UNDERSTANDING SOCIAL PREFERENCES WITH SIMPLE TESTS

GARY CHARNESS AND MATTHEW RABIN

Handout - Kassandra and João

Motivation: Humans do not decide/act in pure self interest. Therefore, economic experiments have recently inspired models of "social preferences". Subjects are motivated by increasing social welfare and reciprocity.

Recent models of social preferences assume that, besides self-interest, individuals also care about others' payoffs:

- Difference-aversion models: players are motivated to reduce the differences between theirs and others' payoffs;
- Social-welfare models: people are motivated to increase the social surplus, mainly through helping those with lower payoffs.

However, authors are concerned about possible confounders in the research behind these models (mainly that the games allow a departure from self-interest only through the predictions done by the models).

Data: 29 different games, with 467 participants (319 in Barcelona and 148 in Berkeley) making 1697 decisions

Social Preferences:

Considering π_A, π_B the money payoffs of Player A and Player B respectively in a two-person game, consider the following formulation of Player B utility function:

$$U_B(\pi_A,\pi_B) \equiv (\rho \cdot r + \sigma \cdot s + \theta \cdot q) \cdot \pi_A + (1 - \rho \cdot r - \sigma \cdot s - \theta \cdot q) \cdot \pi_B,$$

where $r = 1$ if $\pi_B > \pi_A$, and $r = 0$ otherwise;

$s = 1$ if $\pi_B < \pi_A$, and $s = 0$ otherwise;

$q = -1$ if A has misbehaved, and $q = 0$ otherwise.

Player B utility is a weighted sum of his own payoffs and payoffs of Player A, depending this weights on: 1) Whether A is getting a lower or higher payoff than B (dummies r and s); 2) Whether A has misbehaved - fair or unfair behavior (dummy q).

Thus, the parameters ρ and σ capture the distributional preferences of Player B, while θ captures the reciprocity mechanism.

As stated before, different models can be assumed on distributional preferences. In the two-players game, that would come as:

- <u>Narrow self-interest:</u> utility depends only on πB → $(\sigma = \rho = 0)$
- <u>Competitive preferences</u>: people like their payoffs to be high relative to others' payoffs → $(\sigma \leq \rho \leq 0)$
- <u>Difference-aversion</u>: player B likes money and preferes that payoffs are equal, including wishing to lower player A's payoffs when player A does better than player B → $(\sigma < 0 < \rho < 1)$
- <u>Social-welfare preferences</u>: player B prefers more for him and player A, and he is more in favor of getting payoffs to himself when he is behind than when he is ahead. → $(1 \geq \rho \geq \sigma \geq 0)$

However, all models seem to fail to capture some experimental evidence. One way to approach this mismatch is by accounting for reciprocity. Reciprocal preferences are captured by θ and indicate that players change their weights on other payoffs depending on their behavior (e.g.: if Player A pursues self-interest at the expense of social-welfare preferences, Player B will value less the payoffs of Player A than he would otherwise).

Experimental Procedures and Result:

Either one or two participants made decisions, and decisions affected the allocation to either two or three players.

<u>Two-player games</u>: Decision of A, or A and B about allocation of money to A and B.

<u>Three-player games</u>: Decision of C, or both A and C about money allocation to players A, B, and C.

Idea behind: have a great variety of games where different social preferences models predict different behaviors.

<u>Results on Distributional Experiences</u>: assuming no reciprocity $(\theta = 0)$ how many observations are consistent with the predictions for the parameters $(\sigma \; and \; \rho)$ made by each model (of distributional preferences).

As the authors assume no reciprocity, the analysis is more appropriate based on dictator games (since results in response games may be influenced by it).

- Main Result: social-welfare preferences models are much more successful in explaining the results when compared with other models (Table II and III).
- Difference-aversion behavior is verified when there is no sacrifice, but it is far less common when there are sacrifices required (so 31% of individuals B choose (400,400) over (750,400), but none chooses (0,0) over (800,200)). Difference-aversion behavior is also challenged when almost half of players B choose (750,375) over (400,400).

- Barc8 and Barc15 are consistent with both difference-aversion and social-welfare preferences. However, when compared, they show that player B is less willing to give up from 100 to help A in 400 when, by doing so, he receives less than player A (so 73% of players B choose (600,600) over (200,700), but only 33% chooses (700,500) over (300,600)).

When it comes to the behavior of Player A, we then look at response games.

Two ways of interpreting this question :

1. Assuming A's choice is consistent with restriction on preferences if his choice is consistent given any belief about B's behavior.
2. Assuming A correctly anticipates B's behavior.

Some of A's choices are inconsistent with given models. However, difference aversion and social-welfare have some predictive power.

How does the sacrifice of player B affect inequality (Table IV)

Player B took pareto-damaging decisions 17% of the times that he was able to do so.

- Of those, 15% decreased inequality and 22% increased inequality (against the predictions of difference-aversion)
- Conclusion: although authors argue that in a broader setting this pattern (of increasing inequality pareto-damaging decisions) would likely not hold, it brings more questions about the strong link that difference-aversion does between sacrifice and inequality-reduction.

How do the different models help to predict B sacrifice (Table V):

- Sacrifice decisions from player B seem to be more consistent with social-welfare preferences than with difference-aversion. However, the sacrifice required to be consistent with social-welfare preferences was, on average, lower than the sacrifice required to be consistent with difference-aversion, what is likely to partially drive the results (at least in terms of magnitude)

Reciprocity:

Main idea: draw different games with identical choices for B in which in some of those games his decision was preceded by a choice made by A and in others it was not

Results:

- When A took no previous action, B chose (400,400) over (750,400) 31% of the time. When A foregone the option (750,0) to let B play, player B chose (400,400) over (750,400) only 6% of the time. → relative strength of positive reciprocity versus difference aversion when self-interest is not implicated.

- When A foregone (550,500) to let B choose between (400,400) and (750,400), the results are very similar to those when B chooses directly between (400,400) and (750,400) - 69% of the time player B chooses the second option → surprising result, as player B does not punish Player A for his unfair decision.
- Authors find no evidence of positive reciprocity, and only weak negative reciprocity. Nevertheless, there is evidence of concern withdrawal (one specific realization of negative reciprocity): B is less likely to withdraw his willingness to sacrifice to give the social-welfare-maximizing allocation to A if A has behaved selfishly.
- A violation of social-welfare norms plays a stronger role in determining when a person sacrifices to help another than it plays in determining when a player sacrifices to harm another.

Overall, authors find that reciprocity considerations are an important component of behavior.

Multiple Games:

Main interest: how do people feel about changes in the distribution among others' payoffs.

Results in dictator games:

- People care about both the total surplus and the minimum payoffs among others (some players chose the first, many others preferred the second).
- Social efficiency is not the only distributional driving force (players C chooses much more often (1200,0,x) than (750, 350, x), although the first is socially more efficient - Table I, Berc 10/12).
- In contrast to what previous research has shown, individuals care about the allocation among other people: 54% of players C sacrificed to equalize payoffs with each of the other players, even if the difference between a player's own payoff and the average of other players' payoffs kept constant (Table I, Berc 24). Although this is consistent with both social-welfare preferences and difference-aversion models, the authors argue that the percentage of players willing to sacrifice nothing to eliminate disadvantageous inequalities against themselves is lower than 54%, what seems to discard the option that difference-aversion behavior is driving the results.

Results in responsive games:

- Strong evidence in responder behavior: under certain circumstances, reciprocity seems to be a stronger primer for behavior than distributional preferences - when A took a selfish decision, C chose to give A the lowest payoff independently of the distributional consequences (Table I - Berc 16/20)

Summary:

✔ Previous difference aversion models have been confounded by games whose results that departure from self-interest behavior were leading automatically to difference-aversion predictions (what were also the predictions of reciprocity, and so it was not possible to distinguish between both of them)

✔ Authors propose a model with a focus on distributional preferences and reciprocity, which allow different types of preferences:

- competitive preferences
- narrow self-interest
- difference aversion preference
- social welfare preferences

✔ Social welfare preferences have a higher explanatory power than any other distributional preference model.

✔ As suspected by the authors, the role of difference-aversion predictions has been overstated. Its prediction power is lower than social-welfare preferences, and its correct predictions do often coincide with correct predictions of both social-welfare preferences and reciprocity (meaning that a individual might be behaving based on social-welfare or reciprocity preferences, but the results coincide with difference-aversion predictions)

✔ Reciprocity plays a role in decision making. However, negative reciprocity is higher than positive reciprocity. The role of reciprocity is mainly found in a 3 player responsive game.