Behavioural types in public goods games: A re-analysis by hierarchical clustering*

Francesco Fallucchi
University of East Anglia

R. Andrew Luccasen
Mississippi University for Women

Theodore L. Turocy
University of East Anglia

January 12, 2017

Abstract

We re-analyse participant behaviour in standard economics experiments studying voluntary contributions to a public good. Previous approaches were based in part on a priori models of decision-making, such as maximising personal earnings, or reciprocating the behaviour of others. Many participants however do not conform to one of these models exactly, requiring ad hoc adjustments to the theoretical baselines to identify them as belonging to a given behavioural type. We construct a typology of behaviour based on a similarity measure between strategies using hierarchical clustering analysis. We identify four clearly distinct behavioural types which together account for over 90% of participants in six experimental studies. The resulting type classification distinguishes behaviour across groups more consistently than previous approaches.

Keywords: behavioral economics, cluster analysis, cooperation, public goods

JEL Classifications: C65, C71, H41.

*Fallucchi and Turocy acknowledge the support of the Network for Integrated Behavioural Science (Economic and Social Research Council Grant ES/K002201/1). We thank seminar participants at Queen's University Belfast, University of East Anglia, and University of Nottingham for useful comments, as well as the authors of the cited previous studies for generously providing their data. All errors are the responsibility of the authors. Corresponding author: T.Turocy@uea.ac.uk.

1 Introduction

The literature in behavioural economics has developed the concept that people's behaviour can often be classified into qualitatively distinct "types." However, the definition of the set of possible types is often heavily influenced, if not driven, by underlying theories of preference or decision-making process. For example, Stahl and Wilson (1995) proposed the successful level-k model, which distinguishes people by the number of steps of reasoning used in games.

In this paper, we construct a typology of behaviour in public goods voluntary contribution games (VCGs). We re-analyse the data from a number of previous laboratory experiments, using cluster analysis to extract from the data groups of participants, in which members of the same group exhibit similar behaviour, while members of different groups exhibit qualitatively different behaviour. Although our approach does not make a priori reference to models of economic or strategic decision-making, we find that types which emerge endogenously from the data do align with distinct archetypes with intuitive economic or strategic interpretations.

Our work is based on the *P-experiment* protocol for VCGs designed by Fischbacher et al. (2001). This protocol has been widely adopted as a standard methodology, employed by many studies over a decade with results reported for experiments conducted across a variety of languages and continents (Kocher et al., 2008). The protocol is built on the linear VCG (Ledyard, 1997). Upon arrival in the lab, participants are placed into groups with M members. Each participant receives G "tokens." She can allocate all, some, or none of these tokens to a group account; the remainder are kept in her private account. We refer to a participant's allocation to the group account as her *contribution*.[1] A participant receives a point for each token kept in her private account. Each token contributed to the group account yields $P > 1$ points, which are split equally among the group, with P/M tokens received by each group member. This VCG presents a social dilemma: the strategy that maximises individual earnings is to keep all tokens in the private account, while the strategy profile that maximises total earnings of the group is for each member to allocate all G tokens to the group account.

In the P-experiment protocol, contributions are made in two stages. In the first stage, $M - 1$ of a group's members make their contributions. In the second stage, the remaining member observes the average contribution made in the first stage, and then makes her contribution. The decisions of the participants are elicited using the *strategy method* (Selten, 1967). Participants do not know in advance whether they will play in the first or second stage. Instead they indicate what their contribution would be if they are chosen to play in the first stage, and they indicate what their contribution would be if they are chosen to play in the second stage. Because the player who

[1] Because tokens not allocated to the group account are automatically kept in the private account, participants' contributions capture all of the relevant decision information, and so we write the analysis in terms of contributions.

contributes in the second stage knows the average contribution made in the first stage, the strategy method asks participants to specify their contribution in the second stage for each possible result of the first stage. A key objective of the P-experiment protocol is the elicitation of these second stage contributions, which we refer to as a participant's *contribution strategy*, as it is on these that the identification of behavioural types is based.

2 Model

Formally the vector c^i denotes the contribution strategy of a participant i. The component c^i_k is the contribution of participant i to the public account in the second stage, if the other group members contribute k tokens on average in the first stage. (In the P-experiment protocol, the average contribution of other group members is rounded to the nearest integer.) Let $\mathcal{C}$ denote the set of all participants' contribution strategies. We define a *typology* T as a partition of $\mathcal{C}$ into equivalence classes, where each equivalence class defines a distinct type. We write $T(i)$ as the type of participant i in typology T.

Various methods could be used to generate candidate typologies. To evaluate candidate methods for constructing typologies, we propose three criteria:

- If two participants i and j are classified as the same type, $T(i) = T(j)$, then their behaviour should be "similar."
- If two participants are deemed to be different types, $T(i) \neq T(j)$, then their behaviour should be sufficiently dissimilar.
- While classification of behaviour into types is a useful descriptive tool on its own, a typology should be useful in predicting or distinguishing related economic behaviour not used in the classification process.

The existing state-of-the-art in the literature is the typology based on Fischbacher et al. (2001), which we will call T^F. T^F partitions participants into one of four types, which are defined by a set of rules combining stereotypical behaviour and rules of thumb. *Free-riders* (FR) always maximise individual earnings by keeping all tokens in the private account, irrespective of the outcome of the first stage. *Conditional cooperators* (CC) exhibit willingness to increase their own allocations to the group account based on higher allocations by others in the first stage. A participant i is deemed a conditional cooperator by testing whether the Spearman's ρ correlation coefficient between the vector $[0, 1, \ldots, G]$ and the strategy $[c^i_0, c^i_1, \ldots, c^i_G]$ is significantly positive at significance level ≤ 0.001. *Hump-shaped* (HS) contributors are identified based on visual classification of contribution strategies, in which c^i_0 and c^i_G are small, but c^i_k is larger for some intermediate values $0 < k < G$;

these strategies often have a triangular shape when plotted. Any participants whose contribution strategies do not satisfy one of these three sets of criteria are placed in the residual type *others* (OT).

Recent studies have considered adjustments or relaxations of these classification rules. Fischbacher et al. (2012) define conditional cooperators as participants with a monotonic contribution pattern with at least one increase. Rustagi et al. (2010) split conditional cooperators based on the level of statistical significance of the correlation between own contributions and the contributions of other group members, in which strong conditional cooperators are those for whom Spearman's ρ is significantly positive at significance level ≤ 0.001 and weak cooperators are those for whom it is significantly positive at significance level in $(0.001, 0.05]$. They also allow a participant to be a free-rider if at most one entry in the contribution strategy c^i is positive. Adjusting the classification criteria can lead to fewer participants being placed in the residual others type, but also creates the possibility that a participant's strategy could satisfy the criteria for more than one of the types.

We propose an alternative approach to constructing a behavioural typology using hierarchical cluster analysis. We define the distance, or dissimilarity, between two contribution strategy c^i and c^j as the Manhattan distance between the vectors. Economically, this metric can be interpreted as the expected difference between the second stage contributions of participants i and j, if the average contribution k of other group members is chosen uniformly at random. We use Ward's minimum variance method (Ward, 1963) to cluster subjects hierarchically based on this distance measure, with Duda-Hart Je(2)/Je(1) index (Duda and Hart, 1973) determining the number of clusters to produce. We label the typology resulting from this algorithm as T^H.

3 Results

We re-analyse the data from six VCG experiments using the P-experiment protocol, published between 2001 and 2016. The experiments were conducted in four different countries and four different languages, with a total of $N = 551$ participants. All experiments use groups of $M = 4$ participants and each participant has $G = 20$ tokens to allocate.

We break out the classification proportions for each of the six individual studies, and for the population as a whole, in Table 1 for typology T^F and Table 2 for typology T^H. To construct T^F we apply the classification rules from Fischbacher et al. (2001) as listed above; therefore our quoted percentages of types for some later studies vary slightly from those which appear in the corresponding papers, in the case where those authors used a variant on the rules.

The Duda-Hart stopping rule produces 5 clusters in T^H. These clusters are endogenously derived from the data, so for expositional purposes we attach to each of them a label which captures the modal or stereotypical behaviour observed in the cluster. The inspiration for the type labeling

Study	Country	FR	CC	HS	OT	N
Fischbacher et al. (2001)	SUI	29.6%	50.0%	13.6%	6.8%	44
Herrmann and Thöni (2009)	RUS	6.3%	52.5%	6.2%	35.0%	160
Fischbacher and Gächter (2010)	SUI	22.9%	50.7%	12.1%	14.3%	140
Fischbacher et al. (2012)	GBR	14.7%	69.1%	6.6%	9.6%	136
Cartwright and Lovett (2014)	GBR	6.5%	74.2%	3.2%	16.1%	31
Préget et al. (2016)	FRA	22.5%	37.5%	15.0%	25.0%	40
	Overall	**15.6%**	**56.1%**	**8.9%**	**19.4%**	**551**

Table 1: Classification of participants in T^F typology.

Study	Country	OWN	SCC	WCC	UNC	VAR	N
Fischbacher et al. (2001)	SUI	38.6%	29.5%	18.2%	2.3%	11.4%	44
Herrmann and Thöni (2009)	RUS	10.6%	36.3%	19.4%	5.6%	28.1%	160
Fischbacher and Gächter (2010)	SUI	39.3%	36.4%	17.1%	3.6%	3.6%	140
Fischbacher et al. (2012)	GBR	27.2%	50.5%	18.4%	3.7%	0.7%	136
Cartwright and Lovett (2014)	GBR	13.0%	45.2%	29.0%	3.2%	9.7%	31
Préget et al. (2016)	FRA	30.0%	25.0%	17.5%	12.5%	15.0%	40
	Overall	**25.8%**	**38.8%**	**18.9%**	**4.7%**	**11.8%**	**551**

Table 2: Classification of participants in T^H typology.

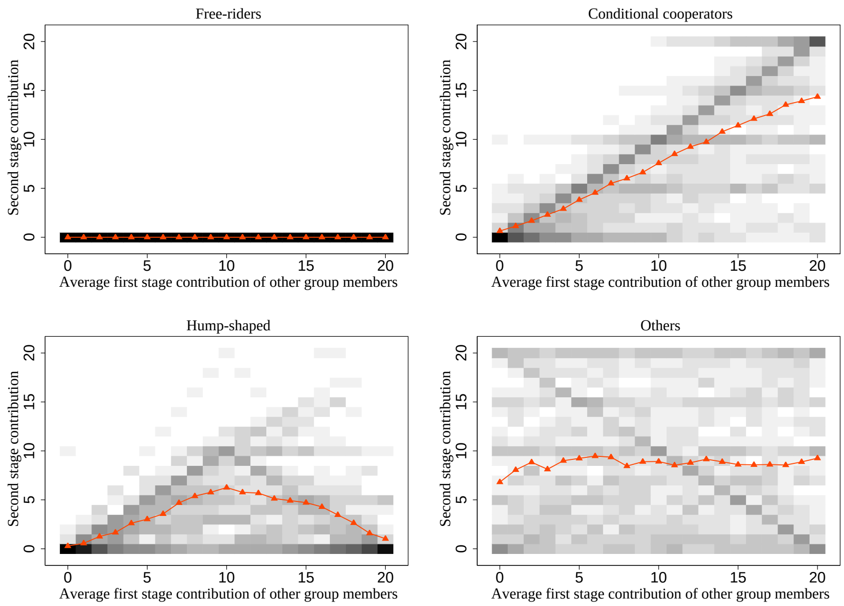

Figure 1: Heatmaps of contribution strategies, typology T^F.

can be seen through the heatmaps of the contribution strategies assigned to each type. The heatmap for a type t is produced by taking the contribution strategies of all participants assigned to that type, and constructing the set $\{(k, c_k^i)\}_{T(i)=t,k=0,\ldots,20}$. The frequencies of these ordered pairs is used to generate the heatmap, with darker shades corresponding to higher frequencies. The heatmaps for the types in T^F are plotted in Figure 1, and those for T^H are in Figure 2. In all figures, red triangles indicate the average contribution of the type, conditional on the first stage outcome.

The modal behaviours which give the types in T^F their names are evident in the heatmaps. We now introduce the types produced in T^H via hierarchical clustering. *Own-maximisers* (OWN, 25.8% of participants) allocate zero or very few tokens to the group account in all contingencies, named because the modal allocation is zero. The modal behaviour among *strong conditional cooperators* (SCC, 38.8%) is to match average allocations on a one-to-one basis. The classification contrasts these with *weak conditional cooperators* (WCC, 18.9%), who have allocation strategies that are increasing in the allocations of other group members, contributing on average about half as much as the other group members. There is a small but distinct group of *unconditional cooperators* (UNC, 4.7%), corresponding to contributions which are at or near full contribution of the tokens to the group account. (Recall that full contribution to the group account maximises the group's total earnings from those tokens.) This labeling parallels that of Kurzban and Houser (2005), where, in a setting in which participants play the game repeatedly, those whose always contribute more than the average of the others in their group are called (unconditional) "cooperators". The final cluster (11.8%) is labeled *various* (VAR), and contribute on average about one-half of the tokens. This group is the most diverse; in addition to a high frequency of contributions exactly equal to 10,

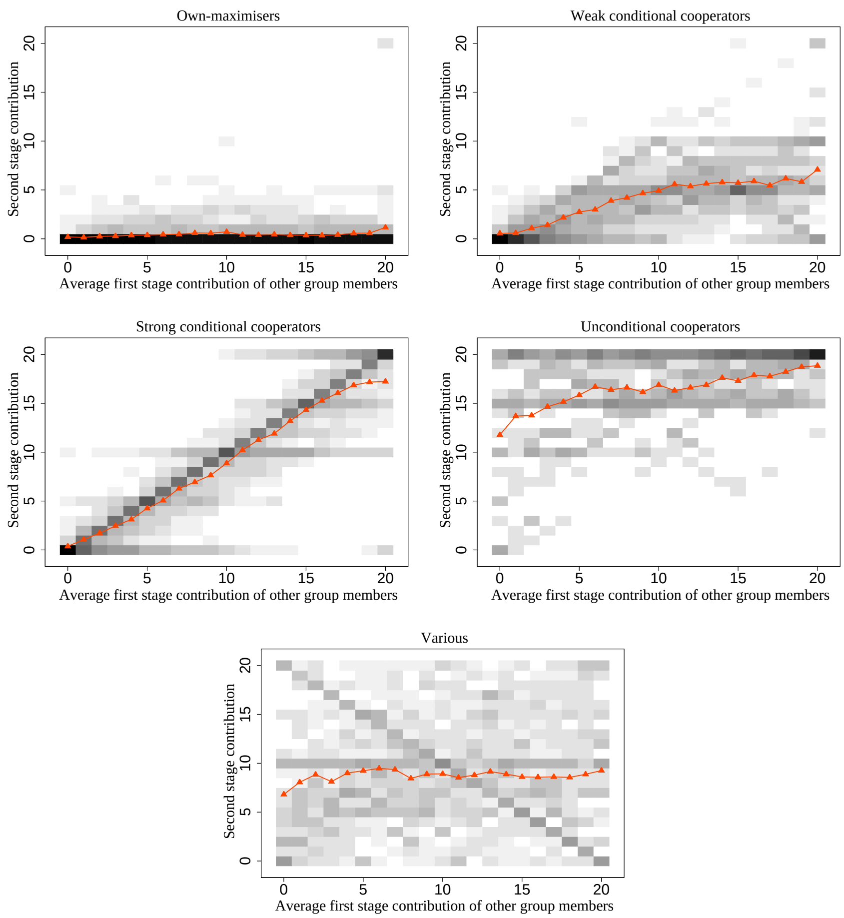

Figure 2: Heatmaps of contribution strategies, typology T^H.

		T^F				
		FR	CC	HS	OT	Total
T^H	OWN	86	24	19	13	**142**
	SCC	0	206	7	1	**214**
	WCC	0	70	21	13	**104**
	UNC	0	10	0	16	**26**
	VAR	0	15	2	48	**65**
	Total	**86**	**325**	**49**	**91**	**551**

Table 3: Comparison of typologies.

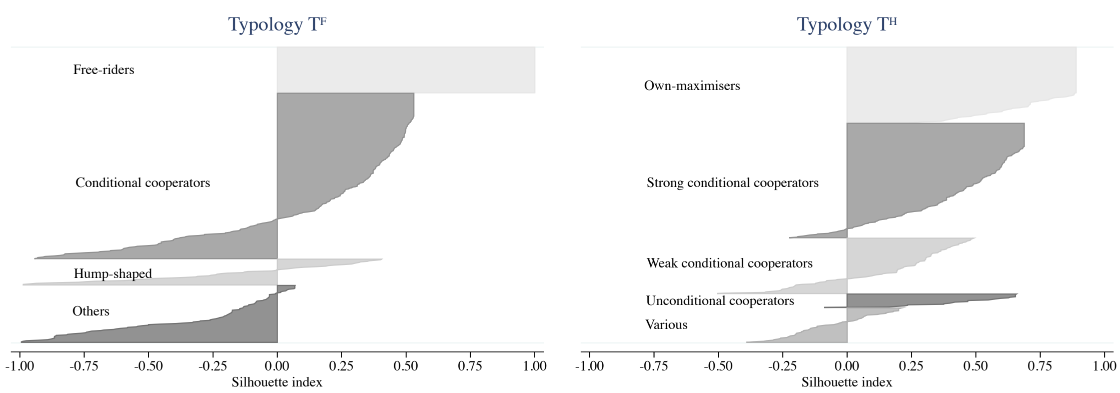

Figure 3: Silhouette plots of type clusters.

there are also "negative conditional contributors" (Burton-Chellew et al., 2016) who decrease their contribution in response to higher contributions by others.

The relationship between the typologies T^F and T^H can be seen in Table 3. Each cell contains the number of participants who are classified under T^F into the column label type, and under T^H into the row label type. Conditional cooperators in T^F split primarily between strong and weak conditional cooperators in T^H, while hump-shaped contributors are distributed across both conditional cooperator types and own-maximisers in T^H. These observations combined with the heatmaps suggest that conditional cooperators and hump-shaped contributors under T^F are not particularly cohesive, and are binning together economically distinct behaviours.

We evaluate the cohesion of the types generated by T^H compared to T^F using silhouette plots (Rousseeuw, 1987) in Figure 3. This method assigns to each participant i an index in $[-1, +1]$. This number compares the average distance between participant i's strategy and the strategies of other participants of the same type, against the average distance to participants who are classified in the "next closest" type. A negative value of the index for participant i indicates there is another type whose strategies are on average closer to participant i's strategy, than the strategies of other participants classified as being the same type as i. Looking at the plot for the T^F typology, a

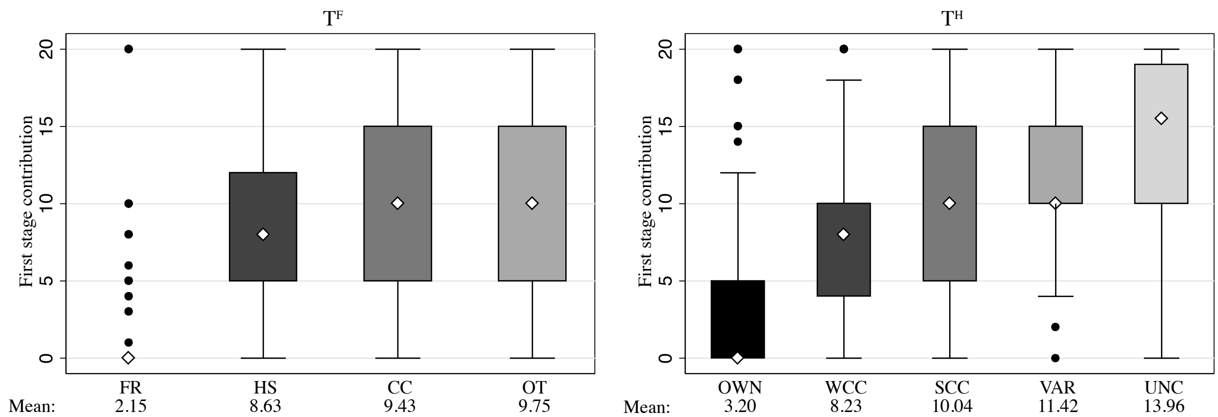

Figure 4: Distributions of first stage contributions by type.

majority of participants identified as hump-shaped contributors (29 of 49) have strategies which are on average closer to one of the other three types' strategies, than to other hump-shaped contributors. Among those identified as others, 93 of 107 have strategies closer on average to one of the other three types than to the rest of those considered others. Many conditional cooperators likewise have negative indices.

We can compare this with the silhouette plot for the types generated by typology T^H. Unconditional cooperators are identified as a distinct and coherent type, with members having large positive indices. Own-maximisers all have positive indices, while almost all strong conditional cooperators (197 of 214) do as well. The distinction between strong conditional cooperators and weak conditional cooperators helps to avoid the large negative indices observed among T^F's conditional cooperators. The heterogeneity of the remaining participants classified as various is evident in the range of indices among the participants; although a majority (39 of 65) have negative indices, the magnitudes are much smaller than those measured for the others type in T^F.

The silhouette plots are thus useful in understanding the properties of the T^F typology, measured in terms of the economic consequences of the contribution strategies of the various types. Consider one of the conditional cooperators who has a silhouette index which is negative and large in magnitude, and recall that the Manhattan distance is one measure of the expected difference in contribution behaviour between two contribution strategies. A strongly negative silhouette index means that the observed contributions of this participant are much closer to the contributions of one of the other types, than they are to the observed contributions of conditional cooperators as a whole.

We turn now to an analysis of contributions in the first stage. Recall that first stage contributions are not used in constructing either the T^F or T^H typologies. The T^H typology does a better job of distinguishing the first stage contributions of participants. Figure 4 shows the distributions of first stage contributions, grouped by the type assignment of the participant based on their second stage strategy. In the T^F typology, free-riders allocate on average 2.15 tokens (with a mode at zero), while the other three types have dispersed distributions of first stage contributions with means and medians near half of the endowment of tokens. The first stage allocation of free-riders is different from other types (all Bonferroni multiple-comparisons test $p < 0.001$), while there is no significant difference in first stage allocations among conditional cooperators, hump-shaped, and others.

Using the T^H typology, the ranking and magnitude of average allocations is consistent with the classification based on the second stage strategies. Own-maximisers contribute the least (3.20 tokens), followed by weak conditional cooperators (8.23), strong conditional cooperators (10.04), various (11.42) and unconditional cooperators (13.96). First stage contributions are significantly different across the five types. The Bonferroni multiple-comparison tests show the mean allocation of own-maximisers is significantly lower than all other clusters (all $p \leq 0.001$). There is a significant difference in allocation between weak conditional cooperators and strong conditional cooperators ($p = 0.088$), but no significant differences between the strong conditional cooperators and various, nor between the various and unconditional cooperators (all other comparisons are significant, $p < 0.011$).

4 Summary

We propose hierarchical cluster analysis as a useful tool for generating classifications of behavioural types bottom-up from experimental data. In VCGs using the P-experiment protocol, we confirm that free-riders and strong conditional cooperators (matching the contributions of others one-to-one) emerge as the cores of clearly-distinguished behavioural groups. In part because cluster analysis considers both within-group similarity as well as across-group dissimilarity, weak conditional cooperation also emerges as a distinct mode of behaviour. This provides an independent justification for a similar distinction among types of conditional cooperator which has been proposed in several previous studies, including Chaudhuri and Paichayontvijit (2006), Rustagi et al. (2010), Gächter et al. (2012) and Cheung (2014).

With the addition of the small group of unconditional contributors, these four types comprise approximately 90% of all participants across the six studies; excluding Herrmann and Thöni (2009) this figure rises to 95%. Notwithstanding the noise inherent in choice data, most participants' behaviour is reasonably well-approximated by one of four stereotypical patterns of behaviour.

Our data-driven approach shares some similarities with Houser et al. (2004), who classify the behaviour of participants in a dynamic decision environment into a small number of distinct types based on Bayesian classification. The output of their method is probabilistic, which is well-suited to their setting. Their participants make multiple, sequential decisions, which indicates the introduction of the noisy shocks on each decision which appear in their model. In the P-experiment protocol, we observe only one contribution strategy for a participant, in which all components are set as part of the same decision. Using the expected difference in contributions as a distance metric is a simple solution that does not require us to impose a noise structure on the setting of the entire $G + 1$-dimensional contribution strategy.

We propose clustering approaches as being complementary to methods based in whole or in part on a priori theoretical considerations. Participants in decision-making experiments may behave in surprising ways not anticipated by existing theories. Methods such as hierarchical cluster analysis allow a look at the data without interposing the prism of a theory of how or why agents make decisions. This offers a practical cross-check that can uncover interesting patterns in the data which might not be salient to a researcher working entirely within a given theoretical framework.

References

Maxwell N. Burton-Chellew, Claire El Mouden, and Stuart A. West. Conditional cooperation and confusion in public-goods experiments. *Proceedings of the National Academy of Sciences*, 113(5):1291–1296, 2016.

Edward J. Cartwright and Denise Lovett. Conditional cooperation and the marginal per capita return in public good games. *Games*, 5(4):234–256, 2014.

Ananish Chaudhuri and Tirnud Paichayontvijit. Conditional cooperation and voluntary contributions to a public good. *Economics Bulletin*, 3(8):1–14, 2006.

Stephen L. Cheung. New insights into conditional cooperation and punishment from a strategy method experiment. *Experimental Economics*, 17(1):129–153, 2014.

Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis*, volume 3. John Wiley & Sons, New York, 1973.

Urs Fischbacher and Simon Gächter. Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, 100(1):541–556, 2010.

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

Urs Fischbacher, Simon Gächter, and Simone Quercia. The behavioral validity of the strategy method in public good experiments. *Journal of Economic Psychology*, 33(4):897–913, 2012.

Simon Gächter, Daniele Nosenzo, Elke Renner, and Martin Sefton. Who makes a good leader? Cooperativeness, optimism, and leading-by-example. *Economic Inquiry*, 50(4):953–967, 2012.

Benedikt Herrmann and Christian Thöni. Measuring conditional cooperation: A replication study in Russia. *Experimental Economics*, 12(1):87–92, 2009.

Daniel Houser, Michael Keane, and Kevin McCabe. Behavior in a dynamic decision problem: An analysis of experimental evidence using a bayesian type classification algorithm. *Econometrica*, pages 781–822, 2004.

Martin G. Kocher, Todd Cherry, Stephan Kroll, Robert J. Netzer, and Matthias Sutter. Conditional cooperation on three continents. *Economics Letters*, 101(3):175–178, 2008.

Robert Kurzban and Daniel Houser. Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. *Proceedings of the National Academy of Sciences*, 102(5):1803–1807, 2005.

John Ledyard. Public goods: A survey of experimental research. In John H. Kagel and Alvin E. Roth, editors, *Handbook of Experimental Economics*. Princeton University Press, Princeton NJ, 1997.

Raphaële Préget, Phu Nguyen-Van, and Marc Willinger. Who are the voluntary leaders? Experimental evidence from a sequential contribution game. *Theory and Decision*, pages 1–19, 2016.

Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.

Devesh Rustagi, Stefanie Engel, and Michael Kosfeld. Conditional cooperation and costly monitoring explain success in forest commons management. *Science*, 330(6006):961–965, 2010.

Reinhard Selten. Die strategiemethode zur erforschung des eingeschränkt rationalen verhaltens im rahmen eines oligopolexperiments. In H. Sauerman, editor, *Beiträge zur Experimentellen Wirtschaftsforschung*. JCB Mohr, Tübingen, 1967.

Dale O. Stahl and Paul W. Wilson. On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1):218–254, 1995.

Joe H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.

A Appendix

A.1 Data description

The studies included in the main analysis were selected from the a larger set of laboratory experiments that met the following criteria:

1. P-experiment (Fischbacher et al., 2001) protocol already published in a peer-reviewed journal as of September 2016.

2. Participants played the VCG game in groups of 4;

3. Participants were endowed with 20 tokens;

4. Marginal per capita return of the public good (MPCR) equal to 0.4 points per token.

We identified a total of 9 studies satisfying these criteria. Six author teams were able to provide us with their data, all of which is reported in our study. For the remaining studies, we either received no response, or the authors were not able to find the data from the paper.

A.2 Silhouette analysis

Let $\mathcal{N}$ be the set of participants, and $\mathcal{T}$ be the set of types into which participants are classified. For any participant i, let c^i denote that participant's strategy, where c^i_k is the contribution of participant i in the contingency where the average contribution of other members of her group is k. The distance between two strategies c^i and c^j is the Manhattan distance, $d(c^i, c^j) = \sum_{k=0}^{20} \left| c^i_k - c^j_k \right|$. Let $T(i) \in \mathcal{T}$ be the type of participant i. We define the average distance from participant i's strategy to the strategies of other participants of a given type $t \in \mathcal{T}$ as

$$a(i,t) = \frac{\sum_{j \neq i : T(j)=t} d(c^i, c^j)}{\sum_{j \neq i : T(j)=t} 1}.$$

For participant i, the distance to the "closest" type which is not $T(i)$ is

$$b(i) = \min_{t \neq T(i)} a(i,t).$$

The participant's silhouette index (Rousseeuw, 1987) is then defined as

$$s(i) = \frac{b(i) - a(i, T(i))}{\max\{b(i), a(i, T(i))\}}.$$

The resulting $s(i) \in [-1, +1]$. Values greater than zero indicate that the members of i's type are closer, on average, than the members of the next closest cluster. Rousseeuw (1987) proposes the silhouette diagram (see Figure 3) to visualise the coherence of clusters. The silhouette is generated by sorting members of each type in decreasing order by $s(i)$ and plotting $s(i)$ against the participant's sorted rank.