Heinrich H. Naxa, Ryan O. Murphya, Kurt A. Ackermann

Interactive preferences

Article (Accepted version) (Refereed)

Original citation:
Nax, Heinrich H., Murphy, Ryan O. and Ackermann, Kurt A. (2015) *Interactive preferences. Economics Letters*, 135. pp. 133-136. ISSN 0165-1765

DOI: 10.1016/j.econlet.2015.08.008

Interactive preferences

Heinrich H. Nax[a], Ryan O. Murphy[a], Kurt A. Ackermann[a]

[a]*Department of Social Sciences, ETH Zürich, Clausiusstr. 37, 8092 Zurich, Switzerland.*

Abstract

Game theory presumes that agents have unique preference orderings over outcomes that prescribe unique preference orderings over actions in response to other players' actions, independent of other players' preferences. This independence assumption is necessary to permit game-theoretic best response reasoning, but at odds with introspection, because preferences towards one another often dynamically depend on each other. In this note, we propose a model of interactive preferences. The model is validated with data from a laboratory experiment. The main finding of our study is that pro-sociality diminishes over the course of the interactions.
Keywords: game theory, social preferences, preference evolution

1. Introduction

Mother Teresa does not defect in prisoners' dilemmas, because she cares for her opponents in ways that transform the games' mixed motives into other games where her and common motives are aligned (e.g., harmony). Cooperation thus emerges as a dominant strategy. The experimental economics literature is concerned with 'subjective expected utility corrections' (Gigerenzer and Selten, 2001) that modify players' utility representations to account for such other-regarding concerns. Numerous corrections have been proposed (e.g., Rabin 1993; Levine 1998; Fehr and Schmidt 1999; Bolton and Ockenfels 2000 in light of laboratory evidence that manifests systematic deviations from narrow self-interest predictions (see Ledyard 1995 and Chaudhuri 2011 for reviews).[1] This route of enquiry is bothersome for many theoretical game theorists who question how these findings generalize beyond the laboratory.[2]

Missing from most alternative utility formulations are interactive components that meaningfully alter the game-theoretic analysis. Standard theory (von Neumann and Morgenstern, 1944) equips players with preferences that prescribe actions vis-à-vis others' actions, independent of others' preferences.

Email address: `hnax@ethz.ch` (Heinrich H. Nax)

[1]With some exceptions (e.g., Saijo and Nakamura 1995; Saijo 2008, many analytical setups have been biased as discussed in, for example, Burton-Chellew and West (2013); Burton-Chellew et al. (2015).

[2]See controversies in JEBO *73*, 2010.

Preprint submitted to EL *July 27, 2015*

Here, inspired by Rabin (1993) and Levine (1998), we propose a model of interactive preferences among players that depend on each other and investigate their dynamic interdependence. The model is validated with laboratory studies involving repeated voluntary contributions games (VCM; Isaac et al. 1985) sandwiched by two sets of dictator games (DG; Kahneman et al. 1986) used to evaluate individuals' social value orientation (SVO; Murphy et al. 2011). Our results show that, independent of unintended behavioral deviations, the proportion of behavior associated with pro-sociality diminishes over the course of the interactions and is replaced by individualism. These patterns carry over between VCMs and DGs. Our model fares predictively well.

2. Methods

2.1. Experimental setup

Experiments were run at ETH's Decision Science Laboratory during February 2013 involving 128 subjects in 6 sessions (4∗20+2∗24). Subjects were informed in detail and in advance of each stage of the experiment using standard instructions.[3] Every decision was monetarily incentivized, and subjects earned over 40CHF>40US$ on average. The experiment lasted roughly 90 minutes.

The experiment had the following three stages:

Stage 1: ***Initial SVO.*** Subjects played 6 DGs choosing allocations in different ranges representing different himself-versus-other tradeoffs; for example, between 100 for himself and 50 for the other, (100,50), and 50 for himself and 100 for the other, (50,100).[4] The 6 decisions are represented as angles in the classical SVO ring (Griesinger and Livingston, 1973), and an individual's initial SVO is taken as the average angle, representing a compact indicator of his *ex ante* SVO.[5]

Stage 2: ***VCM.*** Subjects played 10 VCMs in groups of 4 that were randomly formed in round 1 and then remained fixed for the remainder of rounds. In each round subjects made contributions and guessed others' average contributions (with incentives for accuracy). Before each round, players were informed of the previous-period contributions. (More detail will be provided shortly.)

Stage 3: ***Final SVO.*** Stage 1 is repeated, thus measuring individuals' *ex post* SVOs.

Our analysis focusses on 22 data points p.p., namely his 2 – initial and final – SVOs, plus his 10 contributions and 10 guesses about others' contributions from the VCM, yielding a total of 2,816 data points.

[3] See Murphy and Ackermann (2013) for details.

[4] The remaining 5 choices are amongst linear combinations in the ranges $[(100, 50), (85, 85)]$, $[(50, 100), (85, 15)]$, $[(50, 100), (85, 85)]$, $[(85, 85), (85, 15)]$, and $[(85, 15), (100, 50)]$.

[5] Angles close to $0°$ represent individualistic preferences in the the sense of material self-interest, angles $\geq 22.5°$ indicate pro-sociality.

2.2. The model

2.2.1. Static model

Population $N = \{1, 2, 3, 4\}$ plays a VCM with marginal per capita rate of return $r = 0.4$ and budget $B = 20$. Each $i \in N$ sets a private contribution $c_i \in B$ which, jointly with the others' average contribution, c_{-i}, results in payoff

$$\phi_i = 20 - c_i + 0.4(c_i + 3c_{-i}).$$

We assume i's utility depends on payoffs in Cobb-Douglas form

$$u_i(c) = \phi_i^{1-\alpha_i} * \phi_{-i}^{\alpha_i}, \tag{1}$$

where $\alpha_i \in [0, 1]$ measures player i's concern for others. The nonlinearity of expression 1 distinguishes it from most representations, including Levine (1998), thus rationalizing intermediate contributions in terms of intermediate concerns. We obtain the following expression for α_i by assuming c_i is chosen optimally given his guess about c_{-i} (expressed as $\widehat{c}_{-i}$):

$$\alpha_i = \frac{0.6\phi_{-i}(c_i, \widehat{c}_{-i})}{0.4\phi_i(c_i, \widehat{c}_{-i}) + 0.6\phi_{-i}(c_i, \widehat{c}_{-i})} \tag{2}$$

Note that $\partial\alpha_i/\partial c_i > 0$ and $\partial\alpha_i/\partial\widehat{c}_{-i} < 0$, that is, higher own contributions (holding beliefs about others constant) indicate more concern for others, and higher beliefs regarding others' contributions (keeping own contributions fixed) indicate less concern for others.

The interdependence of preferences results from imposing that, in static equilibrium, $\alpha_i = \widehat{\alpha}_{-i}$, where $\widehat{\alpha}_{-i}$ is i's belief about α_{-i}.[6] The resulting set of equilibria, the general structure of which is under investigation in an ongoing study, contains the standard case (when $\alpha_i = \alpha_{-i} = 0$) but also new ones when $\alpha_i = \alpha_{-i} > 0$ as in fairness equilibria (Rabin, 1993).

2.2.2. Dynamic components

The above game repeats with revelation of past outcomes. Each period t, suppose i contributes to maximize expression 1 so that expression 2 implies α_i^t given $(c_i^t, \widehat{c}_{-i}^t)$. We assume α_i^t is updated in light of evidence by

$$\alpha_i^t = (1 - \beta_i^t)\alpha_i^{t-1} + \beta_i\widetilde{\alpha}_{-i}^{t-1}, \tag{3}$$

where $\widetilde{\alpha}_{-i}^{t-1}$ is i's deduction of α_{-i}^{t-1} from previous-period evidence, and $\beta_i^t \in [0, 1]$ measures i's period-t degree of belief responsiveness.

[6] A weaker assumption in the same spirit would be to weigh this dependence by some parameter as in Levine (1998), something we shall introduce via 'responsiveness' instead.

2.3. Estimation strategy

2.3.1. Classification

Initial SVOs are used to classify individuals as 'individualistic' and 'pro-social'. An individual is pro-social (individualistic) according to the SVO measure if his SVO-angle is ≥ 22.5 (< 22.5) degree.[7] The initial SVO classifications are used to predict initial VCM contributions

'Responsive' and 'unresponsive' types are classified based on the VCM data. Individual i is said to be responsive (unresponsive) if the estimation of expression 3 in light of his VCM decisions from rounds 2-10 yields an average coefficient for β_i^t which is positive (not positive).

2.3.2. Prediction

We use our estimated 2×2 typology (from initial SVO and VCM) to make predictions regarding final SVO classifications, which we shall assess in light of the recorded final SVOs. We shall use the following terminology: an individual is associated with a VCM group matching that is said to be 'individualistic' ('pro-social') if those players he is matched with, on average, contribute less (more) than himself.

We predict unresponsive types (pro-social and individualistic alike) not to change their preferences. We predict responsive types to change their types in the direction of their interaction partners as matched with during the VCM group matching. Hence, a responsive pro-social (individualist) in a VCM group matching that is pro-social (individualistic) will remain pro-social (individualistic). A responsive pro-social (individualist) matched with individualistic (pro-social) others, however, may become individualistic (pro-social), dependant on the action/payoff difference between himself and his opponents. In particular, whichever payoff difference is larger we shall assume will be associated with a preference-change flow of probability one, and the lesser payoff-difference to be proportional to that flow depending on the relative payoff difference.

Table 1: Regressions 1 and 2 (standard errors adjusted for 128 individual clusters)

Regression 1		Regression 2	
'Contribution' (VCM, t=1)		'Responsiveness' (VCM, t=1-10)	
Initial pro-sociality	3.54^* (1.19)	α^{t-1}	-0.35^* (0.04)
Constant	10.76^* (2.72)	$\widetilde{\alpha}_{-i}^{t-1}$	0.44^* (0.15)
Controls	not listed	Controls	not listed
N	128	N	1,152
R^2	0.13	R^2	0.20

* : significance level < 0.01

[7]See Murphy and Ackermann 2013 for a more fine-grained categorization.

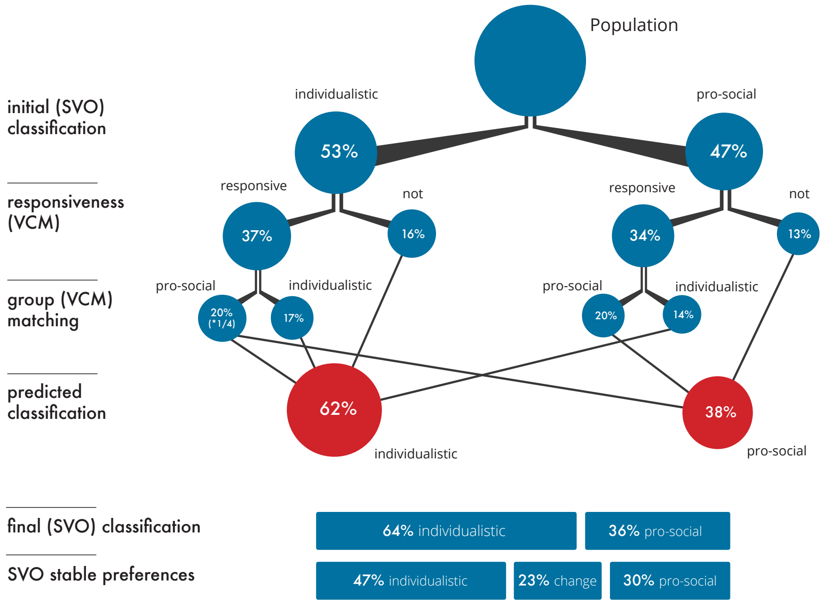

Figure 1: Flow chart of classification and preference prediction.

3. Results

Initial SVOs classify 53% of players as individualistic and 47% as pro-social, and pro-socials give over 30% more in period 1 of the VCM (regression 1). Expression 3 is structurally confirmed at the population level in the VCM data (regression 2). Re-running regressions for expression 3 at individual levels (omitted) for the VCM, we find 71% responsives (34% pro-socials, 37% individualists); 14% (20%) are responsive pro-socials (individualists) matched by chance in individualistic (pro-social) groups.

It is amongst those 34% matched in opposite groups where we expect preference interactions to materialize. An average of 2.3 coins less was earned by the 14% responsive pro-socials in individualistic groups versus 0.6 more by responsive individualists in pro-social groups. Hence, flowing from (responsive) individualistic to pro-social, we expect ca. 1/4 ($\approx$0.6/2.3) of the flow from (responsive) pro-social to individualistic. See figure 1 for a comparison of predicted and actual classifications.

Predictions compare with the data as follows. Final SVOs categorize 64% individualists and 36% pro-socials (62% and 38% predicted). 47% are individualistic in initial and final SVOs, which means that 6% individualists turned pro-socials (5% predicted). 30% were pro-social in both, hence 17% pro-socials turned individualists (14% predicted). The model made two types of errors. First, 7% changed preferences whom we classified unresponsive. Second, we predicted 1% (3%) too few individualists turning pro-socials (vice versa), thus incorrectly predicting flow of 3% responsives. Overall, our model was therefore accurate in predicting global preferences (95%), less in individualizing flow (90%).

4. Discussion

Individuals become less (more) pro-social when interacting with individualists (pro-socials). On average, there is a trend toward individualism over the course of the VCM, independent of the contribution decay. Our result is therefore not a byproduct of learning. Even in the sterile and anonymous context of the laboratory we found evidence for interactive preferences among players that depend on each other and evolve over time. Our model explains indirect reciprocity (Alexander, 1987; Fischbacher and Gächter, 2010) as driven by natural dynamics governing the interactions of preferences. Since stakes and intentions of players certainly matter more outside the laboratory, such phenomena are likely not to be artifacts. Preference dynamics should therefore be studied further, as the long-run predictions of models without preference interactions are potentially misguided.

Acknowledgement. Nax's research is supported by the European Commission through the ERC Advanced Investigator Grant 'Momentum' (Grant No. 324247).

Bibliography

RD Alexander. *The Biology of Moral Systems.* De Gruyter, 1987.

GE Bolton and A Ockenfels. ERC: A theory of equity, reciprocity, and competition. *AER*, 90:166–193, 2000.

MN Burton-Chellew and SA West. Prosocial preferences do not explain human cooperation in public-goods games. *PNAS*, 110(1):216–221, 2013.

MN Burton-Chellew, HH Nax, and SA West. Payoff-based learning explains the decline in cooperation in public goods games. *Proc. Roy. Soc. B*, 282(1801): 20142678, 2015.

A Chaudhuri. Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *EE*, 14(1):47–83, 2011.

E Fehr and KM Schmidt. A theory of fairness, competition, and cooperation. *QJE*, 114:817–868, 1999.

U Fischbacher and S Gächter. Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *AER*, 100(1):541–556, 2010.

G Gigerenzer and R Selten. *Bounded Rationality: The Adaptive Toolbox.* MIT Press, 2001.

DW Griesinger and JW Livingston. Toward a model of interpersonal motivation in experimental games. *Behavioral Science*, 18(3):173–188, 1973.

RM Isaac, KF McCue, and CR Plott. Public goods provision in an experimental environment. *J. Pub. Econ.*, 26(1):51–74, 1985.

A Kahneman, JL Knetsch, and RH Thaler. Fairness and the assumptions of economics. *Journal of Business*, 59(4):285–300, 1986.

JO Ledyard. Public goods: A survey of experimental research. In JH Kagel and AE Roth, editors, *The Handbook of Experimental Economics*, pages 111–194. Princeton University Press, 1995.

DK Levine. Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, 1(3):593–622, 1998.

RO Murphy and KA Ackermann. Explaining behavior in public goods games: How preferences and beliefs affect contribution levels. *SSRN 2244895*, 2013.

RO Murphy, KA Ackermann, and MJJ Handgraaf. Measuring social value orientation. *Journal of Judgment and Decision Making*, 6:771–781, 2011.

M Rabin. Incorporating fairness into game theory and economics. *AER*, 83(5): 1281–1302, 1993.

T Saijo. Spiteful behavior in voluntary contribution mechanism experiments. In CR Plott and VL Smith, editors, *Handbook of Experimental Economics Results*, volume 1, pages 802–816. 2008.

T Saijo and H Nakamura. The "spite" dilemma in voluntary contribution mechanism experiments. *Journal of Conflict Resolution*, 39(3):535–560, 1995.

J von Neumann and O Morgenstern. *Theory of Games and Economic Behaviour*. Princeton University Press, 1944.