Factors of cooperative behaviour in experimental games

Alexis Belianin
Higher School of Economics and IMEMO RAS
icef-research@hse.ru

April 6, 2011

Public goods (PG) game with voluntary contribution mechanism (VCM)

- Factors of cooperative behaviour are of interest to the economists, especially when this behaviour is disequilibrium (e.g. investment game, trust game, ultimatum game, public goods game)
- Recent behavioural interpretations (e.g. McKelvey and Palfrey, 1998; Fehr and Schmidt, 1999; Falk and Fischbacher, 2003) are important, but sometimes lack empirical background
- Empirical attempts (e.g. Camerer e.a., 2003; Stahl, 2008) are more useful, but sometimes restrictive.
- One more of these: estimation of factors of punishment and spite in public goods games using structural model.

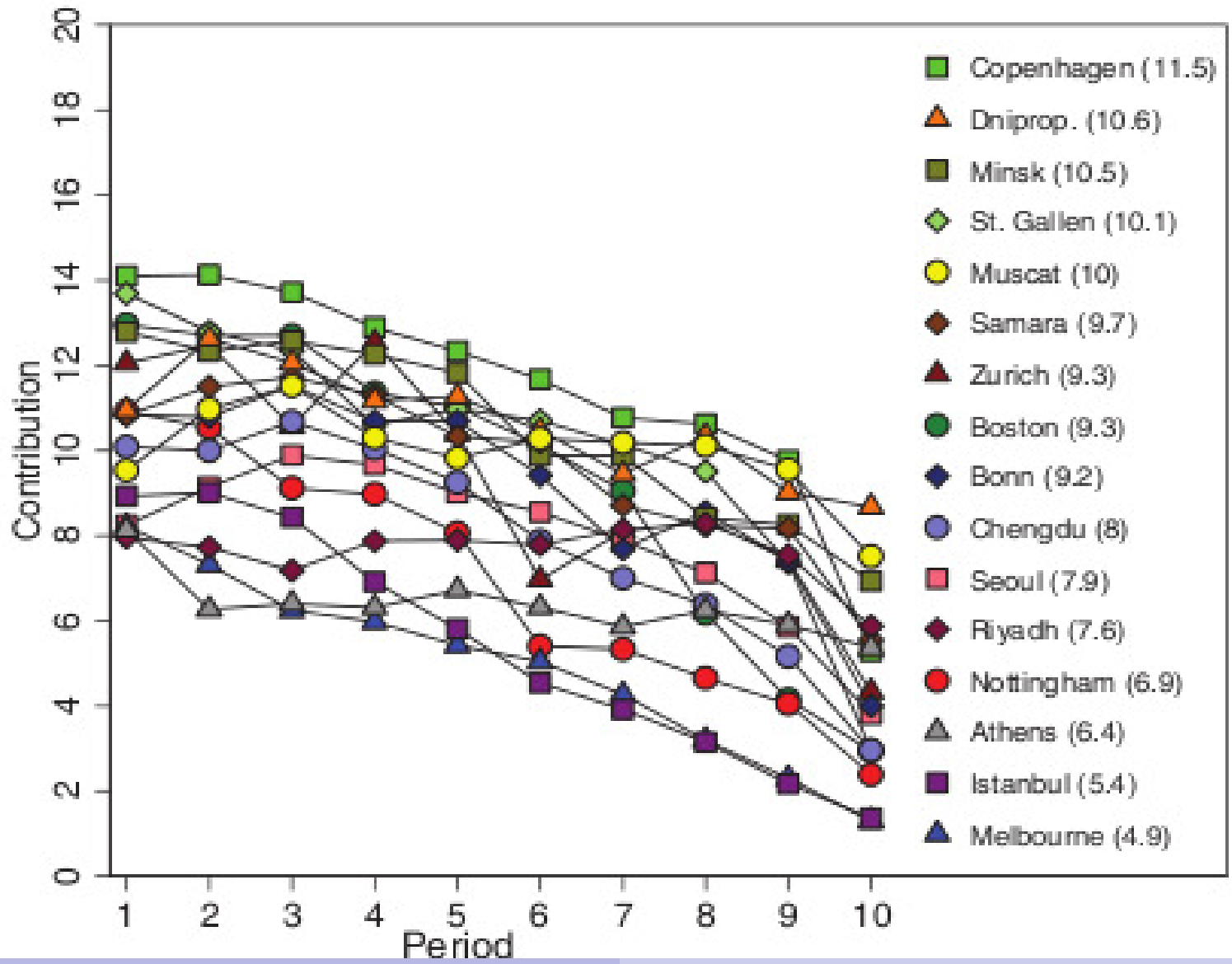
Public goods (PG) game with voluntary contribution mechanism (VCM)

- $n \geq 2$ players endowed with w units per period each (normalized to 1)
- Each player i independently decides what fraction $c_i, 0 \geq c_i \geq 1$ she will contribute to the public good, retaining $1 - c_i$.
- Return from public good is $k \cdot \sum_i c_i = \alpha \bar{c}$, where $\bar{c} = \frac{\sum_i c_i}{n}$ and $\alpha = kn, k < 1 < kn$ is efficiency factor.

$$v_i(c_i, \bar{c}) = 1 - c_i + \alpha \bar{c} = 1 - c_i + k \cdot \sum_i c_i \equiv v_i \qquad (1)$$

The only Nash equilibrium is zero contribution, while Pareto-optimal is 100% contribution

PG with VCM: typical results (Herrmann, Gächter, Thoni, 2009)



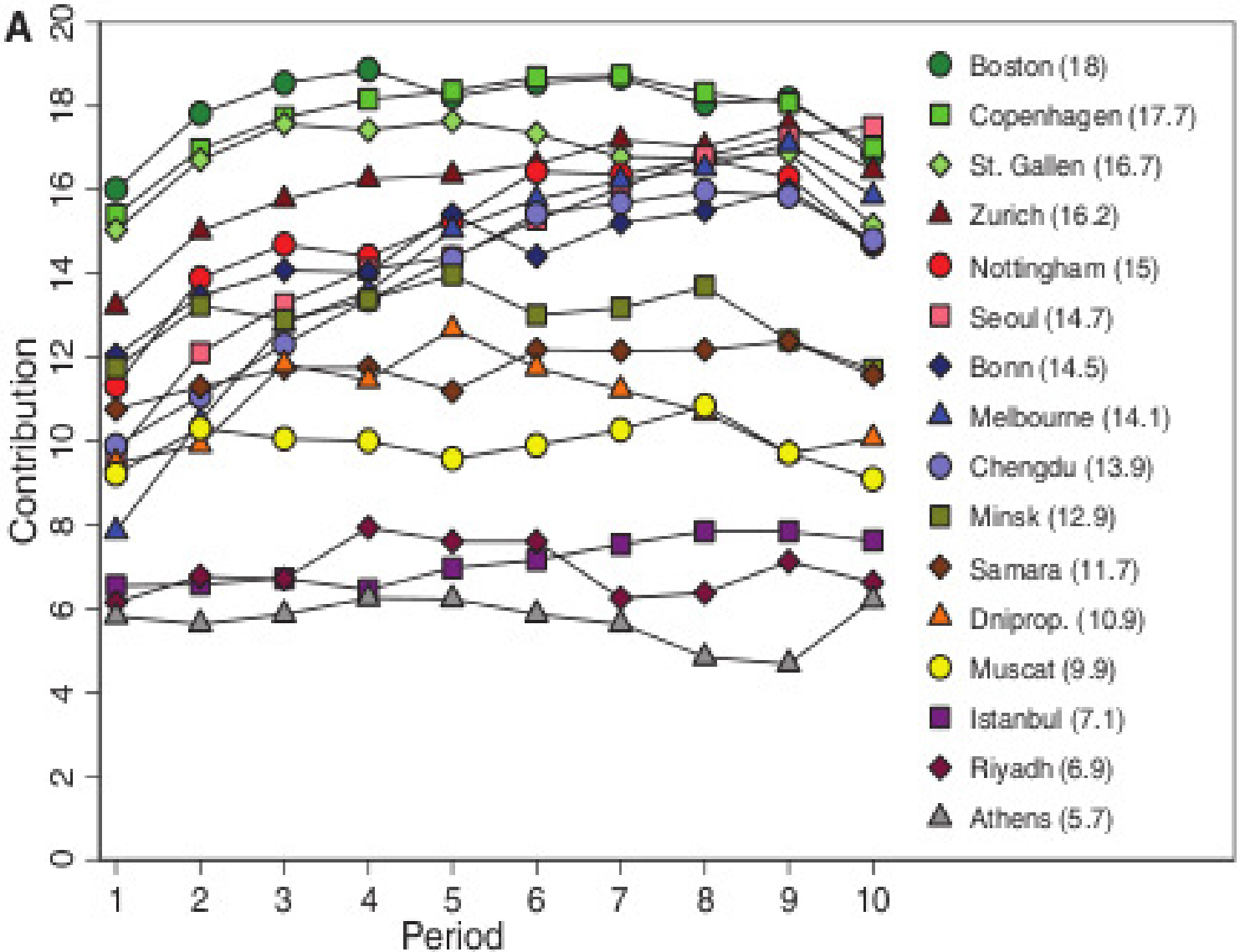
Public goods game with VCM and punishment

After the contribution stage, all players are informed about individual contributions, and can punish each other player j (not herself!) by p_{ij} units at a cost sp_{ij} units to themselves, where $s < 1$. Total payoff to player i is then

$$V_i(\mathbf{c}, \mathbf{P}) = v_i - s \sum_{j \neq i} p_{ij} - \sum_{j \neq i} p_{ji} \tag{2}$$

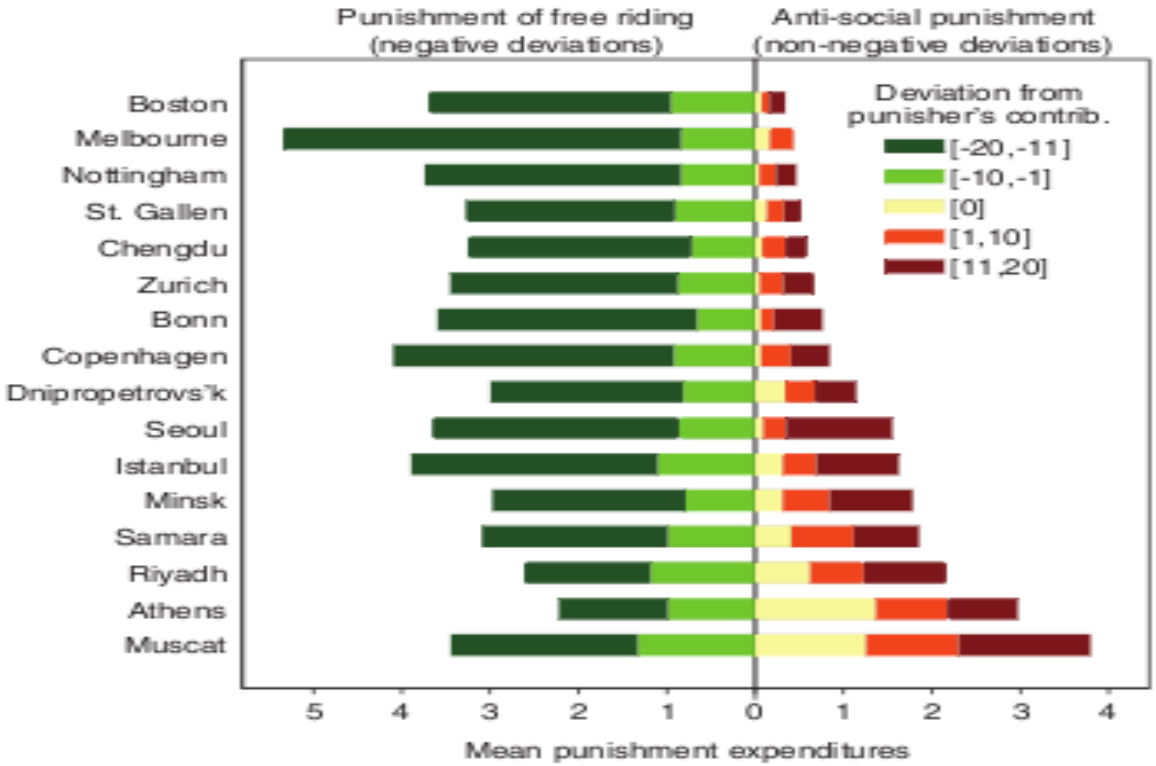
Punishments are known to increase the degree of cooperativeness, especially in with time and in partner treatments.
Mechanism: punishment (threaten, expression of disapproval) of those who free-ride boosts up cooperativeness.

PG with VCM: typical results (Herrmann, Gächter, Thoni, 2009)



Spiteful (antisocial) punishment (Herrmann, Gächter, Thoni, 2009)

Sometimes players punish not only those who contributed less, (free-riders — *prosocial* punishment), but also those who conributed more than they did (*spiteful*, or antisocial punishment)



Middle East, Russia and Eastern Europe are world leaders in spite

Spiteful (antisocial) punishment

...or are they?

- What are the origins for spiteful punishment?

Spiteful (antisocial) punishment

...or are they?

- What are the origins for spiteful punishment?
- More generally: Is punishment necessarily an expression of ethical disapproval (retaliation for low contributions?)

Spiteful (antisocial) punishment

...or are they?

- What are the origins for spiteful punishment?
- More generally: Is punishment necessarily an expression of ethical disapproval (retaliation for low contributions?)
- Yet more generally: what are the motives for punishment behaviour?

Classification of possible motives for punishment

Availability — presense of punishment option is suggestive in itself — the Chekhov motive.

Classification of possible motives for punishment

Availability — presense of punishment option is suggestive in itself — the Chekhov motive.

Tolerance — culturally-defined punishment is something 'customary' and 'acceptable' — the Tjutchev motive.

Classification of possible motives for punishment

Availability — presense of punishment option is suggestive in itself — the Chekhov motive.

Tolerance — culturally-defined punishment is something 'customary' and 'acceptable' — the Tjutchev motive.

Competitiveness — punishment as an efficient way to improve own relative standing in the group.

Classification of possible motives for punishment

Availability — presense of punishment option is suggestive in itself — the Chekhov motive.

Tolerance — culturally-defined punishment is something 'customary' and 'acceptable' — the Tjutchev motive.

Competitiveness — punishment as an efficient way to improve own relative standing in the group.

Preemption — penalizing because one expects penalties from the others.

Classification of possible motives for punishment

Availability — presense of punishment option is suggestive in itself — the Chekhov motive.

Tolerance — culturally-defined punishment is something 'customary' and 'acceptable' — the Tjutchev motive.

Competitiveness — punishment as an efficient way to improve own relative standing in the group.

Preemption — penalizing because one expects penalties from the others.

Upset — negative feeling at what the others have contributed, leading to the desire for retaliation.

$c_i - c_j$, difference between contributions.

$\hat{c}_i - c_j$, difference between believed norm and factual contribution.

$\bar{c} - c_j$, , difference between group norm (mean) and factual contribution.

Classification of possible motives for punishment

Availability — presense of punishment option is suggestive in itself — the Chekhov motive.

Tolerance — culturally-defined punishment is something 'customary' and 'acceptable' — the Tjutchev motive.

Competitiveness — punishment as an efficient way to improve own relative standing in the group.

Preemption — penalizing because one expects penalties from the others.

Upset — negative feeling at what the others have contributed, leading to the desire for retaliation.

- $c_i - c_j$, difference between contributions.
- $\hat{c}_i - c_j$, difference between believed norm and factual contribution.
- $\bar{c} - c_j$, , difference between group norm (mean) and factual contribution.

Spite per se — genuine disapproval of those who behave pro-socially.

Design: baseline after Gächter and Herrmann (2008)

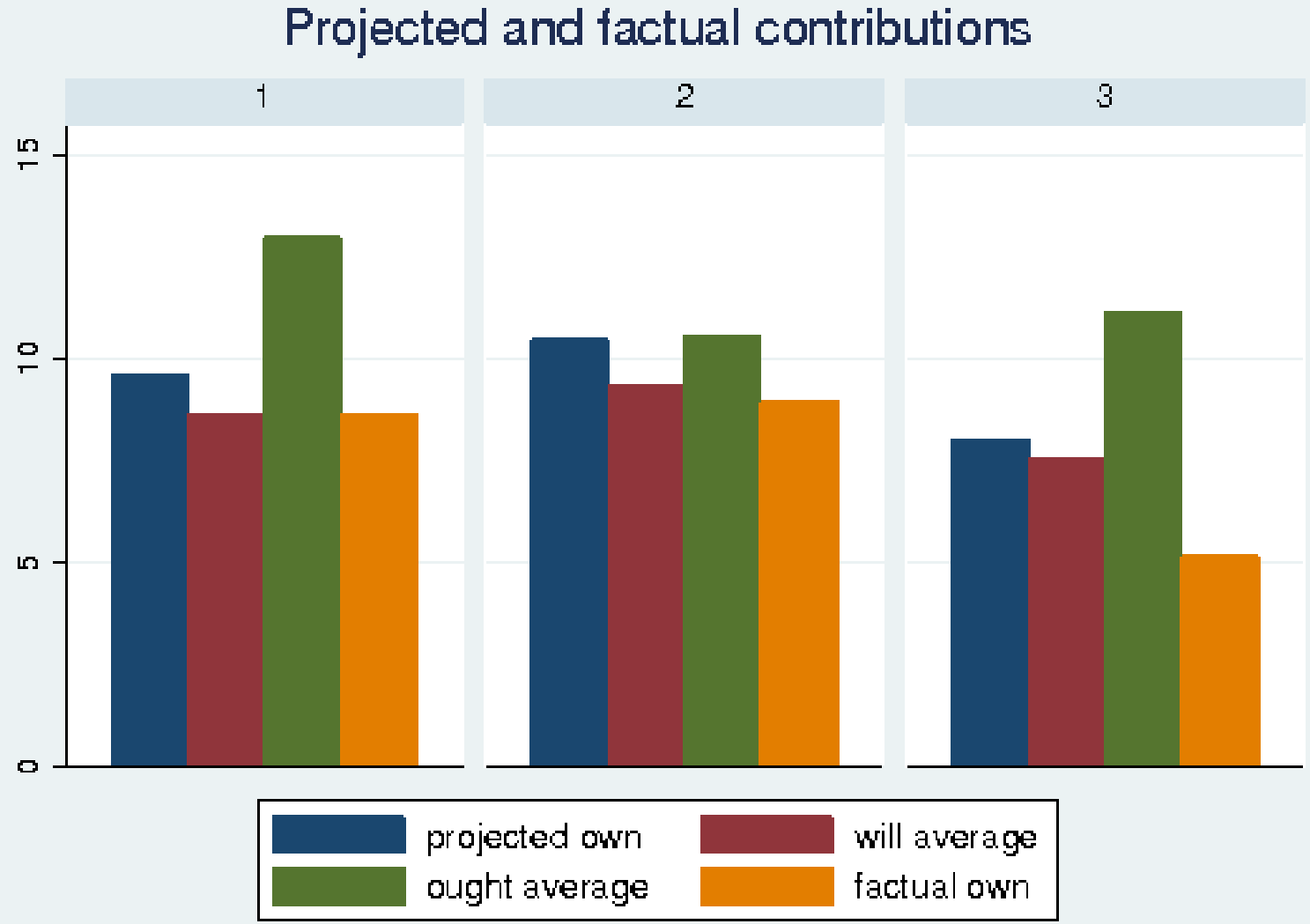
- 2 single-shot games: VCM without punishment, followed by VCM with punishment (2 games altogether).
- Groups of $n = 4$ players, endowment 20, efficiency factor $k = 1.6$ ($\alpha = 0.4$) for all subjects.
- After each contributions stage, participants observe contributions and payoffs of all groupmates.
- Cost of punishment from 0 to 10 either low (0.1) or high (0.5).
- Preceding instructions with worked examples and exercises to check understanding.
- Ex ante intentions questionnaire other than oneself and the punished one, in proportion to their contributions.
- Post-punishment treatments introduced at the end.

Participants: 247 full-time and part-time students from Moscow (75), Perm (76) and Tomsk (96) (sample to be completed).
Average payoff — 208 RuR.

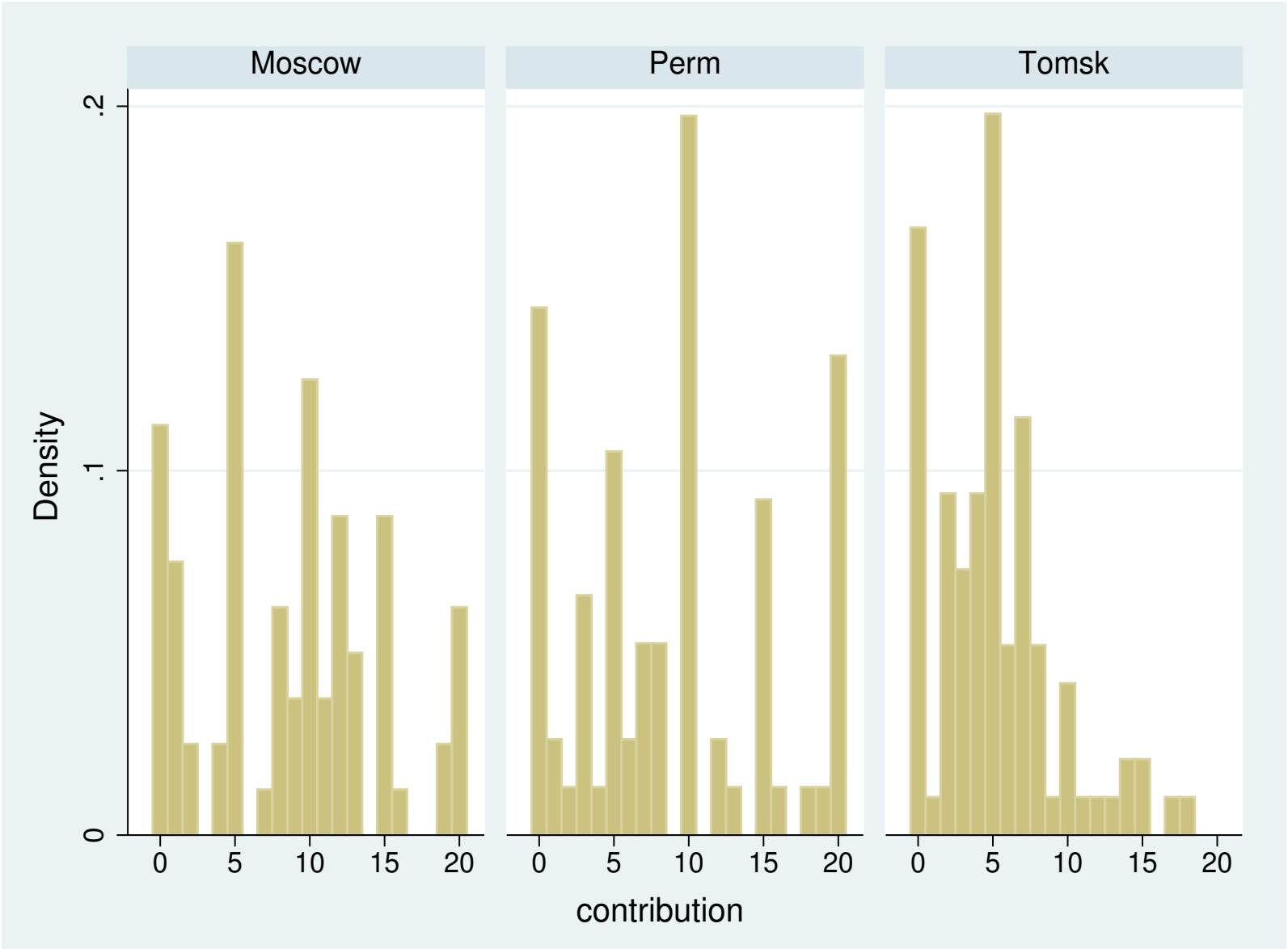
Design: additions

- Intentions questionnaire asks for *planned* own contirubtions, the *due* average and *expected* average contributions in their group, and desired contribution level if the group average turns out to take discrete values of 0, 3, 6, 10, 14 and 17 units, evaluated by strategy method.
- In a separate screen with *yes-no* button shown after the contributions stage, the subject has to choose 'yes' iff (s)he wants to assign deduction points to at least one of his or her group fellows (test for availability).
- After punishment stage, subjects in the low cost of punishment sessions could purchase *insurance against punishment* of up to 10 units from each individual player in her group, at a cost of 0.1 if redistributed from punishment, and 0.2 per unit of insurance.
- Assignment one's punishments to burn it out or to redistribute among the remaining two participants.

Contributions



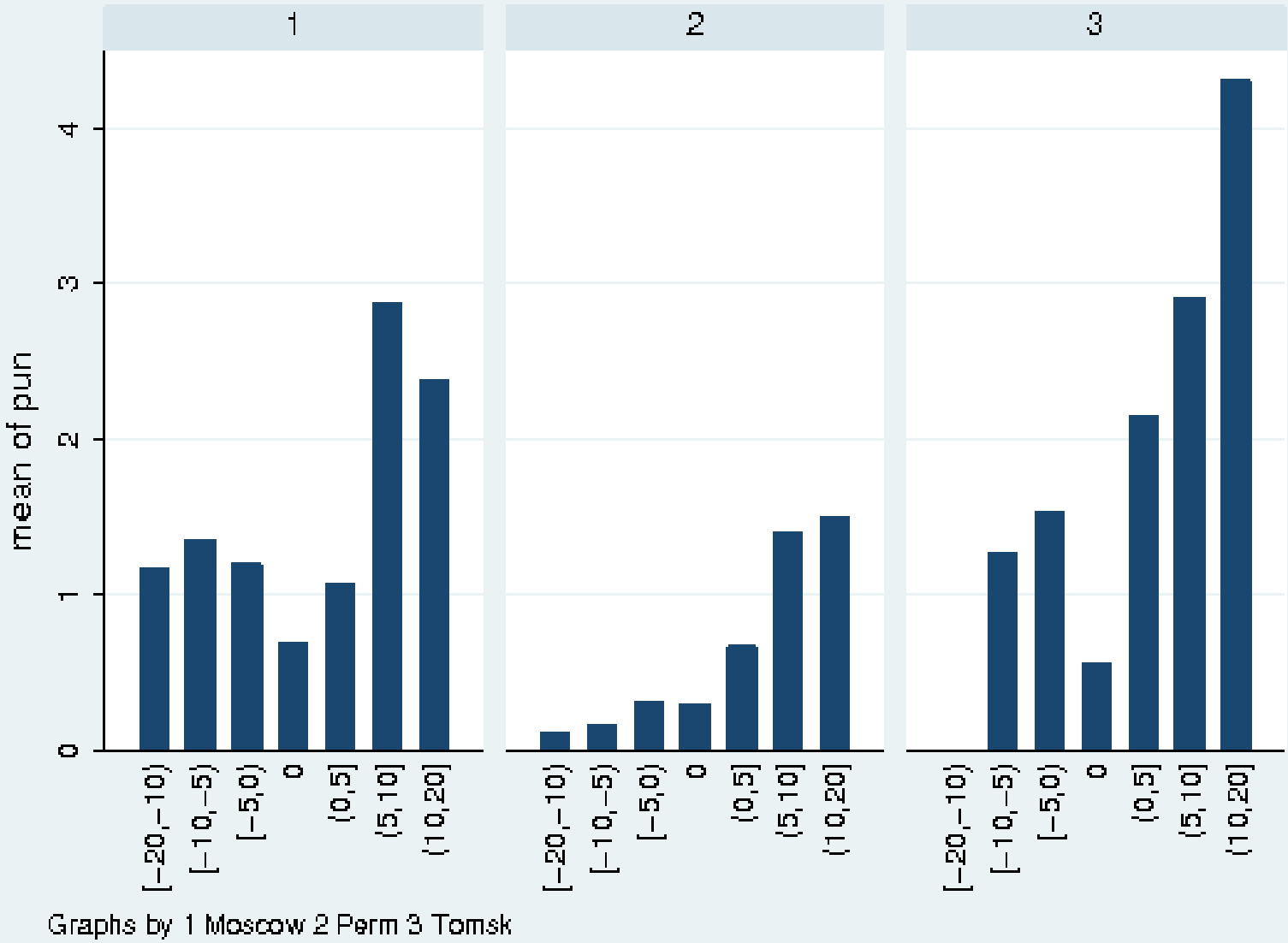
Contributions: distribution



Contributions: trends

- Contributions in Tomsk (5.16) are most dense, and significantly lower than in Moscow (8.66) or Perm (8.94).
- Factual own contributions always lower than projected, especially in Tomsk.
- Normative contributions in Moscow significantly higher than in both Perm and Tomsk.
- Difference between projected normative and expected contributions are lowest in Tomsk — people expect others to be most norm-obedient.
- Difference between normative and own planned contribution is smallest in Perm — people are themselves norm-obedient.

Punishments by differences in contributions



Punishments: statistics

	All	Moscow	Perm	Tomsk
Punish at least once %	56%	60%	45%	60%
# punishments (% pairs)	220 (30%)	65 (29%)	48 (21%)	107 (37%)
Mean punishment size	4.48	4.70	3.06	4.98
Punished 0 players	128 (53%)	40 (53%)	50 (66%)	38 (40%)
Punished 1 player	52 (21%)	17 (23%)	11 (14%)	24 (25%)
Punished 2 players	33 (13%)	6 (8%)	8 (11%)	19 (20%)
Punished 3 players	34 (14%)	12 (16%)	7 (9%)	15 (16%)
# spiteful (% to all pun's)	70 (32%)	54 (23%)	47 (21%)	61 (21%)
$\frac{\text{mean spiteful punishment}}{\text{mean prosocial punishment}}$	0.53	0.61	0.19	0.68

Number of punished players	1	2	3
# of punishment instances	52	66	102
of which spiteful (%)	6 (11%)	18 (27%)	39 (38%)

Punishments: findings

confirmed: Mean frequency and size of spiteful punishments are compatible with those of the previous experiments

Punishments: findings

confirmed: Mean frequency and size of spiteful punishments are compatible with those of the previous experiments

confirmed: Mean punishment size and cost per punishment is the same for prosocial and spiteful punishments (similar rationality)

Punishments: findings

confirmed: Mean frequency and size of spiteful punishments are compatible with those of the previous experiments

confirmed: Mean punishment size and cost per punishment is the same for prosocial and spiteful punishments (similar rationality)

new! Spiteful punishments are typically more serial than prosocial (uniform strategy)

Punishments: findings

confirmed: Mean frequency and size of spiteful punishments are compatible with those of the previous experiments

confirmed: Mean punishment size and cost per punishment is the same for prosocial and spiteful punishments (similar rationality)

new! Spiteful punishments are typically more serial than prosocial (uniform strategy)

new! In the ex post questionnaire, over 80% of spiteful punishers report desire to increase their relative standing as the main motive for punishment (competitive motive)

Punishments: findings

confirmed: Mean frequency and size of spiteful punishments are compatible with those of the previous experiments

confirmed: Mean punishment size and cost per punishment is the same for prosocial and spiteful punishments (similar rationality)

new! Spiteful punishments are typically more serial than prosocial (uniform strategy)

new! In the ex post questionnaire, over 80% of spiteful punishers report desire to increase their relative standing as the main motive for punishment (competitive motive)

new! Factors determining prosocial and spiteful punishments are different (variety of motives)

Punishments factors: Tobit model estimates

	Spiteful		Prosocial		Overall	
Variable	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.
contr	-0.35^{***}	(0.12)			-0.17	(0.12)
dcontr			0.26^{***}	(0.06)	0.51^{***}	(0.15)
rcontr	-0.88^{***}	(0.20)			-0.57^{***}	(0.19)
econtrx	0.20^{*}	(0.12)			-0.02	(0.10)
econtra			-0.11^{**}	(0.05)	0.11	(0.07)
condev	0.28^{**}	(0.11)			0.06	(0.09)
tomsk	3.25^{***}	(1.21)			2.97^{***}	(0.81)
const	4.34^{**}	(1.65)	2.95^{***}	(0.47)	-4.02^{***}	(1.26)

*** — significant at 1%, ** — significant at 5%, * — significant at 10%

contr – c_j of punished

dcontr – $\Delta(c_i - c_j)$

rcontr – $\Delta(c_i - \bar{c})$

econtra – $\Delta(c_i - \mathrm{E}\bar{c})$

econtrx – $\Delta(c_i - \mathrm{E}c_i)$

condev – $\Delta(c_i - \mathrm{E}c_i)$ at group mean

tomsk – dummy for Tomsk, cost 0.1

Punishment factors: interpretations

- Prosocial punishments are caused by **upset**: 1) differences in contributions and 2) over-contribution of the punisher relatively to her normative group standard
- Spiteful punishments are related to 1) low contribution of the punished, 2) low own contributions relatively to group average, 3) large 'unplanned' own contributions, and low cost of punishment (Tomsk), in line with **competitive** explanation.
- None of the explanatory variables for one type of behaviour is significant as explanatory variable for the other
- *Availability* appears to be immaterial: average willingness to punish insignificantly smaller than elsewhere
- *Tolerance* is immaterial: 37% of prosocial and 83% of spiteful punishers have relocated their funds from punishment to insurance at the last stage, suggesting that **preemption** as another reason for 'spite'.

Behavioural model of punishment motives

$$u_i = V_i + \lambda_{1i} \frac{\sum_j \sum_k \gamma_k \varphi_{kij}}{p_{ij}} - \lambda_{2i} \sum_j \frac{\mathrm{E}p_{ji}}{p_{ij}} - \pi \left[\lambda_{1i} \sum_j p_{ji} \left(\sum_k \gamma_k \varphi_{kij} \right) + \lambda_{2i} \sum_j \mathrm{E}p_{ji} \right] \tag{3}$$

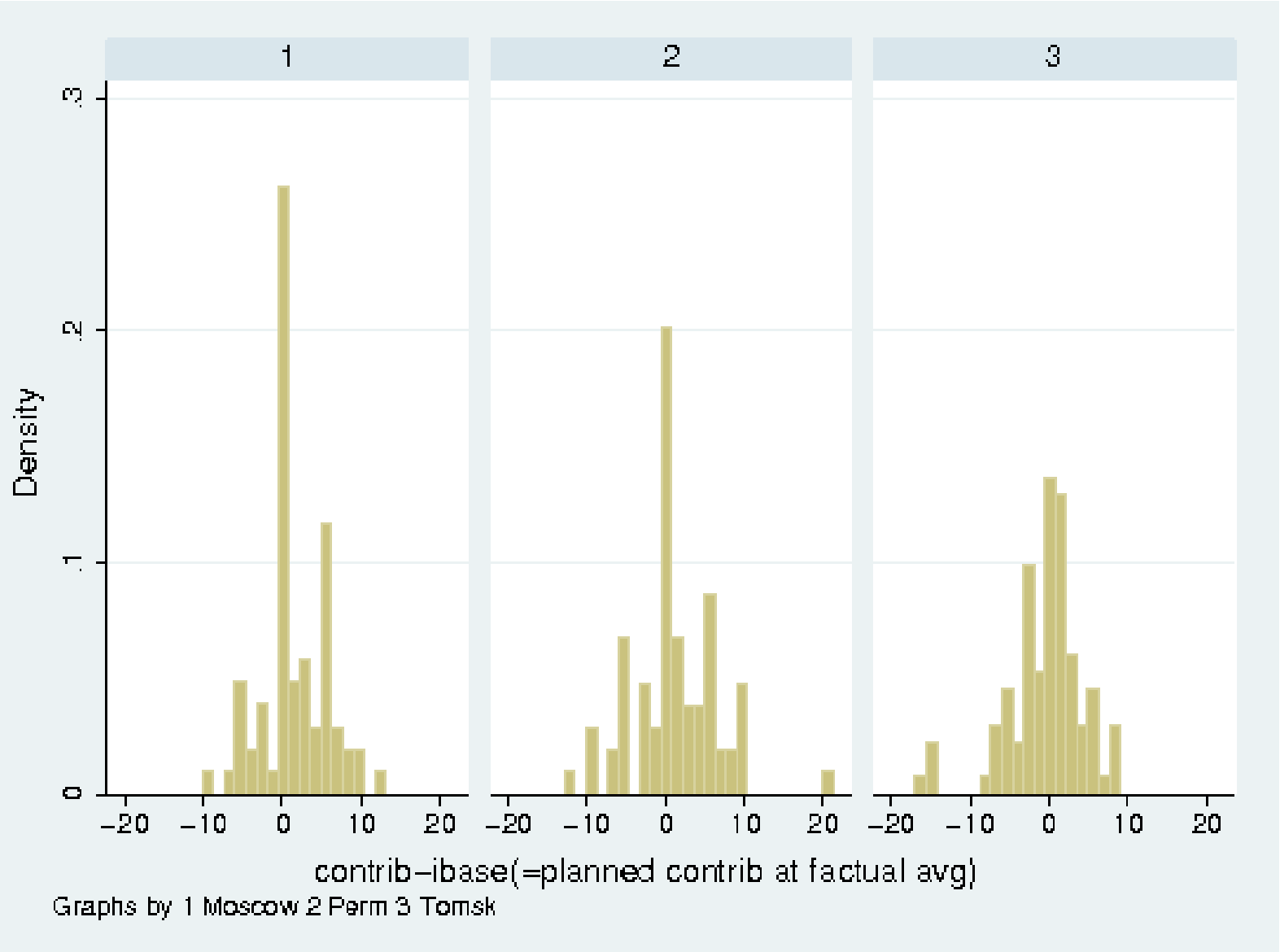
- V_i —material payoff,
- φ — retaliation function of player i at player j,
- $\mathrm{E}p_{ji}$ — expectation of player i of punishment from player j,
- π — cost of punishment,
- λ_{1i} and λ_{2i} — individual-specific weights to retaliation for bad behaviour and preemption for expected punishment

Maximizing (3) wrt punishment p_{ij},

$$p_{ij}^* = \lambda_{1i} \frac{\sum_k \gamma_k \varphi_{kij}}{p_{ij} \pi} + \lambda_{2i} \frac{n-1}{\pi} \tag{4}$$

wherein linear weights λ attached to normal densities of the latent

Factual vs strategic form planned contributions



Model estimates

For prosocial punishment:

$$pun = \alpha + \lambda_1 \phi(prcontr + pcontr) + \lambda_2 \phi(pcons) + \varepsilon \quad (5)$$

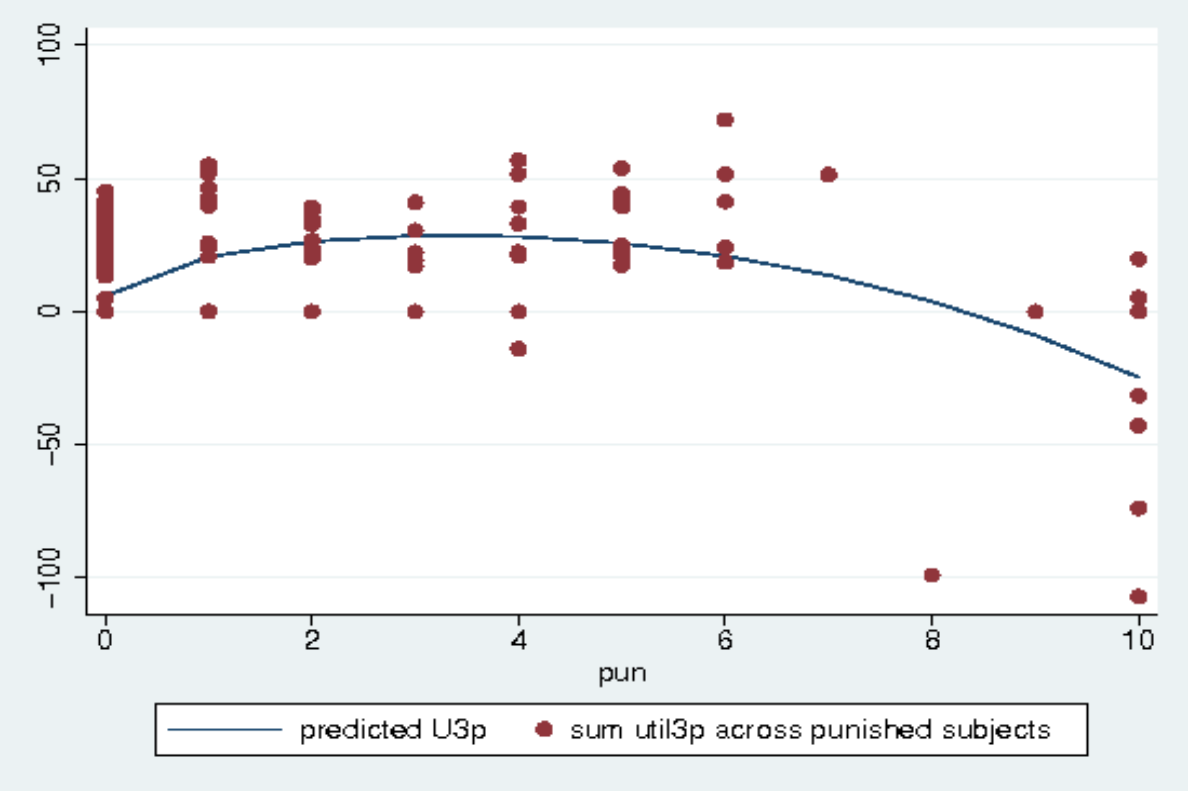
Weights are $\lambda_1^p = 0.207, \lambda_2^p = 0.793$, implying larger proportion of preemptive punishers

For spiteful punishment:

$$pun = \alpha + \lambda_1 \phi(pcondev) + \lambda_2 \phi(pcons) + \varepsilon \quad (6)$$

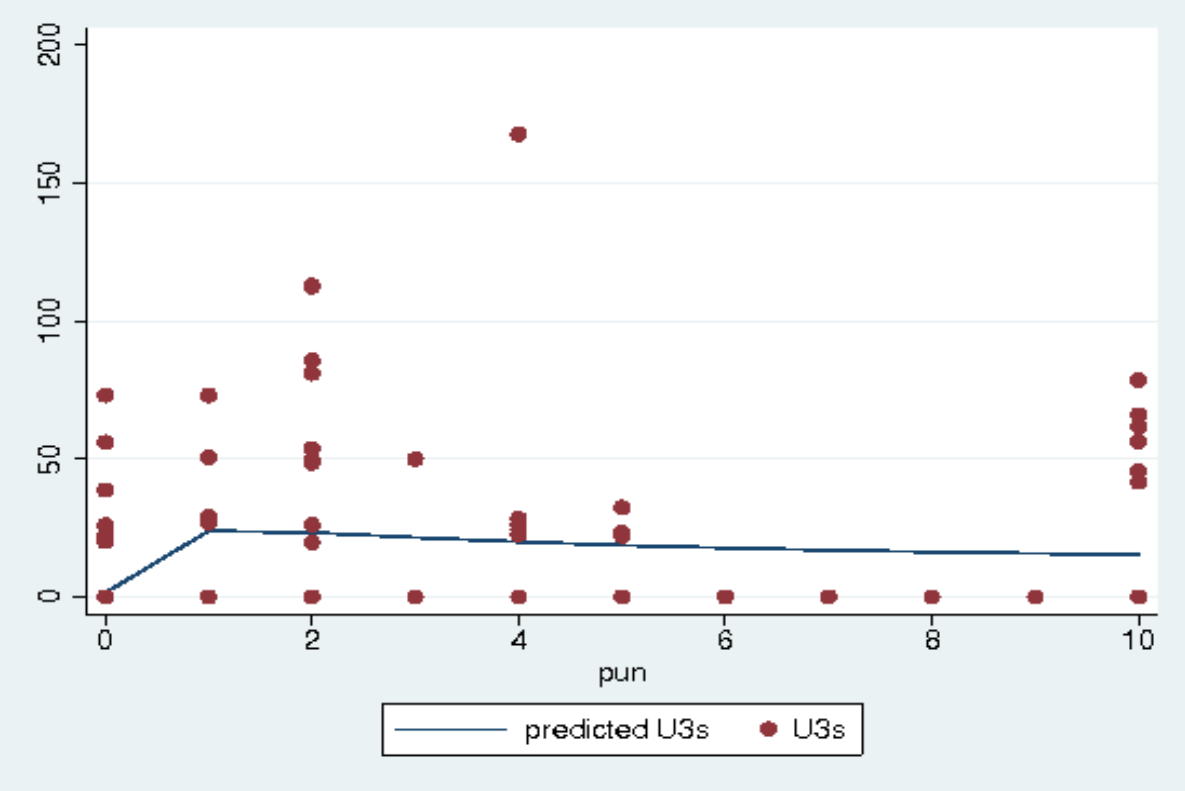
Weights $\lambda_1^p = 0.826, \lambda_2^p = 0.176$, imply larger proportion of retaliators

Estimated utility for prosocial punishers



Inverse U-shape of utility vs. punishment size: at lower levels, larger punishments correspond to low utility of the punisher as they reflect their unhappiness with the social behaviour.

Estimated utility for spiteful punishers



U-shape graph with high dispersion at low punishment levels and large utility for those with extreme punishments.

Classification: the four punishment caterogires

Fair prosocial (15%) Punishments motivated by low contributions of the punished relative to the group standard. Believe they are on their right, punish by a lot (mean 9.78), and almost do not insure (mean 1.28).

Timid prosocial (58%) Fairness motivated, but afraid of expression for fear of preemption and/or cost concerns. Punishment is low (3.51), insurance yet lower (2.5)

Jealous spite (17%) Afraid of being exploited by the society, try to decrease payoffs of more successful players, but not at own cost. Both punishments (2.66) and insurance (2.5) are low.

Active spite (12%) Motivated by competitiveness, but also very afraid of preemption: use maximal punishments (10 in 100% cases) and insurance (7.38%).

The main result so far, to be qualified with more data

Interpretations and extensions

- Punishment in PG context at least, should not always be interpreted as a revelation of dissatisfaction with contributions of the other players: there is a variety of competing explanations.
- These results suggest a multiplicity of principles on which 'punishment' behaviour may rest. In Russia, these were quite heterogeneous, while in Western Europe, for instance, 'spiteful' punishments are minor. Decomposition of punishment motives may be interesting and important for the diagnosis of the state of the respective societies.

Thank you!