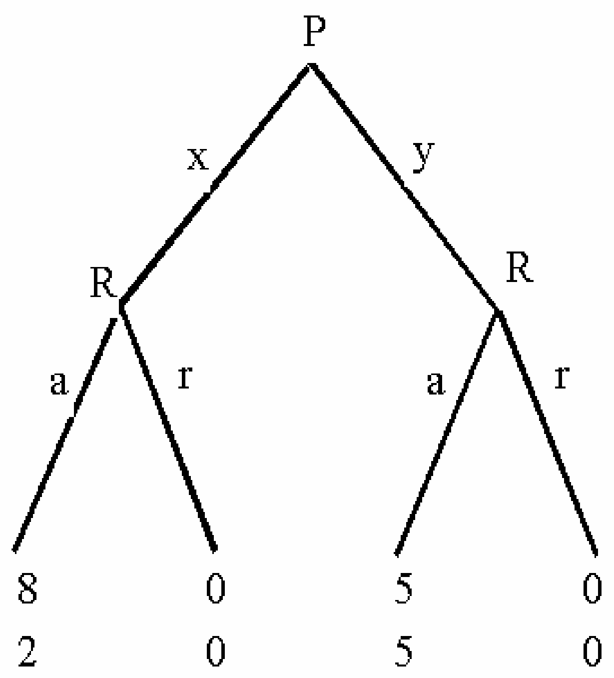
- Güth & van Damme game suggests that maximin is not very important.
- Frechette, Kagel and Lehrer (APS 2003) conducted 5 person voting games. A proposer can propose an allocation, then all five players vote with Yes or No, simple majority vote suffices.
  - Often the proposers propose an allocation that gives the whole payoff to the minimal winning coalition.
- Kagel and Wolfe (Exp. Econ. 2001) conducted a three-player UG similar to the Güth & van Damme game, where a rejection by the responder causes a loss for the passive recipient of $10.
  - Changing the 'consolation price' for the passive recipient, from a baseline level of zero (or a positive amount), so that it produces damage to an innocent third party, does not result in a significant reduction of the rejection rate.
- Conjecture: Maximin is rather unimportant in a strategic context but in allocation games that put the subject in a "moral position" maximin can be quite important.

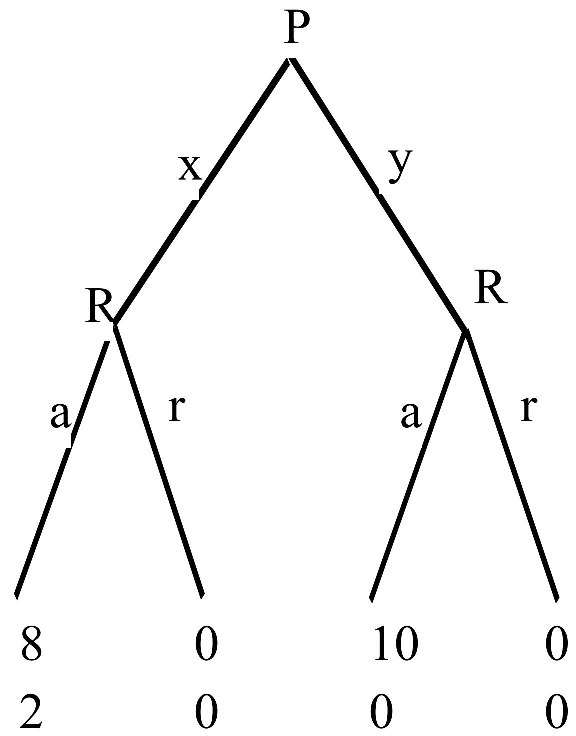
Punishment when income differences remain unchanged? (Source: Falk, Fehr, Fischbacher 2001)



- In the baseline UG 57% of the responders rejected the 8/2 offer.
- In the modified UG only (still) 19% of the responders rejected which is incompatible with FS and BO.

The role of available opportunities (Falk, Fehr, Fischbacher EI 2003)





The role of available opportunities (Falk, Fehr, Fischbacher EI 2003)

- FS (and BO) predict no differences across treatments.
- Interpretation: the 8/2 offer reveals different intentions or allows different inferences about the greediness of the proposer depending on which alternative is available.
- Subjects punish much more in case that a greedy intention or personality can be inferred.
- But even in the condition where the proposer had no choice 18% of the offers are rejected.
- Suggests that outcomes and inferences about intentions (or personality) matter.

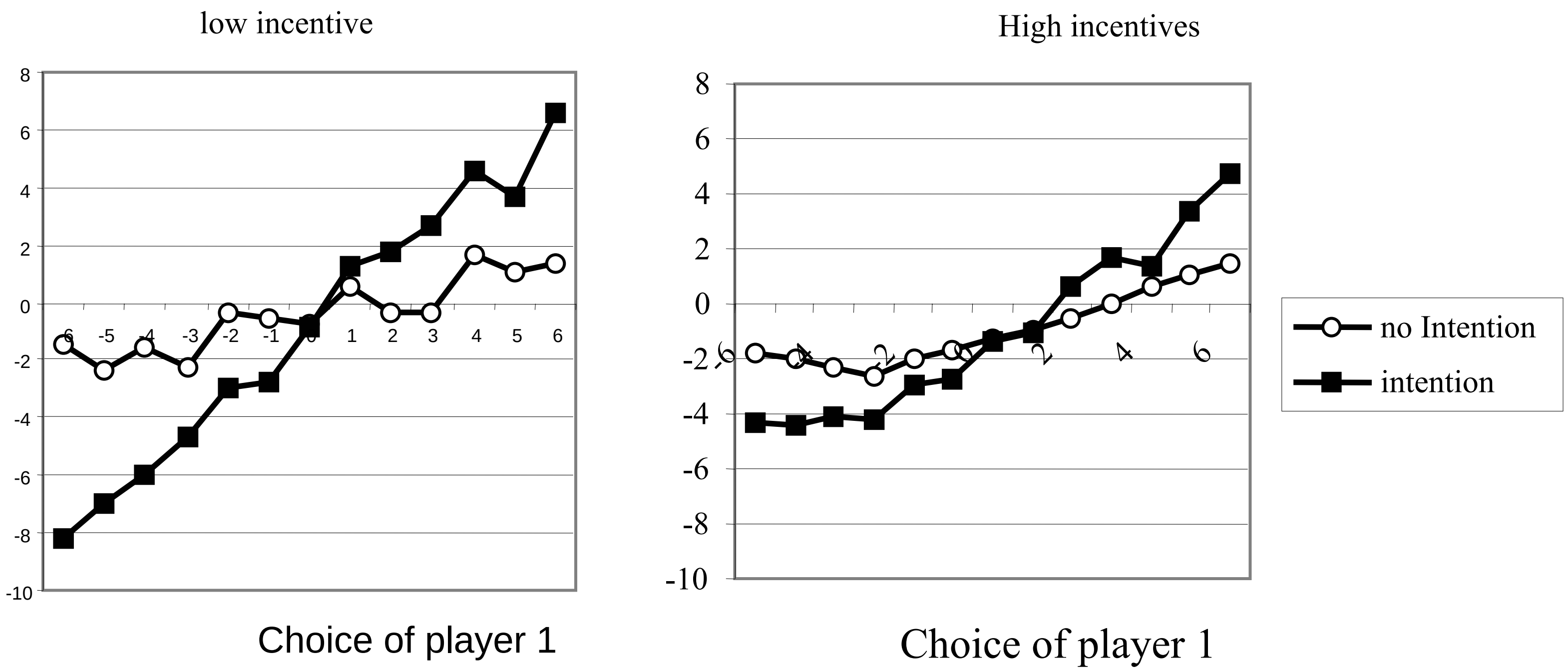
The role of intentions in reciprocal behavior

- Blount (1995)
  - The proposer's decision in the UG is determined by a random device and this is known by the responder.
  - The rejection rate for low offers is significantly smaller in the random device condition compared to the usual condition.
  - Rejections also occur in the random device condition.
- Charness (forthcoming JLE)
  - Wage offers in a gift exchange game are determined by a random device.
  - Weak but significant effects on workers' effort choices. For low wages workers provide less effort if the offer is made by a human proposer.
- Suggests that intentions (inferred personality) and outcomes are determinants of reciprocal behavior.
- Problem: the expectations of the responders are not kept constant across the random device and the human choice condition. In the random device condition the distribution of first-mover choices is not identical across treatments.

The role of intentions - continued

- Each of 2 players receives an endowment of 12 money units.
- Player 1 can give or take from player 2 up to 6 units.
    - Units given are tripled (1:3).
    - Units taken are not tripled (1:1).
- Player 2 observes 1's choice and can reward or punish player 1.
    - Reward: transfer back up to 18 units (1:1)
    - Punishment : 2 assigns "punishment" points to 1, with a cost ratio of 1:3.
- Intentions treatment: player 1 makes the choice
- No-intentions treatment: random device that exactly mimics the distribution of human first mover's determines the first mover choice. Responders know the distribution that underlies the random device.
- Strategy method for second movers.
- High stake and low stake condition. In the high stake condition the subjects' average earnings were CHF 130 ($100) – five times the earnings in the low stake condition.

Results (Source: Falk, Fehr, Fischbacher 2002)



Inequity Aversion and Reciprocity

- The FS inequity aversion model mimics reciprocity
  - Positive reciprocity: if player i's income is higher then j's income than i values j's payoff positively, i.e., she is nice to j.
  - Negative reciprocity: if player i's income is lower then j's income than i values j's payoff negatively, i.e., she is hostile to j.
  - Note, that in principle nothing prevents us from having a reference income that differs from equality.

- Consider the following utility function $U_i = x_i + \sum_j v_i(\kappa_i^j) \cdot (x_j - x_i)$

- $\kappa_i^j = x_i - x_j$ and measures the kindness of player i to player j, i.e., how much inequality to her advantage does player i accept. Define

$$v_i(\kappa_i^j) = \begin{cases} +\beta_i/(n-1) & if \quad \kappa_i^j > 0 \\ 0 & if \quad \kappa_i^j = 0 \\ -\alpha_i/(n-1) & if \quad \kappa_i^j < 0 \end{cases}$$

- Rewarding or punishing does not depend on how the present income distribution has been generated. The intentions of the other players or their personalities are irrelevant.
- It is the aim of player i to reduce the inequality between the players. If this is not possible, no rewarding or punishing occurs.

Problems of a consequentialist reciprocity model

$$U_i = x_i + \Sigma\phi(x_i - x_j)x_j$$

- Assume that ϕ is continuous and monotonically increasing and that $\phi(0) = 0$.
- Here kindness is driven by income differences but it is not the aim to generate equality.
  - Kindness ($\phi > 0$) implies that x_j is valued positively and hostility ($\phi < 0$) implies that x_j is valued negatively.
- Consider the following game: Players start with an equal endowment and 1 can reduce 2's payoff by one unit at a cost of $c<1$. If the non-pecuniary utility term is sufficiently important for 1 she will reduce the payoff of 2.
- Intuition: By punishing player 2 ϕ becomes positive so that 1 experiences a discrete non-pecuniary utility gain.
- Psychological game theory provides a solution to this problem. Is there another one available?

Where do we stand?

- In the class of inequity aversion models the FS-model outperforms the BO model because the latter systematically mispredicts punishment behavior.
- The pure reciprocity models relying on psychological game theory (Rabin, Duwfenberg & Kirchsteiger) are rather complicated and often generate multiple (implausible) equilibria even in the simplest games (e.g. the UG).
- Models combining inequity aversion and intention-based reciproicty might be the right way to go (Falk and Fischbacher 1999) but they are also quite complicated.
- The inequity aversion model of FS predicts well in a large class of games and is simple enough to be used as a building block for applications (e.g contract theory).
- The success of the model derives from two facts:
  - ➢ The fairness of outcomes matters per se.
  - ➢ The model provides a kind of reduced form for intentions or personality driven reciprocity.
  - ➢ Reduced forms greatly simplify the analysis but they have to be used with care. FS is likely to mispredict
    - if the intentionality varies across conditions.
    - if punishment cannot change the income differences.

Issues in Field Applications

- Who are the relevant reference agents?
  - Comparisons within the firms, within organizational units of a firm, across organizational units of a firm, across firms, ...
- What is the relevant reference outcome?
  - What are the established fairness standards (equity need not imply equality!). Which “moral” arguments count. Which are the relevant comparison groups?
- What is the relevant pie that can be shared?
  - A worker’s marginal product or the average product
- What are the information conditions regarding effort, the available pie, etc.
  - Do workers know the quantitative impact of their effort on the firm’s revenue?
  - Is the relation between effort and the marginal revenue of effort deterministic or stochastic? Can the worker hide behind “bad luck”.