Coordination Through Punishment
– An individual Level Perspective –

Felix Albrecht[†], Sebastian Kube[‡,♯]

Preliminary Draft

June 4, 2017

Abstract

JEL-Classification: C9; D03

Keywords: strategy method, conditional contributions, conditional punishment, type classification, public good game

† University of Marburg; ‡ University of Bonn; ♯ Max Planck Institute for Research on Collective Goods;

* Hertie School of Governance

1 Introduction

Coordination failures can be detrimental to economic outcomes for individuals and communities alike (e.g., Huyck, Battalio, and Beil 1990; Brandts and Cooper 2006). Le Lec, Rydval, and Matthey 2015 showed that costly peer-punishment, which repeatedly has been shown to enhance cooperation in in social dilemma situations (e.g., Fehr and Gächter 2002), can alleviate coordination failure in a weakest link environment. Albrecht, Kube, and Traxler 2016 showed that individuals display heterogeneous peer-punishment patterns and that subjects targeting low contributors are crucial to group success in a linear public good game. **Albrecht2017a** further showed that the majority of subjects conditioned peer-punishment on their own contribution level, not punishing larger but severely sanctioning lower contributions.

Given that in weakest-link coordination games only the lowest contribution is payoff relevant this work addresses two issues: 1) do individuals condition peer-punishment on their on contribution or do they restrict peer-punishment solely to the least contributing subject, as she alone determines the size of the joint group project, and 2) do pro-socially punishing subjects affect coordination success in a repeated game similarly as they foster cooperation in public good games.

2 Design and Procedures

The core of our experiment is a 10 period repeated *weakest-link game* (WL) , played in stable groups of 4 subjects, implementing peer-punishment on its second stage. Subjects have to decide how many of their 20 tokens of private endowment to allocate to a *group project*, and how many points d_{ij} to deduct from their peers (maximum 10 each). The size of the group project is determined by the smallest individual contribution in a group. $\pi_i = 20 - g_i + 1.6 \times \min_{i,j,k,l}(g_j) - 1\sum_{j \neq i} d_{ij} - 3\sum_{j \neq i} d_{ji}$, where d_{ij} is the *assigned* and d_{ji} the *received* punishment.

To elicit individual peer-punishment patterns, we implement punishment in the first period as strategy-methods as used by Kube and Traxler 2011, and Albrecht, Kube, and Traxler 2016. Subjects are shown 11 sets of 3 contributions. 10 sets are *hypothetical*, randomly selected from a predefined set of contribution combinations. One is *real*, contributed by the other group members. The exogenous variation of contributions in the strategy method allows for elicitation of individual punishment patterns.

We evaluate data for 228 subjects collected over 10 sessions in the *BonnEconLab* at the University of Bonn, Germany. For every subject we observe 30 punishment decisions with exogenous contribution variation in the strategy method of the WL-game (henceforth *WLS*) and 10 independent group level mean contribution and punishment decisions over the 10 periods of WL.[1]

The experiment was conducted using the experimental software *ztree* (Fischbacher 2007). Experimental subjects were recruited from the BonnEconLab's subject pool using *Hroot* (Bock, Baetge, and Nicklisch 2014). Including a follow-up questionnaire a session lasted approximately 140 minutes with subjects earning on average 12.95 Euros including show-up fee and payoffs of two game played thereafter.

3 Punishment Type Distributions

We follow Albrecht, Kube, and Traxler 2016 in the classification of punishment types by classifying subjects' punishment behavior with respect to their sanctioning behavior towards others' deviation from full contribution (effort).[2] For each of the 228 individuals we estimate model 1 for the 30 punishment observations obtained from the hypothetical contributions in the RPS(WPS)-game.

$$d_{ij} = \alpha_i + \beta_i(20 - g_j) + \varepsilon_i. \tag{1}$$

Like in Albrecht, Kube, and Traxler 2016 subjects are classified into three behavioral categories, i.e., 'Non-Punisher','Pro-social Punisher', and 'Anti-social Punisher'. A fourth category 'Non-classifiable' captures subjects whose punishment pattern is not captured by the definition of the previous three classes. The three behavioral categories are defined in the following ways:

1. A subject is classified as a 'Non-Punisher' (*NPun*) if she assigns zero punishment points in all of the 30 punishment decisions, i.e., $d_{ij} = 0$ for all g_j. In equation (1), this is depicted by $\hat{\alpha}_i = \hat{\beta}_i = 0$.

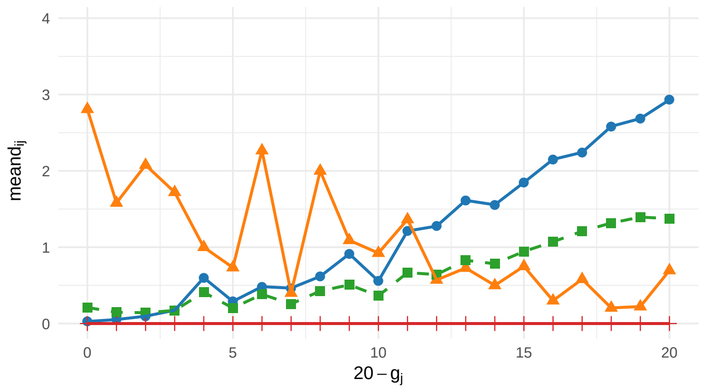
2. Subjects that target their punishment towards those that contribute little or nothing to the public good have a punishment pattern that is upward sloping in $(20 - g_i)$. These subjects, with $\hat{\beta}_i > 0$ and $p \leq 0.01$, are classified as pro-social punishers (*Pun*).

3. Subjects are classified as anti-social punishers (*APun*), if their punishment is either increasing in the other's contribution g_j, i.e., if $\hat{\beta}_i < 0$ and $p \leq 0.01$, or if they display a significant

1. Subjects played three VCM games with and without punishment in the same sessions. As games used very similar payoff functions we took great care to ensure that subjects thoroughly understood the treatment differences by varying wording in the descriptions, the program, and meticulously testing subjects' comprehension of the different tasks.

2. Section C in the supporting online material discusses why we employ others' contributions g_j rather than others' payoff pre-punishment to estimate punishment behavior.

Figure 1: RPS-game Punishment Types



Type	N	%
Pun	111	48.7
NPun	88	38.6
APun	13	5.7
NCL	16	7.0
Total	228	100.0

Notes: Punishment type distribution and average punishment patterns (in the $20 - g_j$-space) in the RPS-game for the different types: pro-social punishers (*Pun*), non-punishers (*NPun*), anti-social punishers (*APun*), and non-classified punishment profiles (*NCL*). To ease illustration, the pattern for the latter is not plotted.

positive but unsystematic level of punishment: $\hat{\alpha}_i > 0$ with $p \leq 0.01$ and an insignificant slope coefficient $\hat{\beta}_i$ with $p > 0.01$.[3]

3.1 Public Good Game Punishment Types

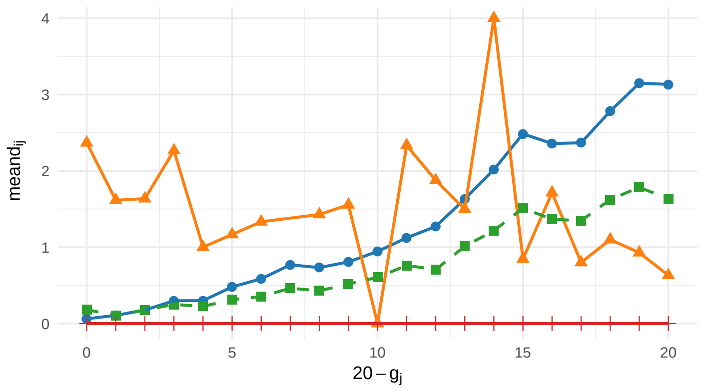
We begin by replicating Albrecht, Kube, and Traxler 2016, classifying individual level punishment patterns in a linear public good game (RPS) with punishment strategy method. Figure 1 presents the distribution of punishment patterns for our 228 experimental subjects. 48.7% of our subjects show pro-social punishment patterns, punishing low contributors more heavily than high contributors. 38.6% of subjects do not invest in peer-punishment in any of the 30 decision situations. 5.7% punish anti-socially deducting more points from high contributors than low contributors. In 7% of the cases subjects' behavior didn't meet one of the three classifications and remained unclassified.

3.2 Weakest Link Game Punishment Types

In a next step we classify individual level sanctioning behavior in a weakest link game setting. As documented by Le Lec, Rydval, and Matthey 2015 subjects engage in costly peer-sanctions in coordination game environments as a means to foster coordination. In line with their findings,

3. The literature typically defines anti-social punishment in reference to a subject's own contribution, i.e., if the punishment-receiving subject contributed a larger or equal amount to the public good compared to the punishing individual (e.g., Herrmann, Thöni, and Gächter 2008). Since our classification does not consider a punisher's own contribution g_i, it deviates from this self-centered notion of anti-social punishment. It nevertheless captures patterns of punishment that is targeted towards high contributors.

Figure 2: WPS-game Punishment Types



Type	N	%
Pun	121	53.1
NPun	69	30.3
APun	7	3.1
NCL	31	13.6
Total	228	100.0

Notes: Punishment type distribution and average punishment patterns (in the $20 - g_j$-space) in the RPS-game for the different types: pro-social punishers (*Pun*), non-punishers (*NPun*), anti-social punishers (*APun*), and non-classified punishment profiles (*NCL*). To ease illustration, the pattern for the latter is not plotted.

we too observe considerable sanctions in the weakest link game setting (WPS). Figure 2 shows the distribution of types and their respective average sanctioning behavior in the WPS-game. We observe an increase in pro-socially sanctioning *Pun*-types compared to the public goods setting. 53.1% of subjects sanction peers displaying low effort levels more strongly than if they display larger efforts. We further observe a reduction in non-sanctioning *NPun* (30.3%) and anti-socially sanctioning *APun* (3.1%) individuals. Finally we observe an increase in non-classifiable individuals (13.6%).

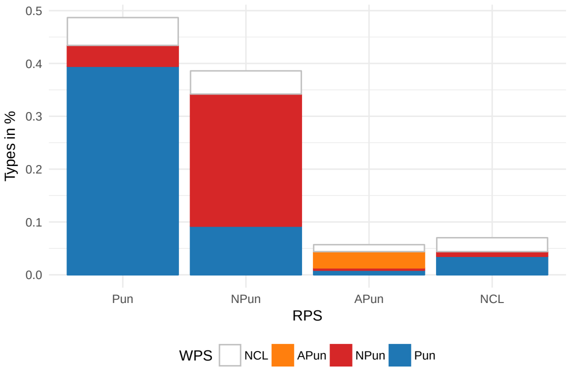
3.3 Individual Cross-Domain Punishment Behavior

Combining the two punishment classifications across the two domains allows us to elicit the *individual punishment type stability*. Figure 3 shows the results. The majority (67.7%) of subjects show a consistent punishment type across the two domains. This includes Pun, NPun, and APun types. Among the switchers the majority of subjects appears to be changing their punishment behavior to be more pro-social. 21 subjects that exert no punishment in the RP-game display a pro-social pattern in the WPS-game.

4 Joint Punishment Demand

It remains to be seen whether this individual behavioral change translates into differences in global punishment demand in the two settings (RPS and WPS). Table 1 presents the results for pseudo panel regressions with individual level fixed effects for the two strategy method settings.

Figure 3: Type Dristribution RPS- and WPS-game



	WPS-game				
RPS ↓	Pun	NPun	APun	NCL	Total
Pun	90	9	0	12	111
%	[39.5]	[3.9]	[0.0]	[5.3]	[48.7]
NPun	21	57	0	10	88
%	[9.2]	[25.0]	[0.0]	[4.4]	[38.6]
APun	2	1	7	3	13
%	[0.9]	[0.4]	[3.1]	[1.3]	[5.7]
NCL	8	2	0	6	16
%	[3.5]	[0.9]	[0.0]	[2.6]	[7.0]
Total	121	69	7	31	228
%	[53.1]	[30.3]	[3.1]	[13.6]	[100.0]

Note: More than 65% of subjects remain consistent across these two games in their punishment behavior.

Game and screen order are used as time variance to construct the pseudo panel and capture potential ordering effects. As figure 1 and figure 2 indicated we find significant positive punishment demands for deviations from fully contributing one's endowment (highest potential effort level) for both settings (RPS and WPS) in the respective estimations in column 1 and 2. The demand for sanctions in the WPS game appears hereby to be larger than the demand for punishment in the RPS game. The estimation in column 3 supports this. The interaction effect $D.WPS \times (20 - g_j)$ in column 3 is positive and significant on the 5 percent level indicating harsher sanctions for every token kept in the private account in the WPS game over the RPS game.

However, this does not reveal from where these additional sanctions stem. As stated above despite remarkable punishment type stability within the experimental subjects we still observe considerable behavioral fluctuations between the two settings. We therefore investigate whether the additional demand for sanctions is an expression of all subjects adapting their sanctioning behavior to the new incentives or just a sub-population, and if this behavioral adaptation is potentially already fully reflected in the two-domain punishment classification.

Column 4 in table 4 presents the results for a pooled RPS and WPS estimation reduced to the 154 *type-stable* subjects (Pun, NPun, APun) along the main diagonal excluding NCL. The demand for punishment $(20 - g_j)$ increases slightly with respect to column 3 and can be explained with the fact that the majority of stable individuals are Pun-type subjects that generate the bulk of positive punishment demand. More interestingly, none of the other coefficients remain significant on any conventional level. This indicates that type-stable subjects are *not* the driver of the increased punishment demand in the WPS game, meaning that the increased demand for sanctions are a

Table 1: Punishment Demand Across Games

	Assigned Punishment d_{ij}			
	RPS	WPS	Joint	Stable
	(1)	(2)	(3)	(4)
$(20 - g_j)$	0.068***	0.085***	0.068***	0.086***
	(0.007)	(0.007)	(0.007)	(0.009)
$D.WPS$			−0.017	0.030
			(0.031)	(0.029)
$D.WPS \times (20 - g_j)$			0.017**	0.001
			(0.007)	(0.006)
Intercept	−0.005	−0.019	−0.000	−0.055
	(0.067)	(0.070)	(0.062)	(0.083)
Observations	6,840	6,840	13,680	9,240
adj.R^2	0.222	0.251	0.201	0.263
AIC	18,036	19,974	41,357	26,656
BIC	18,043	19,980	41,379	26,677

result of individuals that change their sanctioning behavior across the two domains.

5 Contribution and Punishment Domain

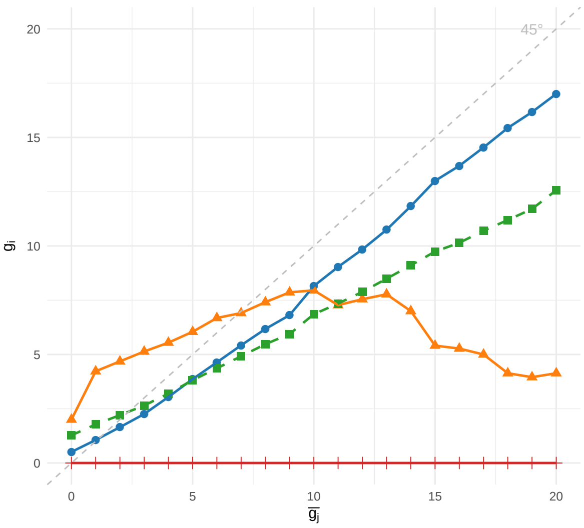
In the previous sections we studied the distribution of punishment types across a linear public good game and a weakest link game. We will now continue to cross-link these punishment classifications with individually elicited contribution types.

5.1 Contribution Types

In similar fashion to Albrecht, Kube, and Traxler 2016 we use the C-game observations to classify contribution types in the tradition of Fischbacher, Gächter, and Fehr 2001. We follow Albrecht, Kube, and Traxler 2016 in that we too classify subjects using linear regression results, deviating from the original Fischbacher, Gächter, and Fehr 2001, who use Spearman rank correlation. In the C-game subjects had to decide upon their *conditional contribution* decision with respect to 21 displayed possible group mean contributions. For each individual we estimate equation 2 in which the average contribution of the other three group member $\overline{g}_j$ explains the individual contribution decision g_i.

$$g_i = a_i + b_i \overline{g}_j + e_i \tag{2}$$

Figure 4: C-game Contribution Types



	Type	N	%
	CC	149	65.4
	FR	42	18.4
	TC	22	9.6
	NC	15	6.6
	Total	228	100.0

Notes: The figure presents the distribution of contribution types, following Fischbacher, Gächter, and Fehr 2001 and **Fischbacher2010** and the average cooperation patterns for the different types: *Conditional Cooperators (CC)*, *Free-Riders (FR)*, *Triangular Contributors (TC)*, and *Non-classified (NC)* cooperation patterns. To ease illustration, the pattern for the latter is not plotted.

Subjects are then classified in accordance with **??**, i.e., *Conditional Contributors* (CC) show a positive slope coefficient b_i, significant on the 1 percent level, *Free Rider* (FR) do not make positive contributions in any of the 21 decision situations, *Triangular Contributors* (TC) show a pattern that initially increases but starts to decrease at a certain point.[4] Similar to the the punishment classification an additional group is introduced to capture behavioral patterns that do not fit with any of the previous groups (NC). Figure 4 presents the results for the contribution type classification. A large majority of subjects (65.4%) show upward sloping contribution patterns and are classified as conditional contributors (CC). Another 18.4% of individuals do not transfer positive amounts to the public goods for any of the 21 group contribution means (FR). Another 9.6% of subjects are classified via eyeballing as triangular contributors (TC). Finally 6.6% individuals did no fit with any of the previous 3 classifications and were labeled an non-classifieds (NC).

Felix: Usefull ?

In a next step we combine these elicited contribution patterns with the previously classified individual punishment patterns to two- and three-dimensional classifications.

4. As in **??** and **??** we classify TC types via eyeballing.

5.2 Contribution × Punishment Types

In the

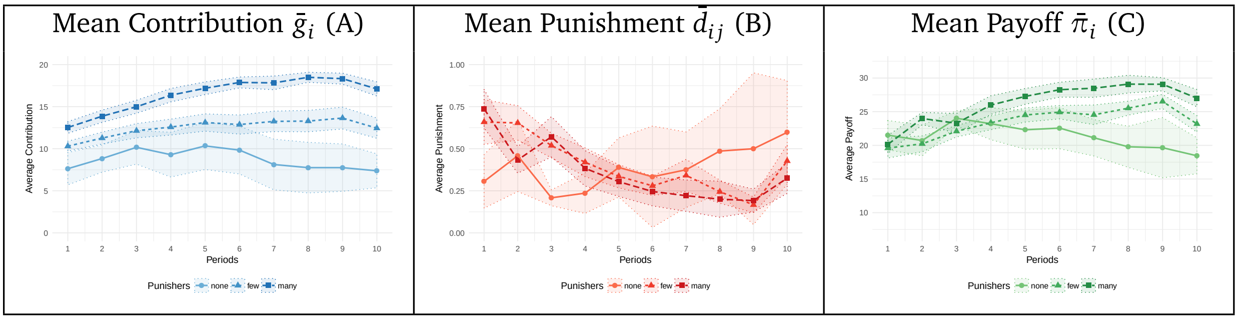
Table 2: Contribution × Punishment Type (C/RPS)

	C-Game				
RPS ↓	CC	FR	TC	NC	Total
Pun	85	18	5	3	111
%	[37.3]	[7.9]	[2.2]	[1.3]	[48.7]
NPun	48	23	9	8	88
%	[21.1]	[10.1]	[4.0]	[3.5]	[38.6]
APun	7	1	3	2	13
%	[3.1]	[0.4]	[1.3]	[0.9]	[5.7]
NCL	9	0	5	2	16
%	[4.0]	[0.00]	[2.2]	[0.9]	[7.0]
Total	149	42	22	15	228
%	[65.4]	[18.4]	[9.7]	[6.6]	[100.0]

Table 3: Contribution × Punishment Type (C/WPS)

	C-Game				
WPS ↓	CC	FR	TC	NC	Total
Pun	89	22	7	3	121
%	[39.0]	[9.7]	[3.1]	[1.3]	[53.1]
NPun	42	13	9	5	69
%	[18.4]	[5.7]	[4.0]	[2.2]	[30.3]
APun	3	0	3	1	7
%	[1.3]	0.0	[1.3]	0.44	[3.1]
NCL	15	7	3	6	31
%	[6.6]	[3.1]	[1.3]	[2.6]	[13.6]
Total	149	42	22	15	228
%	[65.4]	[18.4]	[9.7]	[6.6]	[100.0]

Figure 5: Average Group Outcomes by Type Prevalence in RP-game



Note: Panel A shows the development of mean contributions to the public good over the 10 periods of the linear public good RP-game by Pun-type prevalence per group. Panel B depicts the mean punishment assigned to other group members, panel C the mean individual payoff subjects received. The 20 groups containing 3 or 4 Pun-type subjects are jointly depicted as *many*, the 31 groups with 1 or 2 Pun-types as *few*. The 6 groups without Pun-types are depicted separately. Figure **??** in the supporting online material plots group compositions separately.

6 Impact of Type Prevalence on Group Outcomes

To recap we found that individuals are heterogeneous in their punishment behavior in two different game settings. The majority of subjects, when classified with a fairly simple approach, remain type-stable across the two settings despite differing incentive structures. We further find that the average demand for sanctions within the weakest link setting (WPS) is significantly larger than in the linear public goods setting (RPS) and that this increased demand stems from type-switching individuals, who on average increase their demand for sanctions, while the type-stable individual show no change in sanctioning demand.

We will now investigate the impact of the elicited behavioral types (punishment and cooperation) on group success. In the case of the 10 period repeated linear public good RP-game group success is measured in increased average contributions and increased average payoffs. For the 10 period repeated weakest link WP-game we are concerned with improved efficiency, i.e., improved coordination on *any* or on the *payoff-dominant* equilibrium.

We use the group composition of Pun and CC-type individuals, as classified in the RPS, WPS and C-game, resulting from the randomized assignment to experimental groups at the beginning of the RP and WP-game as exogenous variation to test their respective impact on group success.[5]

Table 4: Cooperation in the RP-game

	Mean Contribution $\bar{g}_i$			Mean Payoff $\bar{\pi}_i$		
	(1)	(2)	(3)	(4)	(5)	(9)
No. of Pun	2.228***		1.997***	1.350**		1.349**
	(0.537)		(0.573)	(0.567)		(0.577)
No. of CC		1.644**	0.817		0.560	0.002
		(0.677)	(0.746)		(0.616)	(0.673)
Constant	9.144***	9.185***	7.457***	21.584***	22.747***	21.580***
	(1.450)	(1.966)	(2.150)	(1.337)	(1.584)	(1.847)
Obs.	570	570	570	570	570	570
R^2	0.222	0.082	0.239	0.091	0.011	0.091
AIC	2,949	2,958	2,950	3,538	3,543	3,540
BIC	2,966	2,976	2,971	3,556	3,560	3,562

Note: Random effects panel estimation for 57 groups of 4 subjects over 10 periods. The explanatory variables can take values from 0 to 4. Columns 1-3 show estimation results for the impact of type prevalence on average contributions. Columns 4-6 show results for the impact on average payoffs. P-values: 0.1 / 0.05 / 0.01 : * / ** / ***. Cluster robust standard errors in parentheses.

6.1 Cooperation Outcomes (RP-game)

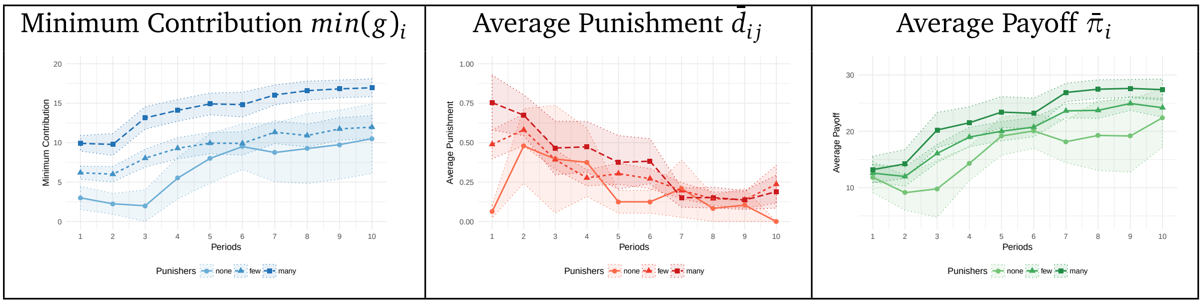
In this chapter we replicate the results of Albrecht, Kube, and Traxler 2016. To elicit the impact of Pun-types on group cooperation we use the individual punishment classification resulting from the RPS-game and the contribution classification from the C-game. Figure 5 shows the development over the ten periods in terms of mean contribution $\bar{g}_i$ (A), mean assigned punishment $\bar{d}_{ij}$ (B), and mean payoff $\bar{\pi}_i$ (C) given the prevalence of pro-social punishers in the respective groups. We reduced the five possible group compositions into straight-forward cluster groups to ease viewing.[6] The 6 groups *without* Pun-type subjects remained separately as 'none'. We joined groups with one (14 groups) or two (17 groups) Pun-types into the category 'few' as these contained less than 50% Pun-types. Further we gathered groups with three (17 groups) or four (3 groups) Pun-types in the category 'many' as they contain more than 50% of Pun-types.

In panel A of figure 5 it is apparent that groups with *many* Pun-type subjects increase their contributions over the course of the 10 periods of the RP-game and also the resulting payoffs appear to be highest in the second half of the RP-game for groups with many pro-social punishers.

5. Section **??** in the supporting online materials shows the hypothetical distribution of Pun and CC-type subjects by running Monte Carlo simulations with 10,000 random draws from the observed individual type distribution in the C, RPS, and WPS-game with randomized assignment into groups of 4. The created hypothetical distributions is tested against our observed distributions using χ^2-tests. We find no significant differences between the hypothetical and observed distributions (**Numbers**).

6. Figure **??** in the supporting online material shows graphs for all group compositions separately.

Figure 6: Average Group Outcomes by Type Prevalence in WP-game



Note: Panel A shows the development of minimum contributions to the public good over the 10 periods of the linear public good RP-game by Pun-type prevalence per group. Panel B depicts the mean punishment assigned to other group members, panel C the mean individual payoff subjects received. The 20 groups containing 3 or 4 Pun-type subjects are jointly depicted as *many*, the 31 groups with 1 or 2 Pun-types as *few*. The 6 groups without Pun-types are depicted separately. Figure **??** in the supporting online material plots group compositions separately.

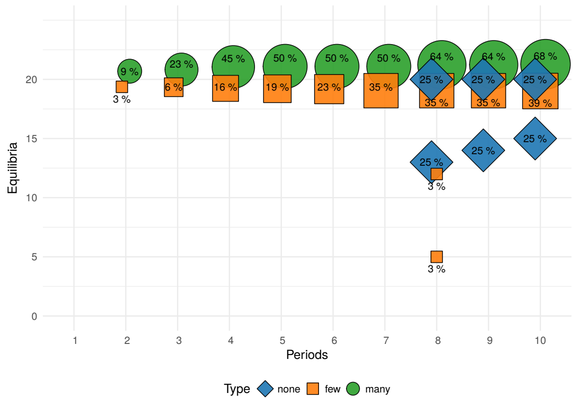
Table 4 presents panel estimation results that support these visual findings. Columns 1-3 show estimates for the impact on mean contributions $\bar{g}_i$ and columns 4-6 for mean payoffs $\bar{\pi}_i$. Increasing numbers of pro-social punishers per group show a significant positive effect on the level of mean contributions and on the level of mean payoffs. Moreover the results suggest that the number of conditional contributors does not enhance within group contribution levels. This is finding is further supported by the low explanatory power (R^2) of models 2 and 5 and the increases in the Akaike and Bayesian information criterion in models 2 and 3 (5 and 6) compared to model 1 (4). Our findings are similar with the difference that *many* CC-types show a significant impact in Albrecht, Kube, and Traxler 2016.[7]

6.2 Coordination

Turning to the weakest link game we now want to investigate group success in a coordination setting. Contrary to the public goods game where zero contribution is the sole subgame perfect Nash-equilibrium, in the weakest link game every joint equal effort level (contribution level) for the public good is a Nash-equilibrium. These equilibria are pareto ranked from contributing nothing being the least efficient to contributing fully being the payoff-dominant Nash-equilibrium.

7. Contrary to **??** we used the number of Pun and CC-type subjects in their linear form to perform our estimations. This was necessary as the distribution of group compositions in the margins – 6 groups without Pun-types, 1 group without CC-types – did not warrant separation by clusters using dummy variables. Nevertheless, the graphical representations support the general findings of the regression models.

Figure 7: Distribution of Equilibrium Coordination by Type Prevalence



Note: Fraction of successful equilibrium coordination by the 4 groups *without* Pun-types (none), 33 with *few*, and 20 groups with *many* Pun-type subjects, as classified based on WPS behavior. One group *without* Pun-type subjects coordinates on the payoff dominant equilibrium in period 8, 9, and 10. A second group without Pun-types manages to coordinate on the not payoff-dominant equilibria 13, 14, and 15 in periods 8, 9, and 10. Groups with *few* Pun-types coordinate twice on equilibria that are not payoff-dominant in period 8. In the majority of cases groups with *few* Pun-types coordinate on the payoff-dominant equilibrium when successfully coordinating. Groups with *many* Pun-types exclusively coordinate successfully on the payoff dominant equilibrium. All groups fail to tacitly coordinate on an equilibrium in the first period.

Table 5: Equiblibrium Coordination in the WP-game

	Any Potential Equiblibrium			Payoff-Dominant Equilibrium		
	(1)	(2)	(3)	(4)	(5)	(9)
No. of Pun	0.285		0.220	0.697***		0.635**
	(0.188)		(0.210)	(0.255)		(0.281)
No. of CC		0.278	0.164		0.487*	0.164
		(0.216)	(0.238)		(0.292)	(0.314)
Intercept	−1.443***	−1.572***	−1.734***	−2.822***	−2.624***	−3.120***
	(0.452)	(0.609)	(0.625)	(0.652)	(0.841)	(0.880)
Obs.	570	570	570	570	570	570
AIC	564	565	566	470	475	472
BIC	577	578	583	483	488	489

Note: Random effects panel probit estimation for 57 groups of 4 subjects over 10 periods. The explanatory variables can take values from 0 to 4. Columns 1-3 show estimation results for group coordinating on *any* possible equilibrium in the game. Columns 4-6 show results for coordination on the *payoff-dominant* equilibrium. P-values: 0.1 / 0.05 / 0.01 : * / ** / ***. Standard errors in parentheses.

References

Albrecht, Felix, Sebastian Kube, and Christian Traxler. 2016. "Cooperation and Punishment: The Individual-Level Perspective."

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch. 2014. "hroot: Hamburg Registration and Organization Online Tool." *European Economic Review* 71:117–120.

Brandts, Jordi, and David J. Cooper. 2006. "A Change Would Do You Good .... An Experimental Study on How to Overcome Coordination Failure in Organisations." *American Economic Review* 96 (3): 669–693.

Fehr, Ernst, and Simon Gächter. 2002. "Altruistic punishment in humans." *Nature* 415 (6868): 137–140.

Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-made Economic Experiments." *Experimental Economics* 10 (2): 171–178.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are people conditionally cooperative? Evidence from a public goods experiment." *Economics Letters* 71 (3): 397–404.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter. 2008. "Antisocial punishment across societies." *Science* 319 (5868): 1362–7.

Huyck, John B. Van, Raymond C. Battalio, and Richard O. Beil. 1990. "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure." *American Economic Review* 80 (1): 234–248.

Kube, Sebastian, and Christian Traxler. 2011. "The Interaction of Legal and Social Norm Enforcement." *Journal of Public Economic Theory* 13 (5): 639–660.

Le Lec, Fabrice, Ondrej Rydval, and Astrid Matthey. 2015. "Efficiency and Punishment in a Coordination Game: Voluntary Sanctions in the Minimum Effort Game."

Supporting Online Material

(*For Online Publication Only*)

Coordination Through Punishment – An individual Level Perspective –

Felix Albrecht[†], Sebastian Kube[‡,♯]

June 4, 2017

† University of Marburg; ‡ University of Bonn; ♯ Max Planck Institute for Research on Collective Goods;

* Hertie School of Governance

Contents

List of Tables

List of Figures

A Contribution Triplets

Below we list the contribution triples that were used within each combination of g^L, g^M and g^H (see Table **??**). Before the experiment, these 10 × 8 triples were randomly generated by sampling with replacement from the corresponding sets g^L, g^M, g^H. Each player then faced a randomly selected triple within each combination 1 – 10. If the selected triple would by chance correspond to the real triple, the subject would not face this situation but instead another one of the pre-defined contribution triples for the corresponding combination.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	(g^L, g^L, g^L):	(0,0,0)	(0,2,3)	(1,1,3)	(1,2,2)	(1,2,3)	(1,2,4)	(1,3,3)	(1,3,4)
(2)	(g^L, g^L, g^M):	(0,1,5)	(0,2,8)	(0,2,14)	(1,2,10)	(1,2,12)	(1,3,14)	(2,2,6)	(2,3,12)
(3)	(g^L, g^L, g^H):	(0,3,18)	(1,2,20)	(1,3,19)	(1,4,20)	(2,2,18)	(2,2,19)	(3,3,18)	(4,4,17)
(4)	(g^L, g^M, g^M):	(0,9,11)	(0,5,12)	(0,13,14)	(1,10,15)	(2,6,8)	(2,9,11)	(2,10,15)	(3,13,14)
(5)	(g^L, g^M, g^H):	(0,6,19)	(0,14,17)	(2,6,17)	(2,8,20)	(2,11,19)	(3,7,18)	(4,8,17)	(4,10,20)
(6)	(g^L, g^H, g^H):	(0,18,19)	(1,19,19)	(2,18,19)	(2,18,20)	(2,19,19)	(3,18,20)	(3,19,19)	(4,19,20)
(7)	(g^M, g^M, g^M):	(5,7,12)	(5,14,16)	(6,6,9)	(6,10,10)	(7,8,9)	(7,10,13)	(7,14,16)	(8,9,11)
(8)	(g^M, g^M, g^H):	(5,5,17)	(5,8,18)	(6,11,20)	(8,15,17)	(9,12,18)	(9,15,18)	(11,15,19)	(12,15,19)
(9)	(g^M, g^H, g^H):	(5,18,20)	(7,18,19)	(9,18,20)	(11,17,17)	(12,17,18)	(12,18,18)	(14,17,20)	(15,17,19)
(10)	(g^H, g^H, g^H):	(17,17,19)	(17,18,19)	(17,18,20)	(17,19,19)	(17,19,20)	(18,18,19)	(18,18,20)	(20,20,20)

B Instructions

Below you find the English set of instructions used at FSU. The German and Japanese set of instructions can be requested from the authors. The first part describes a public good game without punishment in the tradition of (Fischbacher, Gächter, and Fehr 2001), which was played ahead of the core games with punishment. The second part describes the *One-Shot* game, the third the *Repeated Interaction* game respectively.

C Contribution vs. Payoff Discussion

Table 1: Contribution versus Payoff as Explanatory Variables

	Individually Assigned Punishment d_{ij}			
	RPS		WPS	
	(1)	(2)	(3)	(4)
$(20 - g_j)$	0.068***		0.085***	
	(0.007)		(0.007)	
π_j		0.062***		0.038***
		(0.006)		(0.003)
Intercept	−0.005	−0.962***	−0.019	0.182***
	(0.067)	(0.158)	(0.070)	(0.057)
Observations	6,840	6,840	6,840	6,840
Adjusted R^2	0.222	0.098	0.251	0.048
AIC	18,036	19,054	19,974	21,619
BIC	18,043	19,061	19,980	21,626

Note: Pseudo panel estimation with individual level fixed effects for 228 subjects. Cluster robust standard errors in parentheses. P-values: 0.1 / 0.05 / 0.01 : * / ** / ***. Column (1), (3), and (5) only RP-game, (2), (4), and (6) only WP-game observations with 30 observations per subjects in each game.

Table 2: Contribution Difference vs. Payoff Difference as Explanatory Variables

	Individually Assigned Punishment d_{ij} (1)	(2)	
$max(\pi_j - \pi_i, 0)$	-0.008	-0.008	$max(g_j - g_i, 0)$
	(0.006)	(0.006)	
$max(\pi_i - \pi_j, 0)$	0.120***	0.120***	$max(g_i - g_j, 0)$
	(0.012)	(0.012)	
$D.WPS \times max(\pi_j - \pi_i, 0)$	-0.003	-0.003	$D.WPS \times max(g_j - g_i, 0)$
	(0.006)	(0.006)	
$D.WPS \times max(\pi_i - \pi_j, 0)$	-0.004	-0.004	$D.WPS \times max(g_i - g_j, 0)$
	(0.012)	(0.012)	
$D.min(g_{jkl})$	0.278***	0.278***	$D.min(g_{jkl})$
	(0.057)	(0.057)	
$D.WPS \times D.min(g_{jkl})$	0.043	0.043	$D.WPS \times D.min(g_{jkl})$
	(0.072)	(0.072)	
Intercept	0.061	0.061	Intercept
	(0.056)	(0.056)	
Observations	13,680	13,680	Observations
Adjusted R^2	0.319	0.319	Adjusted R^2
AIC	39,175	39,175	*AIC*
BIC	39,220	39,220	*BIC*

Note: Pseudo panel estimation with individual level fixed effects for 228 subjects. Cluster robust standard errors in parentheses. P-values: 0.1 / 0.05 / 0.01 : * / ** / ***. Column (1), (3), and (5) only RP-game, (2), (4), and (6) only WP-game observations with 30 observations per subjects in each game.

D Additional Analyses

D.1 Self-Centered Punishment Classification

Table 3: Type Dristribution RP- and WP-game

t_sdiff_bz41		t_sdiff_61 -12	-10	-1	0	2	20	23	100	Total
-11	No.	0	1	0	0	0	0	0	0	1
	%	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.4
-10	No.	0	50	0	6	0	0	0	6	62
	%	0.0	21.9	0.0	2.6	0.0	0.0	0.0	2.6	27.2
-1	No.	0	0	38	3	0	0	0	5	46
	%	0.0	0.0	16.7	1.3	0.0	0.0	0.0	2.2	20.2
0	No.	0	8	14	57	0	0	0	9	88
	%	0.0	3.5	6.1	25.0	0.0	0.0	0.0	3.9	38.6
2	No.	0	0	0	1	1	0	0	1	3
	%	0.0	0.0	0.0	0.4	0.4	0.0	0.0	0.4	1.3
20	No.	1	0	0	0	0	2	0	3	6
	%	0.4	0.0	0.0	0.0	0.0	0.9	0.0	1.3	2.6
100	No.	0	7	2	2	0	0	1	10	22
	%	0.0	3.1	0.9	0.9	0.0	0.0	0.4	4.4	9.6
Total	No.	1	66	54	69	1	2	1	34	228
	%	0.4	28.9	23.7	30.3	0.4	0.9	0.4	14.9	100.0

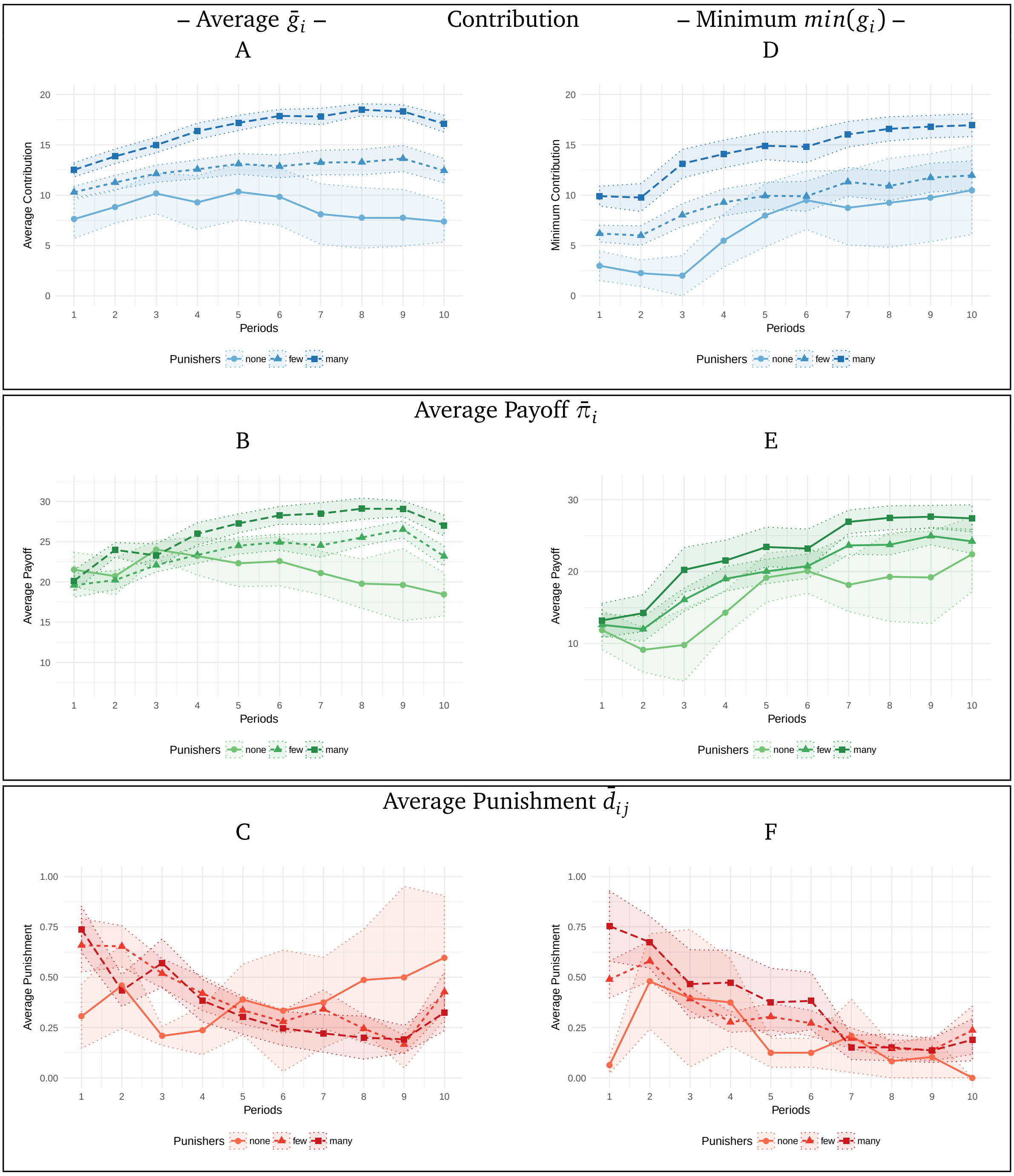
Note:

D.2 Payoff Comparison

Table 4: Punishment Demand

	Assigned Punishment d_{ij}							
	RPS	WPS	Joint	stable	RPS	WPS	Joint	stable
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$(20 - g_j)$	0.068***	0.085***	0.068***	0.086***	0.056***	0.072***	0.052***	0.073***
	(0.007)	(0.007)	(0.007)	(0.009)	(0.007)	(0.007)	(0.007)	(0.009)
$D.WPS$			−0.017	0.030			−0.025	0.026
			(0.031)	(0.029)			(0.031)	(0.030)
$D.WPS \times (20 - g_j)$			0.017**	0.001			0.019***	0.001
			(0.007)	(0.006)			(0.007)	(0.006)
$D.min(g_{jkl})$					0.473***	0.424***	0.580***	0.450***
					(0.051)	(0.045)	(0.064)	(0.066)
$D.WPS \times D.min(g_{jkl})$							−0.130*	−0.030
							(0.076)	(0.078)
Intercept	−0.005	−0.019	−0.000	−0.055	0.025	0.009	0.036	−0.027
	(0.067)	(0.070)	(0.062)	(0.083)	(0.065)	(0.069)	(0.061)	(0.082)
Observations	6,840	6,840	1,3680	9,240	6,840	6,840	1,3680	9,240
adj.R^2	0.222	0.251	0.201	0.263	0.246	0.267	0.223	0.280
AIC	18,036	19,974	41,357	26,656	17,821	19,827	40,978	26,442
BIC	18,043	19,980	41,379	26,677	17,835	19,841	41,016	26,477

Figure 1: Average Group Outcomes by Type Prevalence



Note: More than 65% of subjects remain consistent across these two games in their punishment behavior.

Table 5: Mean Contribution g_i in RP- and WP-game

	μ_{g_i}	SE
RP	10.798	(0.457)
WP	13.447	(0.397)
t-test		0.000

Note: