Punishment without crime: a tale of cooperation and competition in public goods game

Alexis Belianin
Laboratory of Experimental and Behavioural Economics and ICEF,
Higher School of Economics
icef-research@hse.ru LCSR Research Workshop, April 25, 2012

April 25, 2012

Punishment in public goods game (PG VCM)

- Factors of cooperative behaviour are of interest to economists, sociologists, psychologists... all social scientists.

Punishment in public goods game (PG VCM)

- Factors of cooperative behaviour are of interest to economists, sociologists, psychologists... all social scientists.
- Experimental measures are some way to make cross-country comparisons in identical conditions (e.g. investment game, trust game, ultimatum game, public goods game)

Punishment in public goods game (PG VCM)

- Factors of cooperative behaviour are of interest to economists, sociologists, psychologists... all social scientists.
- Experimental measures are some way to make cross-country comparisons in identical conditions (e.g. investment game, trust game, ultimatum game, public goods game)
- Recent behavioural explanations (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 useful, albeit restrictive.

Punishment in public goods game (PG VCM)

- Factors of cooperative behaviour are of interest to economists, sociologists, psychologists... all social scientists.
- Experimental measures are some way to make cross-country comparisons in identical conditions (e.g. investment game, trust game, ultimatum game, public goods game)
- Recent behavioural explanations (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 useful, albeit restrictive.
- One more of these: estimation of factors of punishment in public goods games using experimental technique and structural model.

In this lecture we

1. Discuss the cross-country evidence of cooperation in public goods games
2. Claim that conventional attribution of punishment to 'dissatisfaction with low contribution' (and by the same token, to disapproval of antisocial behaviour) is too quick/impudent: In the PG game context, people may punish each other for different (strategic) reasons driven by the experimental institution.

In particular, this may explain the divergence between the fractions of spiteful behaviour (punishing those who contributed *more* than you did) obsereved in some (developing) countries to a much more substantial extent that in other (developed).

Contributions:

- New experimental design (insurance against punishment)
- Behavioural model of strategic incentives for punishment
- Empirical estimates of latent classes of motives in a convenience sample of

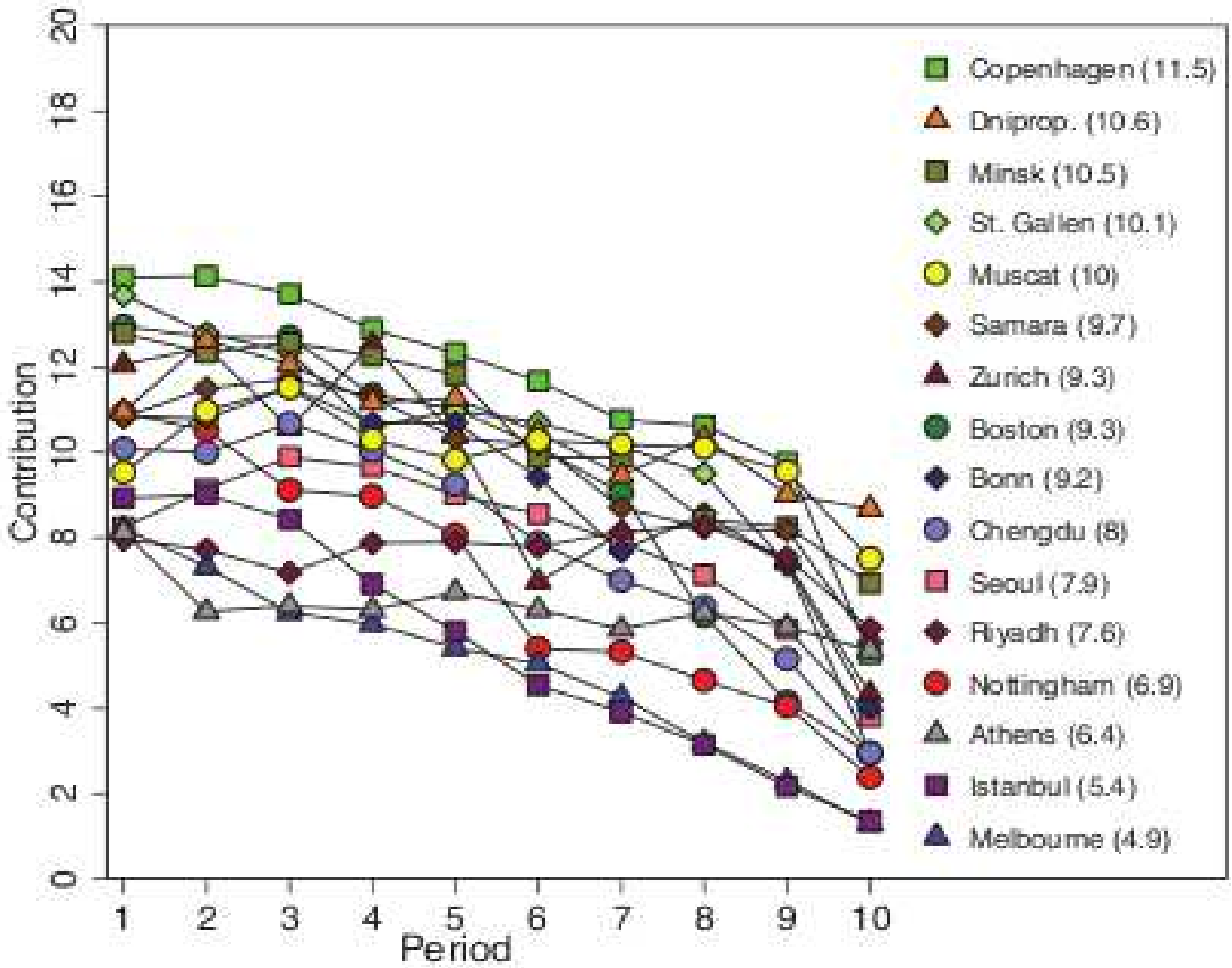
Public goods (PG) game with voluntary contribution mechanism (VCM)

- $n \geq 2$ 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.

$$v_i = 1 - c_i + \alpha \bar{c} = 1 - c_i + k \cdot \sum_i c_i \tag{1}$$

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

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



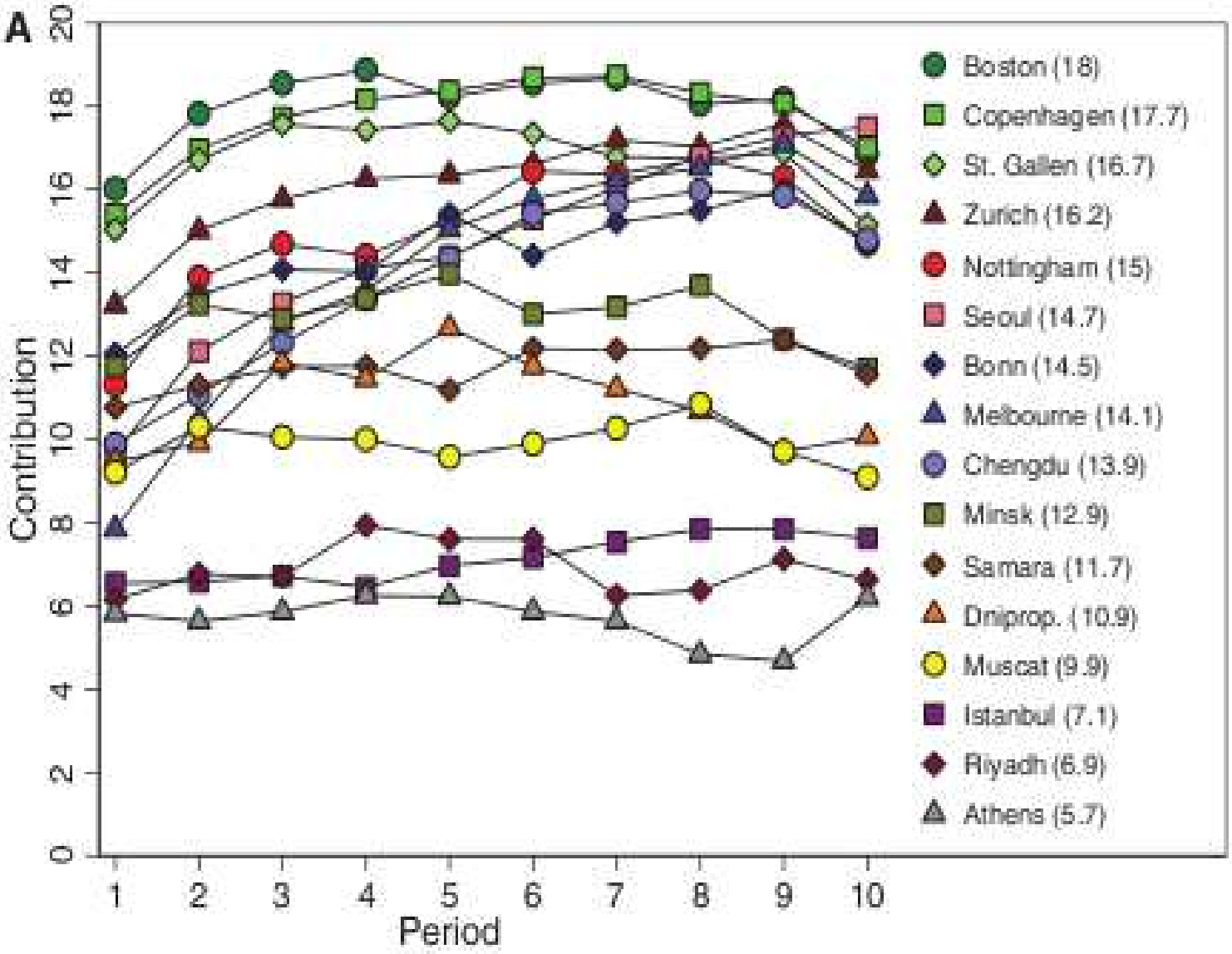
Public goods game with VCM and punishment

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

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

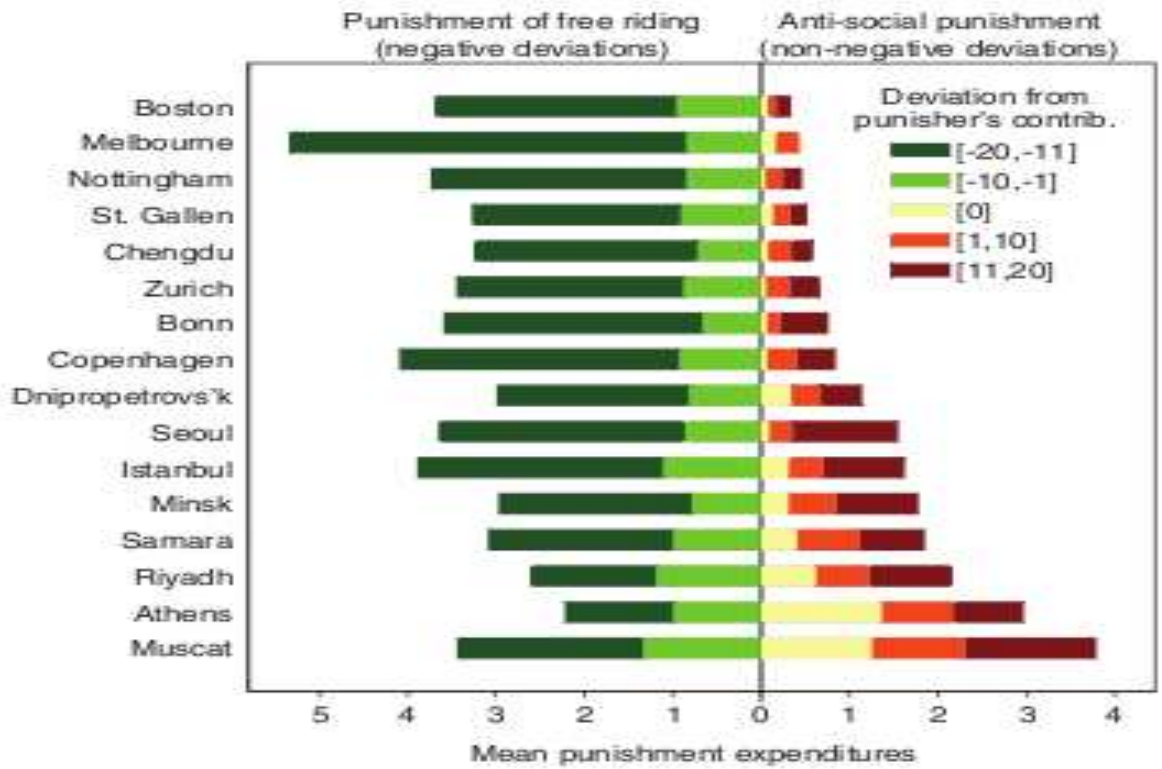
Punishments are known to increase the degree of cooperativeness, especially in with time and in partner treatments.
Mechanism: punishment (threaten, expression of disapproval) of those who free-ride boosts up cooperativeness.

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.

'If in the first scene of the play, there is a gun on the wall, by the third scene it mush shut'

Classification of possible motives for punishment

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

'If in the first scene of the play, there is a gun on the wall, by the third scene it mush shut'

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

'The entire Russian history before Peter the Great is an entire commemoration cervice, and after Peter the Great — an entire criminal case'

Classification of possible motives for punishment

Availability — presense of punishment option is suggestive in itself — the Chekhov motive.
'If in the first scene of the play, there is a gun on the wall, by the third scene it mush shut'

Tolerance — culturally-defined punishment is something 'customary' and 'acceptable' — the Tjutchev motive.
'The entire Russian history before Peter the Great is an entire commemoration cervice, and after Peter the Great — an entire criminal case'

Competitiveness — punishment as an efficient way to improve own relative standing in the group — the Dostoyevsky motive.
'Am I a trembling biest, or I daresay?'

Classification of possible motives for punishment

Availability — presense of punishment option is suggestive in itself — the Chekhov motive.
'If in the first scene of the play, there is a gun on the wall, by the third scene it mush shut'

Tolerance — culturally-defined punishment is something 'customary' and 'acceptable' — the Tjutchev motive.
'The entire Russian history before Peter the Great is an entire commemoration cervice, and after Peter the Great — an entire criminal case'

Competitiveness — punishment as an efficient way to improve own relative standing in the group — the Dostoyevsky motive.
'Am I a trembling biest, or I daresay?'

Preemption — penalizing because one expects penalties from the others — the Brodsky motive
'A man is more frightening than its skeleton'.

Classification of possible motives for punishment

Availability — presense of punishment option is suggestive in itself — the Chekhov motive.
'If in the first scene of the play, there is a gun on the wall, by the third scene it mush shut'

Tolerance — culturally-defined punishment is something 'customary' and 'acceptable' — the Tjutchev motive.
'The entire Russian history before Peter the Great is an entire commemoration cervice, and after Peter the Great — an entire criminal case'

Competitiveness — punishment as an efficient way to improve own relative standing in the group — the Dostoyevsky motive.
'Am I a trembling biest, or I daresay?'

Preemption — penalizing because one expects penalties from the others — the Brodsky motive
'A man is more frightening than its skeleton'.

Retaliation — 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.
'If in the first scene of the play, there is a gun on the wall, by the third scene it mush shut'

Tolerance — culturally-defined punishment is something 'customary' and 'acceptable' — the Tjutchev motive.
'The entire Russian history before Peter the Great is an entire commemoration cervice, and after Peter the Great — an entire criminal case'

Competitiveness — punishment as an efficient way to improve own relative standing in the group — the Dostoyevsky motive.
'Am I a trembling biest, or I daresay?'

Preemption — penalizing because one expects penalties from the others — the Brodsky motive
'A man is more frightening than its skeleton'.

Retaliation — 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.

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.

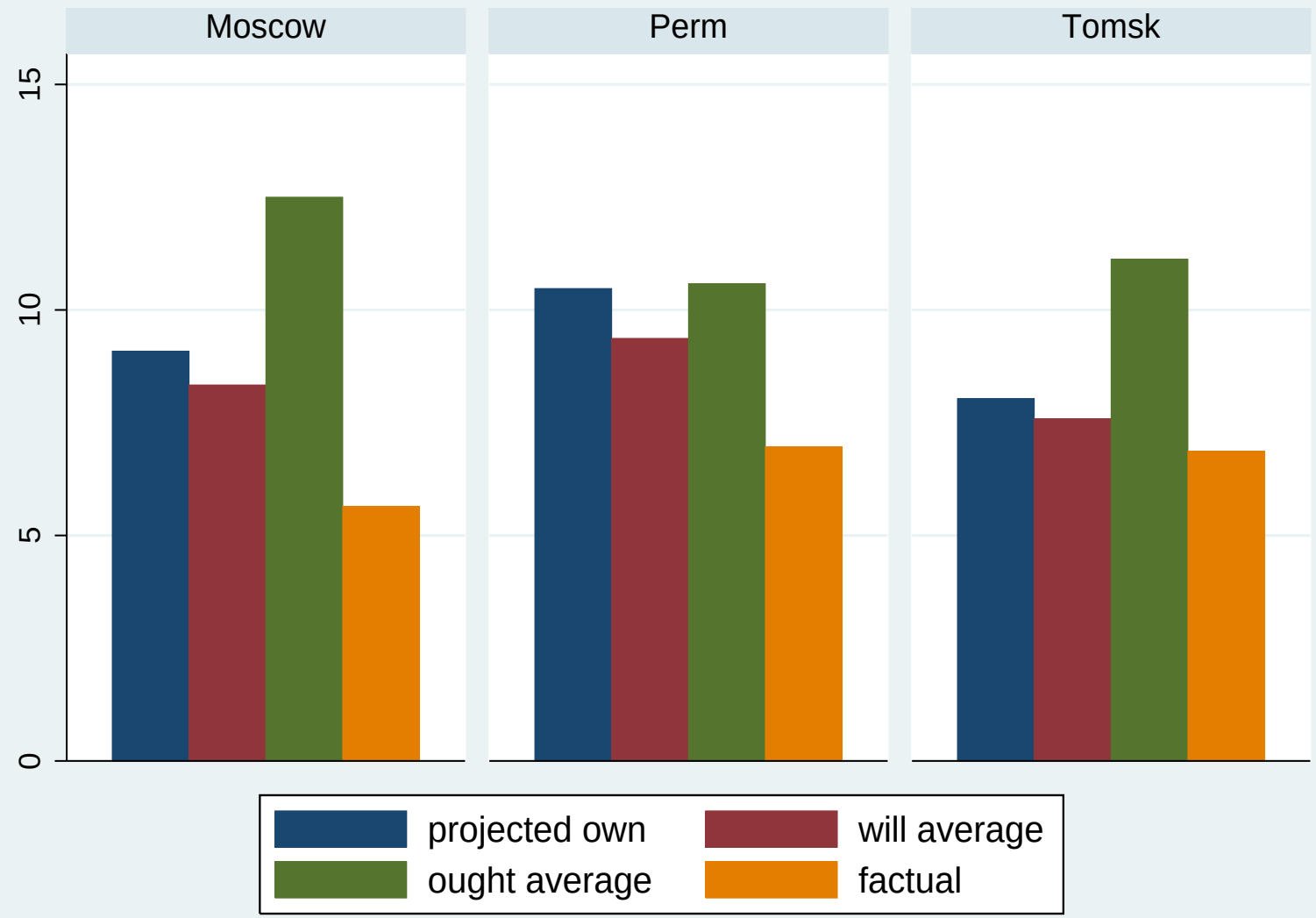
Participants: 300 full-time and part-time students from Moscow (128), Perm (76) and Tomsk (96). Gender composition — 50/50, average payoff — 208 RuR.

Experiment on the map

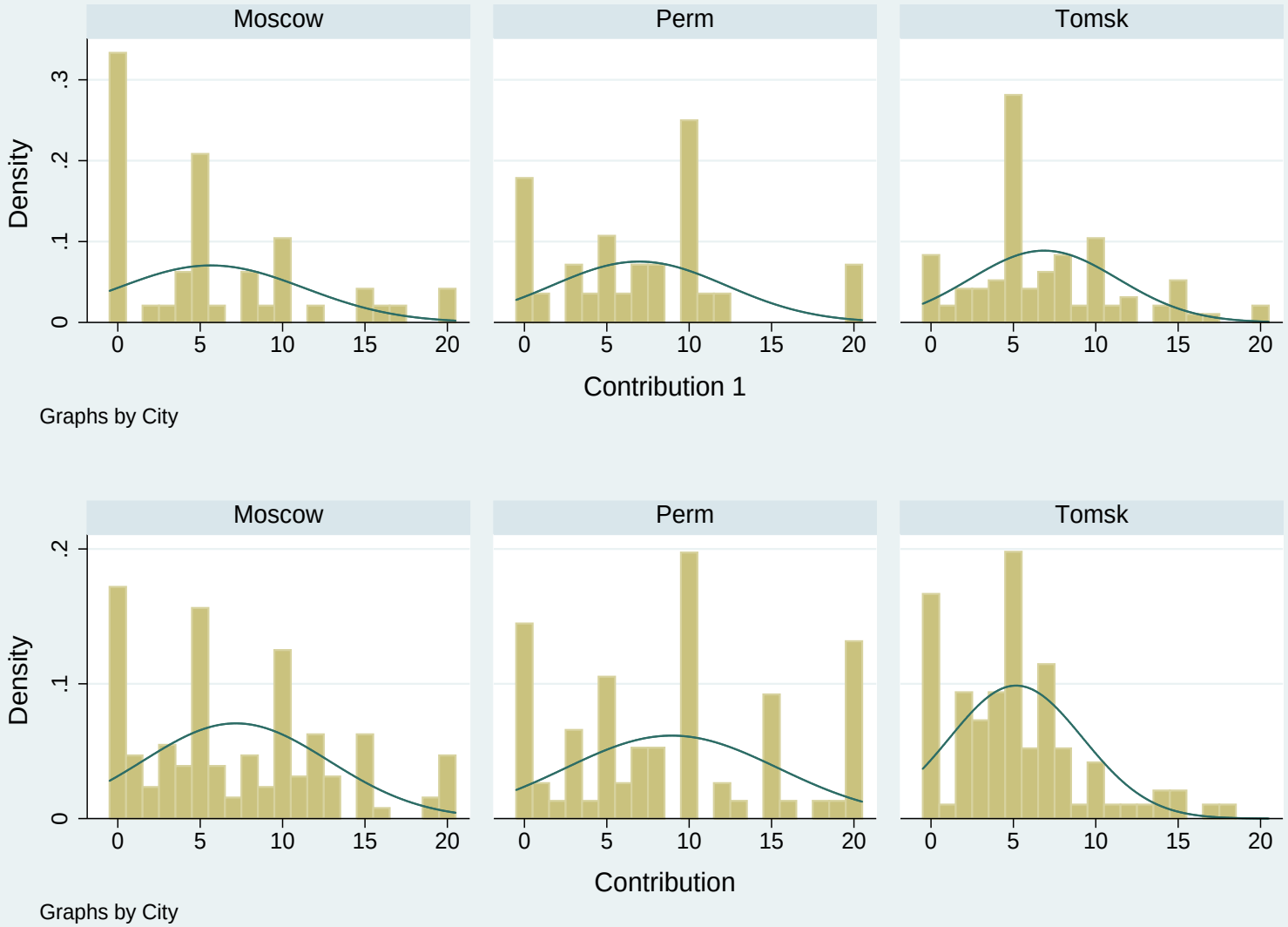
Design: additions

- 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.

Contributions



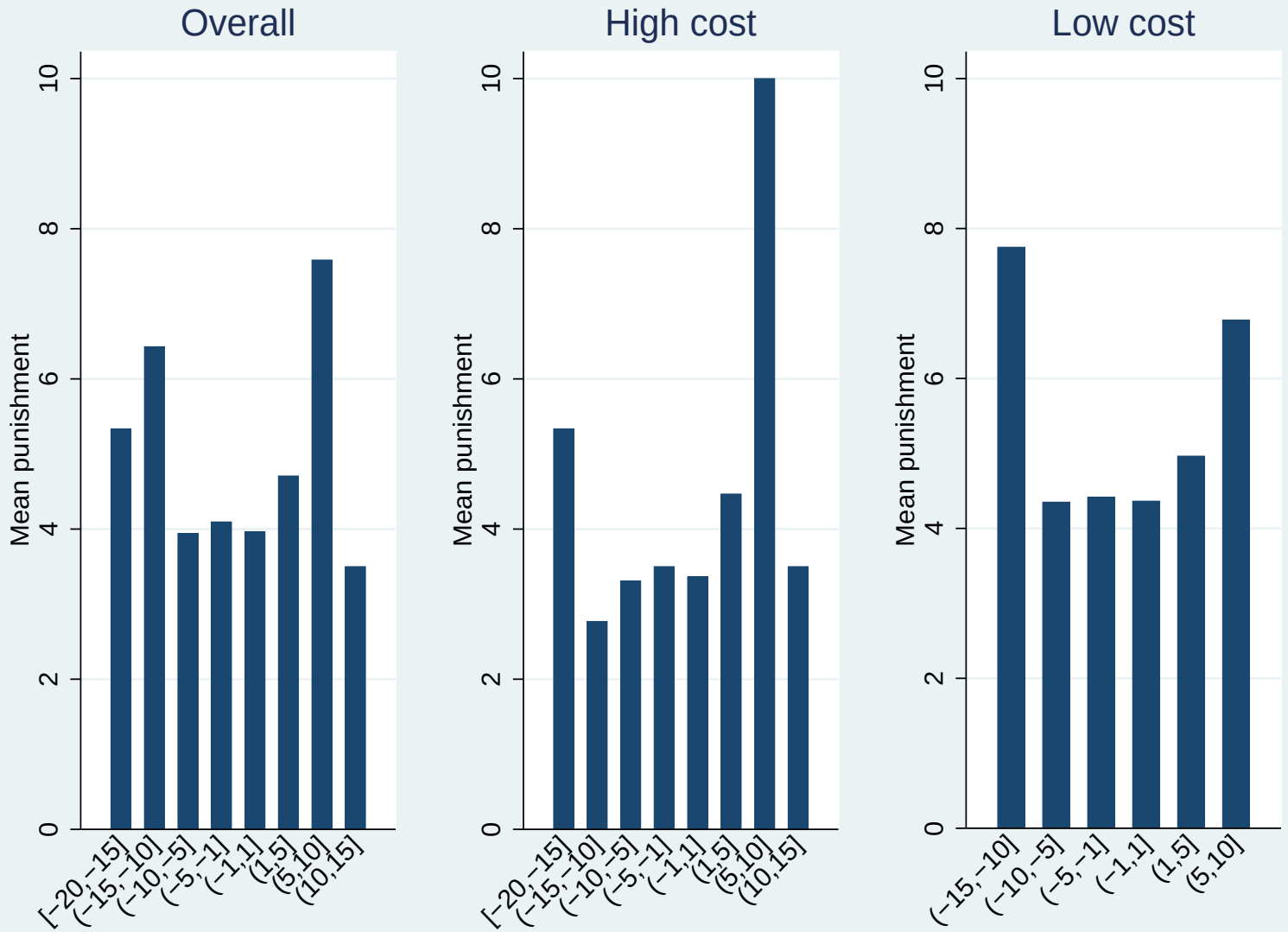
Contributions: first (upper) and second (lower) stage



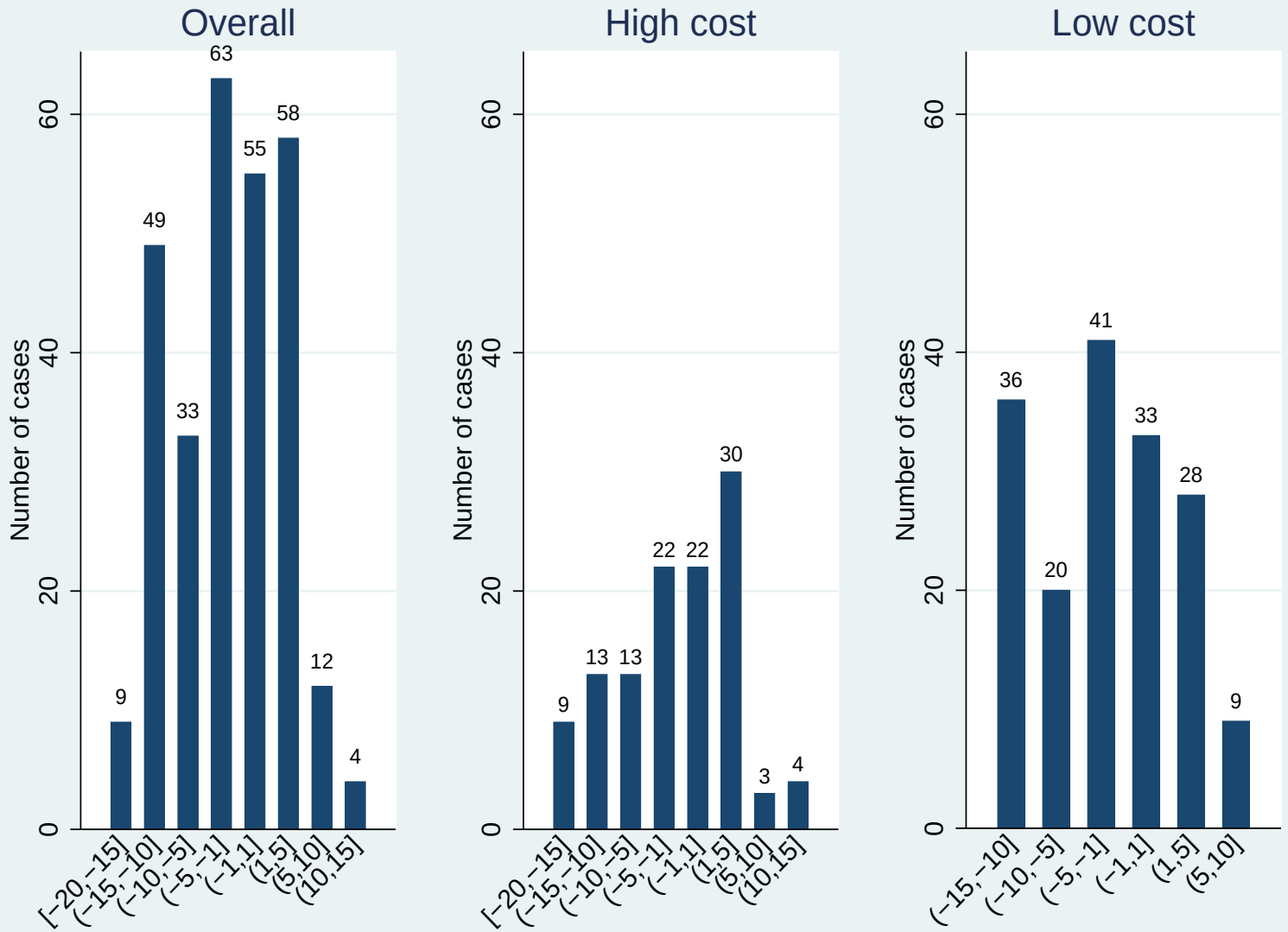
Contributions

- Contributions in line with the previous experience.
- Factual own contributions always lower than projected and (especially) normative.
- Expected undercontribution.
- In one-round span, disciplining role of punishment is limited at best.
- Second-stage contributions are stable across cities at low (median 5) and high (median 9) costs.

Mean punishments by treatments



Number of punishments by treatments



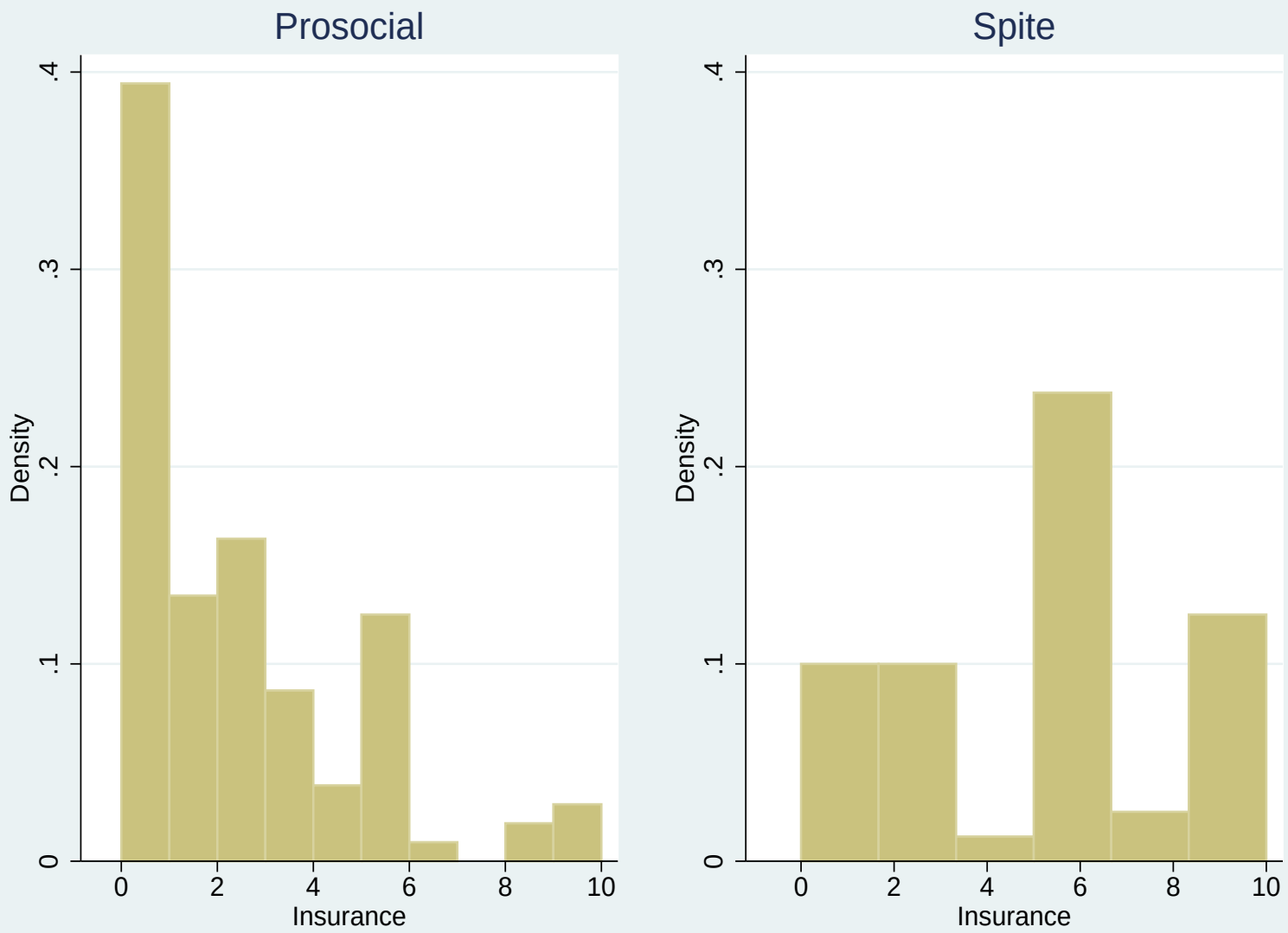
Punishments: statistics

statistics	*contrib*	*wish*	*qpun*	*pun*
Overall (N=295)				
mean	6.99	.55	.95	4.75
median	5	1	0	4
sd	5.60	.49	1.14	3.23
Low cost = 0.1 (N=143)				
mean	5.09	.59	1.16	5.33
median	5	1	1	5
sd	4.17	.49	1.17	3.34
High cost = 0.5 (N=152)				
mean	8.78	.51	.76	3.93
median	9	1	0	3
sd	6.16	.50	1.07	2.89

Punishments trends by treatments

stats	spun	apun	devcap	cavg	difavg	insp	insour	assign
Spiteful, 1 punishment (6 players)								
mean	2.5	2.5	-3.83	-3	0.75	2	.333	.167
median	2.5	2.5	-3.83	-3.5	1	3	0	0
Spiteful, 2 punishments (6 players)								
mean	9.5	4.75	-4.77	2.16	1.20	3.17	.167	.5
median	8.5	4.25	-4.91	-0.5	0.63	2	0	.5
Spiteful, 3 punishments (13 players)								
mean	20.5	6.82	-4.03	-0.72	1.53	5.74	.148	.385
median	30	10	-3.33	0	1	5	0	0
Prosocial, 1 punishment (50 players)								
mean	4.2	4.2	-1.16	-8.86	-4.69	1.92	.4	.58
median	3.5	3.5	-1.16	-8.5	-4.5	1	0	1
Prosocial, 2 punishments (27 players)								
mean	8.56	4.28	0.31	-9.98	-3.28	2.17	.583	.5
median	7	3.5	0	-9.5	-3.25	2	1	.5
Prosocial, 3 punishments (20 players)								
mean	14.3	4.77	-0.85	-8.31	-1.67	1.28	.41	.567
median	12	4	-0.66	-8	-0.87	1	0	1

Insurance decisions



Motives, % in ex-post questionnaire

Reasons		
Variable	Prosocial (N=121)	Spiteful (N=53)
Lower (than average) contribution	47.1	20.8
To stop them lowering our revenues	13.2	7.5
To gain more than they will	12.4	43.4
Afraid of them reducing my revenue	11.8	9.4
To equalize revenue within group	9.1	15.1
Intuitively/to experiment	7.5	1.9

Size determinants		
Variable	Prosocial (N=121)	Spiteful (N=50)
Inverse to their contribution	29.0	6.0
Maximal to the smallest contributor	18.5	8.0
To average out revenue	15.5	16.0
To put all revenues down to mine	11.5	–
Intuitively	8.7	14.0
Depending on my costs	6.8	–
Maximal to all	2.9	38.0

Preliminary conclusions

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

Preliminary conclusions

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

confirmed: Mean punishment size decreases with cost, and is on average the same for prosocial and spiteful punishments (similar rationality)

Preliminary conclusions

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

confirmed: Mean punishment size decreases with cost, and is on average the same for prosocial and spiteful punishments (similar rationality)

new! Spite increases in low-cost conditions

Preliminary conclusions

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

confirmed: Mean punishment size decreases with cost, and is on average the same for prosocial and spiteful punishments (similar rationality)

new! Spite increases in low-cost conditions

new! Spiteful punishments are more serial and larger on average than prosocial punishments

Preliminary conclusions

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

confirmed: Mean punishment size decreases with cost, and is on average the same for prosocial and spiteful punishments (similar rationality)

new! Spite increases in low-cost conditions

new! Spiteful punishments are more serial and larger on average than prosocial punishments

new! Spiteful punishers insure significantly more often and use more extra money than prosocial punishers

Preliminary conclusions

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

confirmed: Mean punishment size decreases with cost, and is on average the same for prosocial and spiteful punishments (similar rationality)

new! Spite increases in low-cost conditions

new! Spiteful punishments are more serial and larger on average than prosocial punishments

new! Spiteful punishers insure significantly more often and use more extra money than prosocial punishers

new! In the ex post questionnaire, over 3/4 of spiteful punishers report desire to increase their relative standing as the main motive for punishment

Punishments factors: Tobit model estimates

	Spiteful		Prosocial		Total	
Variable	**Coef.**	**Std.Err.**	**Coef.**	**Std.Err.**	**Coef.**	**Std**
contr			−0.409***	(0.103)	-0.658	(0.
difcontr	−0.865***	(0.224)	1.312***	(0.122)	0.695***	(0.
relcontr	−1.583*	(0.947)			−0.451**	(0.
homxavg			0.175*	(0.112)	0.029	(0.
cost	−22.17***	(6.263)	−6.290***	(1.575)	−8.753***	(1.
Intercept	−20.025**	(4.859)	−5.216***	(0.606)	−4.259***	(0.
Log pseudolik.	-368.55		-739.23		-1167.29	
N	958		1060		1148	

Cluster-robust standard errors in parentheses. *** –1%, ** –5%, * – 10% sign.level

contr – c_j, contribution of punisher, difcontr – $c_i - c_j$, relcontr – $c_i - \mathrm{E}c_j$, homxavg – $\mathrm{E}c_i - \mathrm{E}\bar{c}_j$, cost – cost treatment dummy

Punishment factors revisited

- *Availability* appears to be immaterial: average willingness to punish insignificantly smaller than elsewhere.
- *Tolerance* is immaterial: most punishers insure, 51% of prosocial and 75% of spiteful punishers have relocated their funds from punishment to insurance.
- Prosocial punishments driven by **retaliation**: differences in contributions are the major explanatory factor.
- Spiteful punishments driven **competition**: willingness to beat the others prevails.
- Separate factor of **preemption** (or being afraid of self-expression) may apply to both.

How can we disentangle competitive/retaliation and preemption motives for prosocial and spiteful punishments, respectively?

Behavioural model of punishment motives

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

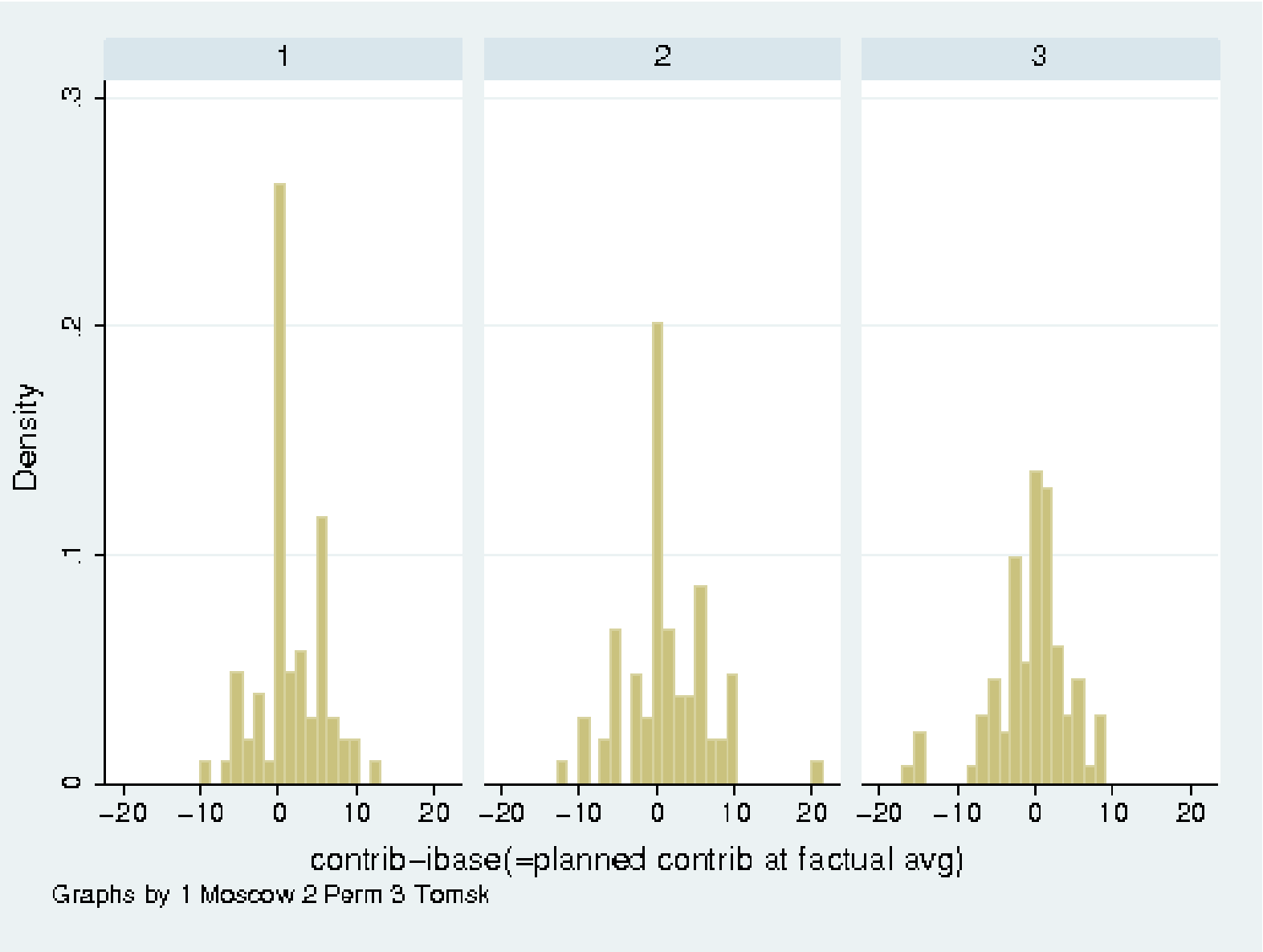
- V_i —material payoff,
- φ — dissatisfaction function of player i at player j,
- $\mathrm{E}p_{ji}$ — expectation of player i of punishment from player j,
- π — cost of punishment,
- η_{1i} and η_{2i} — individual-specific weights to retaliation and preemption for expected punishment (η's are zero in case of no punishment)

Maximizing (3) wrt punishment p_{ij} and rearranging,

$$p_{ij}^* = \eta_{1i} \frac{\sum_k \gamma_k \varphi_{kij}}{p_{ij} \pi} + \eta_{2i} \frac{n-1}{\pi} \quad (4)$$

wherein linear weights η attached to normal densities of the latent factors are estimable using GLLAMM

Factual vs strategic form planned contributions



Model estimates

For prosocial punishment:

$$pun = \alpha + \eta_1 \phi(prcontr + pcontr) + \eta_2 \phi(pcons) + \varepsilon \quad (5)$$

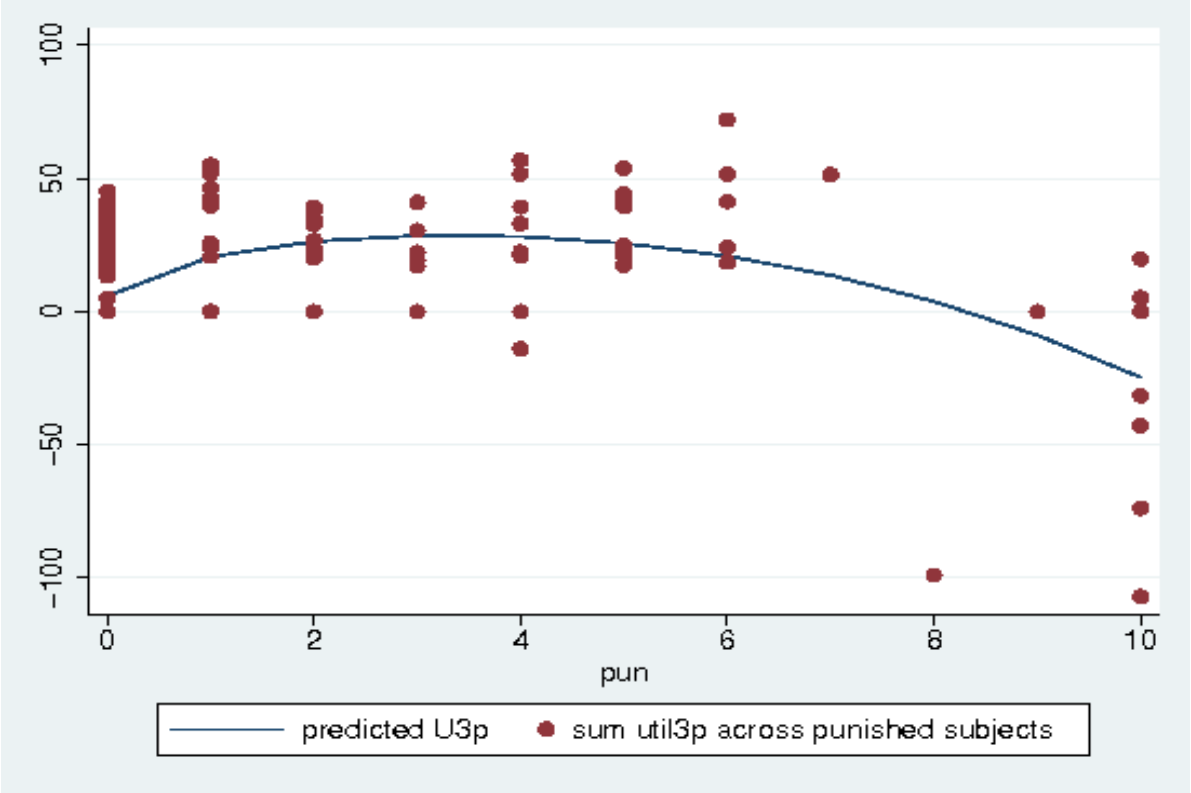
Weights are $\eta_1^p = 0.26, \eta_2^p = 0.73$, implying larger weight on preemption

For spiteful punishment:

$$pun = \alpha + \eta_1 \phi(pcondev) + \eta_2 \phi(pcons) + \varepsilon \quad (6)$$

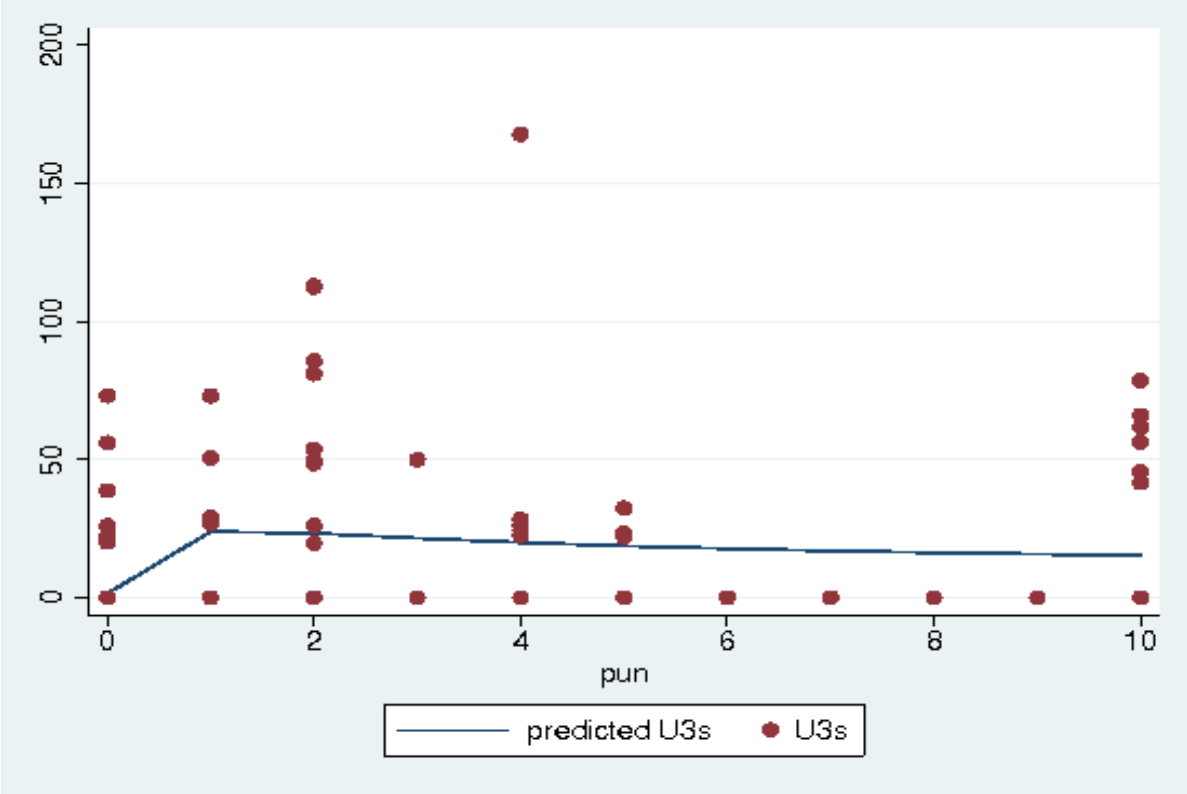
Weights $\eta_1^p = 0.16, \eta_2^p = 0.82$, imply larger weight on preemption

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.

Cluster statistics

Table:

stats	contrib	cexpx	homexp	pun	insp	from
Retaliating prosocial — 12% (M 13%, P 4%, T 13%)						
mean	10.16	-10.51	4.61	9.54	1.9	.65
p50	10	-9	4	10	0	1
sd	4.04	5.868	5.37	1.09	3.59	.48
Preemptive prosocial — 59% (M 56%, P 67%, T 58%)						
mean	8.85	-7.02	3.00	3.43	2.03	.42
p50	8	-6	2	3	2	0
sd	4.96	4.85	4.72	1.94	2.09	.49
Competitive spite — 11% (M 15%, P 0%, T 12%)						
mean	1.37	1.82	2.06	9.65	6.5	.23
p50	1	1	2	10	5	0
sd	2.029	6.25	5.16	1.284	2.74	.42
Preemptive spite — 18% (M, T 16%, P 30%)						
mean	4.69	2.85	1.85	2.65	2.25	.3
p50	5	2	0	2	2.5	0

Classification: the four punishment caterogires

Retaliate prosocial (12%) Punishments motivated by low contributions of the punished relative to the group standard (retaliation). Believe they are on their right, punish by a lot (mean 9.54), and almost do not insure (mean 1.9), skeptical (ought - will contribute is max), redistribute from punishment to insurance (peaceful!).

Preemptive prosocial (60%) Fairness motivated, but afraid of expression for fear of preemption and/or cost. Punishment is low (3.43), insurance yet lower (2.03)

Competitive spite (11%) Motivated by competitiveness, use maximal punishments (9.65 of 10) and insurance (6.5).

Preemptive spite (18%) Undercontribute and know it (contribute - promise max), but afraid of self-expression in both punishments (1.85) and insurance (2.25).

Conclusions 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.
- Most important reasons for Russia are preemptive motives (together, over 3/4), followed by competitiveness (18%) and retaliation for undercontribution (12%)
- Cross-city and cross-country variety is interesting: In Russia, spiteful punishments are large, while in Western Europe, they are minor. However, if we exclude strategic punishments from apparently spiteful ones in Russia, its 'spitefulness' would substantially shrink.
- Decomposition of punishment motives may be interesting and important for the diagnosis of the state of the respective societies.

Thank you!

PS: Full version of the paper available at http://epee.hse.ru