Document de travail du LEM / Discussion paper LEM
2018- 01

Group formation and cooperation in social dilemmas: A survey and meta-analytic evidence

Andrea GUIDO
LEM UMR 9221 - Université Catholique de Lille - Anthropo-Lab / andrea.guido@univ-catholille.fr

Andrea ROBBETT
Middlebury College, Department of Economics / arobbett@middlebury.edu

Rustam ROMANIUC
LEM UMR 9221 - Université Catholique de Lille - Anthropo-Lab / rustam.romaniuc@univ-catholille.fr

http://lem.cnrs.fr/IMG/pdf/dp2018-01.pdf

http://lem.cnrs.fr/

Les documents de travail du LEM ont pour but d'assurer une diffusion rapide et informelle des résultats des chercheurs du LEM. Leur contenu, y compris les opinions exprimées, n'engagent que les auteurs. En aucune manière le LEM ni les institutions qui le composent ne sont responsables du contenu des documents de travail du LEM. Les lecteurs intéressés sont invités à contacter directement les auteurs avec leurs critiques et leurs suggestions.

The goal of the LEM Discussion Paper series is to promote a quick and informal dissemination of research in progress of LEM members. Their content, including any opinions expressed, remains the sole responsibility of the authors. Neither LEM nor its partner institutions can be held responsible for the content of these LEM Discussion Papers. Interested readers are requested to contact directly the authors with criticisms and suggestions.

Group formation and cooperation in social dilemmas: A survey and meta-analytic evidence[*]

Andrea Guido[†], Andrea Robbett[‡] & Rustam Romaniuc[§]

October 30, 2017

Abstract

In the last two decades, many laboratory experiments have tested the hypothesis that groups that are composed of "like-minded" subjects maintain higher cooperation levels than randomly formed groups. We survey the growing literature on group formation in the context of three types of social dilemma games: public goods games, common pool resources, and the prisoner's dilemma. The 62 selected papers study the effect of different sorting mechanisms – endogenous, endogenous with the option to play the game, and exogenous – on cooperation rates. In games with an endogenous sorting, the introduction of costly signals about one's type prevents cooperation to unravel over time by allowing cooperators to stay away from free-riders. The introduction of the option to play the game also achieves a separating equilibrium but its success depends on the attractiveness of the outside option. When sorting is exogenous, cooperation is higher with common knowledge that the group is composed of like-minded individuals. Additionally, the dimension on which people are sorted plays a fundamental role in how successful groups are in sustaining cooperation. To compare the effects of endogenous and exogenous sorting on cooperation, we conduct a meta-analysis with 431 group-level observations. We find that cooperation rates are higher with endogenous sorting than with exogenous and that punishment is effective in the former but not in the latter.

[*] For valuable insights at various stages of this work, we wish to thank Rachel Croson, Roman Sheremeta and the participants at the IMEBESS meeting in Barcelona and at the ASFEE meeting in Rennes. We are also grateful to the authors that responded to our various requests on the ESA forum.

[†] Université Catholique de Lille, Anthropo-Lab and LEM-UMR CNRS. Contact: *andrea.guido@univ-catholille.fr*

[‡] Middlebury College, Department of Economics. Contact: *arobbett@middlebury.edu*

[§] Université Catholique de Lille, Anthropo-Lab and LEM-UMR CNRS. Contact: *rustam.romaniuc@univ-catholille.fr*

1 Introduction

There is overwhelming evidence from laboratory and field experiments that many people cooperate even when this is individually costly and does not necessarily lead to a better treatment by a third party. However, an equally frequent observation is that cooperation declines over time because some people switch to free-riding when they observe selfish behavior in their group. In other words, research in experimental economics demonstrates that the co-existence of "conditional cooperators" and "free-riders" leads to socially undesirable outcomes over time (e.g. Ehrhart and Keser [1999], Keser and van Winden [2000], Burlando and Guala [2005], Gächter and Thöni [2005], de Oliveira et al. [2015]). Thus, the experimental literature suggests that cooperation is fragile if groups or communities are composed of people who vary in their cooperative attitudes.

In the last two decades or so, many experiments manipulated the way groups are formed in order to find a solution to the decay of cooperation over time. Chaudhuri [2011] notes that group formation is one of the most effective non-punitive mechanisms to sustain cooperation. We survey the growing literature on group formation in the context of three types of social dilemma games: public goods games, common pool resources, and the prisoner's dilemma. The 62 selected papers implement three types of sorting mechanisms:

1. endogenous sorting: all subjects play the social dilemma game where subjects have some ability to form or leave groups on their own accord;

2. endogenous sorting with the option to play the game: subjects can exit a specific relationship or avoid the game entirely;

3. exogenous sorting: sorting is undertaken by the experimenter with or without the subjects' knowledge.

For each sorting mechanism, we highlight the main design features and findings. The bottom line from the surveyed literature is that the presence of subjects with pro-social motivations is not sufficient for groups to sustain cooperation. With the endogenous sorting, cooperative behaviors can survive over time only if the freedom to form or leave groups is coupled with a pre-determined rule that allows conditional cooperators to stay away from free-riders. In games where participants can opt not to play the game, there is a selection effect: cooperators are more likely to opt-in to a social dilemma and have higher expectations about the likelihood that others will cooperate. The downside is that those who opt-out do not learn to cooperate. When sorting is exogenously implemented, it turns out that cooperation is higher when there is common knowledge about the cooperative attitudes of the group mates. Additionally, the method of measuring people's cooperative attitudes (i.e., the dimension on which they are sorted) plays a fundamental role in how successful groups are in sustaining cooperation. Some methods achieve a higher degree of homogeneity than others.

To further explore the differences in the sorting mechanisms and whether the differences can be explained by some of the environmental and design variables, we conducted a meta-analysis on 11 studies involving 18 treatments and 431 group-level observations. We find that on average cooperation levels are higher in experiments with an endogenous sorting mechanism than in the experiments with an exogenous sorting. Given that with endogenous sorting cooperation is bound to be fragile – cooperators are chased by free-riders – we find that monetary punishment is highly effective in increasing cooperation in these experiments but does not affect contribution decisions when sorting is exogenous.

The remainder of the paper is organized as follows: section 2 presents the endogenous sorting literature where all subjects play the social dilemma game, section 3 adds the literature with an option to play the game, section 4 presents the exogenous sorting papers, in section 5 we present the selection criteria and the main results for the meta-study, and section 6 concludes.

2 Endogenous sorting

Chaudhuri [2011] defines endogenous sorting as the mechanism that allows subjects to leave or form groups on their own accord. The endogenous sorting mechanism has been studied in more than 40 laboratory experiments. The objective of this section is to identify the different ways through which endogenous sorting mechanisms are implemented and their effects on cooperation.

We will begin by presenting the simplest mechanism that allows subjects to freely enter and exit groups – hereafter, *free migration*. This mechanism is plagued by a persistent, fundamental problem: conditional cooperators cannot dodge defectors and consequently cooperation unravels over time. Several means have been proposed to prevent free-riders from entering groups. Some settings allow genuine cooperators to send costly *signals* of their willingness to cooperate. Others study the effects of decentralized enforcement mechanisms such as monetary punishment, ostracism, entry-exit restrictions, or a combination of them. In the following, we provide an overview of all these mechanisms.

2.1 Free migration

In this category, we include all the experimental studies where subjects are free to migrate across groups. These designs are rooted in Charles Tiebout's canonical local public finance model Tiebout [1956], which proposed that residents will "vote with their feet" in response to differences in local public good provision and move to the communities where the public good and other local features best matched their preferences. This model was initially proposed as a solution to the demand revelation problem, since residents who move in response to the bundles of taxes and public goods in each community would reveal their true preferences for provision and could then be taxed according to their demand. However, the model has subsequently been applied to problems far beyond public finance to capture the fundamental idea that people will *sort themselves* into the communities, organizations, or firms that best suit them. Ehrhart and Keser [1999] conducted the first experiment testing Tiebout's hypothesis. In their experiment, subjects are allowed to freely move or create new groups.[1] The game consists of a 30-round linear public goods game with the marginal per capita return (MPCR) from the public account decreasing in the group size.[2] The declining MPCR implies that the presence of free-riders imposes a cost on the other group member. In each session, 18 subjects play two different independent games of 9 subjects each.

The authors find that the composition of groups is not *stable* and dynamics are *non-monotonic*: highly cooperative groups in the current period tend to grow in following round, while those characterized by low contributions shrink. However, when the size of the group remains unchanged or increases in the actual round, contribution levels unravel afterwards. This unstable dynamic is due to cooperative subjects trying to flee from defectors who in turn try to catch up with them to benefit from high contributions.

Robbett [2016] studies the dynamics of community formation when members have different marginal per capita returns from the public good and also varies whether these returns decline over group size. The population is split into two different types: High Types who benefit more from the provision of the good and may experience congestion and Low Types who have a low marginal per capita return.[3] Six groups or locations are randomly formed at the beginning of the experiment and subjects can move across them.

[1] Moving across groups was costly. Subjects were endowed with 10 Experimental Currency Units (ECUs) and paid 5 ECUs for creating or switching groups.

[2] In the standard linear public goods game, each participants is endowed with tokens or currency that they can allocate between a private account (which benefits on the individual) and a public account, which benefits all group members. The marginal return from the private account is typically 1, while the per-person return from the public account, the MPCR, is less than one.

[3] Agents are aware of the heterogeneity but not of the distribution.

Two different games are implemented: a congestible public goods (CPG) game, in which the presence of free-riders reduces High Types' payoffs (as in Ehrhart and Keser [1999]) and a pure public goods (PPG) game, where the most efficient outcome is for all participants to form a single group and High Types have no financial incentive to move to avoid the Low Types. The results show that preference heterogeneity is a relevant factor to understand players movements and contributions. High Types are pioneers who are generally the first to enter empty communities and they contribute high amounts over time. Low Types enter highly cooperative groups later, contribute less, and provision declines. While there is a substantial difference in the movement patterns between the CPG and PPG games, the congestion drives only about half of the movement. High Types continue to flee Low Types even when their presence does not lower their financial payoff, suggesting that much of the chasing pattern observed by Ehrhart and Keser [1999] and in the CPG game is driven by an intrinsic preference for avoiding free-riders.

Signaling These papers therefore suggest that cooperation may only be preserved from unraveling by screening and separating defectors from conditionally cooperative individuals. The introduction of a costly signal, for example a higher group entry fee, could allow pro-social agents to be recognized as such by others and self-select into groups.

Brekke et al. [2011] aim at testing whether social commitments, such as charity donations, could serve as a costly screening device. In their 3-person public goods game, subjects can choose to adhere to two group types, which differ in their payoff schemes. In the blue group, each member receives an extra fixed payoff while in the red the extra sum is donated to the Red Cross. The experiment is divided into three parts: part I consists of a one-shot public goods game with groups formed randomly;[4] in part II subjects play 10 rounds of the public goods game choosing beforehand their preferred group type; part III is the same as part II with the difference that players can change groups in each round and the number of rounds is increased to 20 periods.[5] The authors expect more cooperative subjects to join the red group since socially beneficial commitments can serve as costly screening devices to help pro-social subjects meet each other.

Results show that higher contributors in the one-shot game choose the red group while those who contribute less prefer the blue one. This correlation between willingness to cooperate and choice of group demonstrates the validity of the signal. As a consequence

[4] The subjects were informed that the experiment consists of 3 parts and that choices made in part I will not affect their earnings in the two other parts of the experiment.

[5] In part II and III, subjects are randomly chosen to create groups, implementing partner and stranger design, respectively. What they can choose is the pool of subjects from which their partners can be drawn, i.e. red or blue.

of the successful sorting, cooperation levels in red groups are higher and stable over time compared to those of blue groups, which present the usual decreasing trend.

Furthermore, when Brekke et al. [2011] run a placebo treatment with no mention of the Red Cross, few subjects, if any, chose the red group and cooperation unraveled as there was no separation between types. The validity of the signal seems strictly related to its *social meaning*.

Other experiments have demonstrated the effectiveness of endogenous mechanisms without reliance on social meaning. For instance, Aimone et al. [2013] design a laboratory experiment, absent any group identity or doctrinal construct.[6]. In their experimental setting, subjects are endogenously sorted by revealing their willingness to relinquish a fraction of their private return from the public good. The experiment considers two one-shot games: a standard and a "sacrifice" public goods game. The latter differs from the former in that subjects can express their preferences regarding the group's private return by choosing one value within [0.55, 0.95] in 0.05 increments, while it is fixed at 1 in the standard game. After all subjects express their preferences, the four subjects with the highest private return are grouped together, the four subjects with the next highest private return are placed together in another group, and so on. The private return chosen in each group is the average number chosen by the 4 group members. Two treatment orderings are used: the normal public good game in the first round followed by the "sacrifice" in the second (Inexperienced treatment) and vice versa (Experienced treatment). In both orderings, a third round of the Sacrifice game is played. Subjects are aware of the assortative mechanism and unaware whether there would be a subsequent round in each stage.

Results show that more cooperative individuals choose a lower private return (higher unproductive cost), screening out defectors, regardless of the game-play order. To further investigate the causes of the sorting mechanism's success, three control sessions with an exogenous (random) grouping mechanism were run. From this comparison, it is clear that the successful provision of the public good is driven by the possibility of shunning defectors by signaling subjects willingness to cooperate, which is impossible in random exogenous mechanisms.

Signaling mechanisms have been tested even under uncertain conditions regarding their validity. Grimm and Mengel [2009] implement a prisoner's dilemma game introducing the concept of *viscosity*: an increased probability of interacting with others of one's type or group. Subjects can choose to play in groups that differ in the defector gain of the

[6] The authors base their work on Iannaccone [1992] related to "sacrifice and stigma" mechanisms – i.e. unproductive costs, employed by religious groups to dodge free-riders.

associated payoff scheme. Choosing the group whose defector gain is lower represents a signal of one's cooperativeness. However, a high viscosity reduces the benefit of the signal as it is likely that subjects might be paired with others from the opposite group. Subjects play the game for 100 rounds, repeatedly choosing between two groups: group A, whose defector gain in the related payoff scheme is lower than that of group B. They conduct three treatments that differ in their viscosity levels and three control conditions (varying feedback and payoff given to subjects and ruling out the possibility of sorting). Prior to playing the game, subjects were informed about the percentage of subjects in groups A and B, and their individual probability of meeting members of each group.

The data suggest that the level of viscosity affects the sorting of subjects into different groups. In treatments characterized by high levels of viscosity, a large fraction of subjects choose group A (59.2% and 36.8%), while subjects opt for group B when viscosity is lower. The authors also show that viscosity positively affects the number of subjects cooperating in the prisoner's dilemma. Members in group A cooperate significantly more than those in group B and this difference becomes more marked as viscosity increases.

There is also experimental evidence showing that costly signals do not always work to screen out free-riders in the long run. For instance, the experiment reported in Robbett [2016] also varies the entry fees among the six available communities. Since the Low Types receive lower returns from public good provision, the entry fees are relatively higher for them and they should only wish to enter one of these communities if the provision is much higher than in their own. The high entry fees could thus serve as a mechanism to help High Types avoid the presence of Low Types. The results show that higher entry fees can initially promote separation between High and Low Types and average contributions are higher in these locations. However, these group experience the same declining provision level and are ultimately no more stable than the low entry fee locations.

To summarize, the common feature of the experiments where cooperative subjects succeed in separating themselves from defectors is that the dimension of the group is not endogenously determined. This cap impedes defectors' ability to easily reach highly cooperative groups.[7]

[7] The literature on endogenous group formation where subjects get payoff-relevant information about the others (or about other groups) includes Bayer [2016] who shows that allowing subjects to break-up a partnership based on the current and past history of the partner increases cooperation. Also, Coricelli et al. [2004] investigate the impact of introducing costly partner selection on the cooperation levels in a 2-person public goods game. Subjects are informed about each other's contributions and are asked to select their future interaction partner, employing two mechanisms: unidirectional and bidirectional. Unidirectional mechanism uses auctions to define pairs, while bidirectional mechanism uses an algorithm based on point assignments. They find that cooperation is higher under the unidirectional mechanism than under the bidirectional mechanism. Contributions are not affected by revealing individual contribution histories.

Beliefs, voting and market institutions Alternative solutions that do not directly rely on costly signals have been implemented to support the separation between subject types.

Cabrera et al. [2013] study the role of a promotion-demotion system in overcoming the free-rider problem. The between-subject design involves 24 subjects in a standard 10-rounds public goods game and 80 subjects in a treatment where the population is split into two groups or "leagues" – the *major* and the *minor* league. The highest contributor in the minor league is promoted to the major league, while the lowest contributor in the major is relegated to the minor league. The parameters and payoffs are otherwise identical in both the baseline and treatment. The perfect-Bayesian Nash equilibrium involves free-riding as long as beliefs are such that both leagues have the same expected payoff.

The authors find, however, that contribution levels in the major league are significantly higher than those in the minor and both perform better than the baseline. Contributions are therefore treatment dependent. The promotion-demotion system works as mechanism to separate free-riders and cooperators into minor and major leagues, respectively. Low contributors seldom enter the major league given the positive correlation between the number of rounds a subject is in the major league and her contributions. Despite the successful self-selection, the contribution dynamics are the same irrespective of the treatment. Cooperation unravels over time and even though free-riders are set apart in the minor league, subjects in the major league lower their contributions over time. The separation of subjects into different leagues seems to increase the overall cooperation levels but leaves unaffected the contribution dynamics.

Different results are shown in Bohnet and Kübler [2005], which reports the results of a two-person prisoners dilemma, in which the right to play a form of the game offering insurance against defection is auctioned off. Players play a one-shot prisoners dilemma for five rounds and are randomly rematched each round with new counterparts who have chosen the same version.[8] The authors vary the number of rights available and the timing of the auction[9]. Four treatments are conducted: a control where versions of the game are exogenously assigned; two treatments where the rights are auctioned off in every period, differing in the number of available rights; and a treatment where auctions take place in the first period. In all treatments, except the control, subjects are initially assigned game

[8] Versions of the game are labeled A and B. In the payoff of the former version, the "sucker payoff," which the player receives by cooperating when the other player defects, is higher – hence the reference to insurance against defection.

[9] The number of rights is chosen so as to be higher or lower than the expected number of conditional cooperators in each group. Thus, auction prices should be higher when group B is smaller as cooperators can easily meet each other.

A and can bid for game B afterwards.

In game B, cooperation levels are higher than those in game A in all treatments except the control. Cooperators especially tend to prefer game B when they were previously defected on in game A and, after playing game B, tend to bid higher prices to keep the right to remain in that version. The first period results suggest that most subjects sort themselves into games A and B according to their type. When the auction takes place only in the first period, however, contributions show a descending pattern as a consequence of an incomplete sorting that leads cooperation to decrease steeply in later rounds.

We have thus far seen that experiments with free-migration across groups are plagued by a "chasing" phenomenon, in which free-riders pursue cooperators across communities (Ehrhart and Keser [1999]). Additionally, cooperators often refuse to remain in a group where other people are free-riding on their contributions, even when the presence of free-riders doesn't harm their payoffs (Robbett [2016]). This indicates that group stability may rely on either a fixed group size (as highlighted in the previous subsection) or on institutions require all participants to contribute equally to the public good or that allow participants to endogenously enforce a certain behavior.

Robbett [2014] highlights the relevance of institutions in ceasing the chasing of cooperators by defectors. If instability is caused by subjects' unwillingness to remain in communities where others contribute less than they do, then stable group dynamic may be reached when agents contribute equally, for example under a *taxation system*. Subjects play a 20-round non-linear public goods game and, in each period, can choose where to reside among 6 communities. There are two types of agents, who differ in their optimal bundle of taxes and public good expenditures. The author designs four experimental conditions each representing a different institution for determining contributions. The first is a standard voluntary contributions mechanism, in which subjects can contribute however much they wish. The remaining three impose restrictions on agents' contribution levels. Under the setting *Fixed Tax*, each community is linked with a fixed, posted tax (i.e., public good contribution) that has to be paid in each round as long as the subject resides in that community. The provision of the public good depends, in turn, on the number of residents in that community. Under the experimental condition *Fixed Quantity*, each community was associated with a fixed, posted provision of the public good. Subjects therefore bear a per-capita tax in each community, depending on the number of its residents. Since preference ares induced, in both experimental settings the optimal bundle for each agent type is provided among locations. The fourth experimental setting, *Voting institution*, allows subjects to vote on the local tax policy in each period, with the median vote implemented.

Subjects tend to sort themselves into homogeneous communities under all institutions, except in the standard voluntary contributions mechanism where the usual “catch-up" problem still emerges. Despite the efficient sorting, subjects in the Fixed Tax and Fixed Quantity conditions, who can only “vote with their feet," often sort into communities with a suboptimal taxation policy, such that they either under-or over-provide the public good relative to the optimal tax-provision bundle for the population. Under the Voting institution subjects converge towards their own optimal consumption of public good by voting on their local taxation policy. In other words, by voting with their feet they are able to sort by their preference for the public good and by voting with their ballots they are able to ensure that the community they sort into enacts the optimal tax-provision policy. This evidence highlights the fact that pre-designed optimal policies do not perform better than an endogenous process for tailoring the policies of communities.Additional institutions for endogenously overcoming the “chasing” problem, such as peer punishment and group entry rules, are highlighted in the subsequent sections.

2.2 Monetary punishment

A way to prevent defectors from entering groups of cooperators is to give to the latter means to screen free-riders out. In the previous section, we have shown that signals can solve the separation problem to some extent. However, subjects can also individually *enforce* cooperative social norms. We distinguish three main enforcement mechanisms: monetary punishment, ostracism, and entry and exit restrictions.

Gürerk et al. [2014] paper is the first to investigate the role of punishment under free migration. The authors aim at addressing the following question: would a sanctioning institution deliberately be adopted when individuals can choose between a sanctioning and a sanctioning-free environment? In their experiment, subjects play a 30-round public goods game. Each round consists in three stages: an institution choice stage, a voluntary contribution stage and a sanctioning stage. In the first, subjects choose between joining a group with the presence of punishment and reward (SI) or one without (SFI). In the second stage, all the participants interact with subjects in their own group in a public goods game. Eventually, in the third stage, participants grouped in SI are asked to reward or sanction other members of the same group by assigning from -20 to 20 tokens to other participants. Each token negatively assigned induces a cost of 3 ECU to the punished subject and 1 ECU to the punisher. Conversely, in case of tokens positively assigned, the cost and the reward are symmetrically set at 1 ECU. Subjects receive feedback regarding other participants' individual contributions from the same group as well as from the other groups.

One third of the population chooses SI in the first round. The initial choice of the institution correlates with different types of behavior. More than half of the members in the group that choose SI in the very first round are classified as "high-contributors". Three-quarters of them use tokens to punish and establish norms of cooperation. This fraction of subjects can be classified as "strong reciprocators". Only 5% of subjects sorted in SI group is classified as free-riders (namely contributions lower than 5 ECUs), which is not surprising because opting for a sanction and not contributing would be an example of shooting oneself in the foot. This fraction hikes up to 44% in the SFI group, revealing free-riders' preference for a sanction-free environment.

Free-riders in the SFI environment initially earn more and therefore many subjects join this group starting from the second round. As cooperation unravels in the immediate periods, subjects find it optimal to migrate towards SI where payoffs are higher. Subjects switching from SFI to SI significantly increase their contributions and even former free-riders become full cooperators. Punishment becomes an accepted social norm in SI and subjects use it only occasionally over time.[10]

In a follow-up paper, Gürerk [2013] investigates whether *social learning* can increase the willingness to accept punishment institutions. Subjects' awareness of the "long-term" beneficial effects of punishment may foster general acceptance. The experimental design is similar to Gürerk et al. [2006] with the difference that rewarding is not considered. Three treatments are implemented: a baseline game where subjects receive no information regarding previous sessions, a treatment (SHT) where participants receive a report about the decisions done by participants of a previous experiment (see Gurerk et al. [2010]) and a treatment (SHT-Half) where subjects are provided with only a subset of the social history. The social history provided consists of the average number of community members, their contributions, the received punishment tokens, and the payoff for each round in a treatment with punishment implemented by Gurerk et al. [2010]. In the SHT-Half treatment, only the history of the institutional choice is provided to subjects.

It turns out that social history increases the initial acceptance of the punishment institution and achieves full participation in the community with punishment. Contributions

[10] Rewards are not perceived as encouragement to increase contributions, as they target those who already abide by the social norms. Gürerk and co-authors disentangle this combined effect in a follow-up work. The authors separate the combination of reward and punishment in two different treatments.

steadily increase when social history is provided and stabilize near the social optimum. This in turn lowers punishment expenses and reduces the initial inefficiency loss seen in Gürerk et al. [2006]. Furthermore, the significant difference between SHT and SHT-Half leads to the conclusion that subjects do not merely imitate the institutional choice made by other subjects but also pay attention to other relevant factors.

2.3 Non-monetary punishment

A different form of punishment, other than directly reducing the payoffs of fellow group members, is "ostracism" from the group. Cinyabuguma et al. [2005] study the extent of cooperation in a public goods game under the threat of ostracism. In each session, groups of 16 subjects are endowed with 10 ECUs and play a series of 15-round public goods game. The experiment has two treatments: the *baseline* treatment consists in a standard public goods game, while in the *Expulsion* treatment subjects can vote to expel members of their own group after being informed of each others' past contributions. The expulsion is implemented only if the majority votes in favor. The ostracized subjects are banished to another group (called the "Blue group") playing the same public goods game with a reduced endowment of 5 ECUs for the remaining rounds. Each subject voting in favor of expulsion faces a cost of 25 ECUs in case of a majority vote to expel the targeted subject. The authors implement two conditions: BE where subjects play the Expulsion treatment after the Baseline, and EE where subjects play the Expulsion twice.

The authors find that expulsion is used sporadically, on average between one and four times per session in each treatment. The related dynamic fluctuates over time: it peaks in the first round where ostracism is used to initially discipline defectors and in later rounds due to the end-game effect when cooperation unravels. Votes are targeted towards defectors. Targeted players immediately raise their contribution levels after having faced the threat of expulsion.[11] However forms of antisocial punishment also emerge over time, with high contributors being the target of defectors' votes. When ostracism is possible, contributions are higher starting from the first round because subjects anticipate this punishment, and are stable over time until the end game effect in the final rounds. In the EE condition, contributions are higher the second time the expulsion treatment is played than the first time.

Maier-Rigaud et al. [2010] find similar results in an experimental setting involving

[11] Masclet [2003] finds similar results. In his public goods experiment, subjects exclude their peers for two reasons. Subjects are willing to punish unfair behaviors and expect behavioral changes in responses to exclusions.

groups of 6 subjects playing a 10-period public goods game with a partner matching design. They differ from Cinyabuguma et al. [2005] in that they don't allow ostracized players to play in a group of likewise expelled players. (Instead, they are simply excluded from all group activities, earning their endowment in each round). They also implement smaller group sizes and consequently a higher MPCR, and the vote to expel other group members is costless. Two treatments characterize the design of this experiment:[12] a *baseline* consisting in a simple public goods game, and a treatment with the presence of ostracism.[13]

Ostracism has a positive effect on contributions, particularly if included in the first treatment of the session.[14] However, despite the differences in the experimental settings, results are similar to Cinyabuguma et al. [2005]. Average contributions in the baseline and the ostracism treatment significantly differ in all periods, apart from the first rounds when the social norm of cooperation has not been well established yet, and last rounds of each treatment, with the end-game effect.[15]

In some experimental settings, participants can not only ostracize group members but also freely migrate across groups and/or be forgiven. In Charness and Yang [2014] experiment, 9 subjects are randomly assigned to 3 groups and play a 3-player public goods game for three periods. At the end of it, subjects learn about IDs and individual contributions of their own group members. The authors then allow subjects to exit their groups, to exclude other members and to merge already existing groups in three separated stages for 15 rounds. An additional 15-period segment with new player IDs is then played. The experimental design considers three treatments. In the "Main" treatment, the MPCR is decreasing over group size but the total social benefit of contributing (i.e., MPCR times group size) is increasing, such that the social optimum is for all nine members to form a single group. The "Capped Efficiency" efficiency is identical to Main with the only difference that the social benefit to of contributions is capped after the group size is equal or higher than four, such that there is no efficiency gain to forming groups greater than size four. In the Baseline treatment, subjects are randomly assigned to fixed groups (3-person, 6-person and 9-person groups) and no regrouping is possible.

They find that contribution levels over the 15 rounds are higher in Main and Capped Efficiency than in the control treatment Baseline. In the Main treatment, subjects manage

[12] Participants received information about the structure of the current treatment only.

[13] Authors control for the order of implementation by switching the order of treatments. Non-parametric tests show that the order of implementation matters in contribution levels, apart from the last periods.

[14] Average contributions increase to around 85% of the initial endowment when ostracism is implemented as the first treatment, and around to 80% when second.

[15] Indeed, authors show that subjects cast votes mainly at the beginning and the end of each treatment so as to tame defectors.

to create grand coalitions, while under Capped Efficiency group size is on average four, consistent with the cap on groups' return. Group dynamics tend to be more stable in the Main treatment than in the Capped Efficiency treatment: subjects are more likely to exit and expel other members in the latter than the in the former treatment. The novelty of this experiment is the presence of redemption. Redemption gives the possibility for individuals who have made early mistakes to later join successful groups and to become highly productive members.

Another contribution that focuses on the role of ostracism and competition under free migration is Sääksvuori [2014]. The author disentangles the effect of group competition from the threat of ostracism. The between-subject experimental design involves 4 different treatments. In each of them, participants play a 20-period public goods game, being randomly sorted in 12-person societies. They are endowed with 20 ECUs and the MPCR decreases in the group size. Treatments involve a chance to exit the current group and become a free-agent, enabling the person (1) to apply for a membership in any other remaining group, (2) to create a new group jointly with other free-agents, or (3) to stay out of any group arrangement. The baseline treatment consists of a standard public goods game. Whenever there is ostracism, subjects can expel other group members by casting a majority vote. The author also crosses ostracism with a competition between two groups.

In general, ostracism enhances cooperation levels, a result congruent with the findings from the other studies. In particular, the group size is strongly affected by the presence of ostracism. Whenever ostracism is allowed, the group size is larger and more stable over time than without ostracism.

While the expulsion of group members is costly, preventing free-riders from entering the group in the first place is another way to sustain high cooperation levels. Ahn et al. [2008] and Ahn et al. [2009], restricted entry with free exit and free entry with restricted exit. The game consists in a 20-period non-linear public goods game, with a dominant strategy of contributing 3. Each period begins with subjects being asked if they wish to change groups before making any decision regarding their contributions for the provision of the public good. They are informed about the aggregate contribution at the group level in the previous 5 rounds, and the number of subjects in each group who had chosen to remain in the group from the previous period. The between-subjects experimental design includes three different treatments that differ for in entry and exit rules.

In the "Free Entry/Exit" treatment, subjects are free to migrate among groups without any restriction as in Ehrhart and Keser [1999] and Robbett [2016]. In the "Restricted

Entry" treatment, exit is unrestricted but entering a new group is conditioned to the majority approval of its members. On the other hand, in the "Restricted Exit" treatment, entering is free, but exiting a group is conditional to the approval of groups' members. In both cases, those voting can see the past five contributions of the subject attempting to enter or exit the group.

In the first study Ahn et al. [2008], the public good was pure and subjects therefore have an incentive to form the grand coalition with all twelve participants in a single group. They find that the "Restricted Entry" treatment has the highest levels of contributions and the group size is smaller than in the other treatments. The restriction on entry decisively teaches applicants to increase their contributions. Once a low-contributor is rejected by members of another group, he/she would raise her contributions until he/she gets accepted. A restriction on exit instead is detrimental for cooperation as high-contributors are prevented from exiting groups of defectors. In this case, cooperation unravels from the early rounds as subjects denied exit retaliate lowering their contribution levels. The higher one's past contributions, the greater the likelihood of being accepted. The opposite holds when the subject is attempting to exit a group.

However, since the restricted entry mechanism promoted small, exclusive groups, even though they also achieved high average contributions, their members did not benefit from the positive externalities that come from being in a large group with a pure public good. Thus, despite being successful in promoting cooperation, the restricted entry mechanism also had the effect of reducing the average earnings of the subjects in comparison to the Free Entry/Exit and the Restricted Exit conditions. In a follow-up study, Ahn et al. [2009] incorporate congestion and thus reduce the incentive of forming large groups. They find that Restricted Entry raises average contributions and total earnings, eliminates the earnings disadvantage of contributors in comparison to free-riders, and reduces the level of congestion within the groups in comparison to the Restricted Exit and Free Entry/Exit conditions (where congestion is defined as surplus members in a group beyond the optimal membership level, given current contributions).

To summarize, we have seen that free-migration is plagued by a persistent, fundamental problem: conditional cooperators cannot dodge defectors and consequently cooperation unravels over time (Ehrhart and Keser [1999]). The introduction of costly signals, such as a higher group entry free (Brekke et al. [2011], Aimone et al. [2013]), allows pro-social agents to be recognized as such by others. Costly signals also deter free-riders from chasing cooperators. Alternative solutions, such as auctioning out the right to play the game (Bohnet and Kübler [2005]), creating separating leagues (Cabrera et al. [2013]) or allowing subjects to vote on the adoption of different institutions (Robbett [2014]) may also improve the overall contributions rates to the provision of public goods. Agents can also signal their willingness to cooperate by voting for specific institutions that sanction free-riding. We have seen that individuals voluntarily switch over time from sanctioning-free environments to sanctioning institutions (as in Gürerk et al. [2006]). The adoption of sanctioning institutions particularly affects the behavior of conditional cooperators who move from free-riding to full cooperation. Another interesting observation is that allowing subjects to exclude group members (Cinyabuguma et al. [2005]) or prevent them from entering the group (Ahn et al. [2008]) also has a positive effect on cooperation levels. However, this mechanism promotes small groups, thus preventing their members from enjoying the benefits from large communities.

3 Voluntary association and the option to play the game

A closely related strand of the endogenous group formation literature considers situations in which participants can exit a specific relationship or avoid the game entirely. In such scenarios, agents may be able to opt-out of the social dilemma permanently (instead accepting payoffs that do not depend on the actions of others), "sit out" temporarily and then re-enter to continue interacting with the same partner(s), or exit the relationship and be re-matched with new partners.

The earliest experiment that we are aware of concerning this type of voluntary interaction is reported in Orbell et al. [1984].[16] They consider an n-player prisoner's dilemma game in which each of nine players chose whether to cooperate or defect, with each individual's payoffs strictly increasing in the number of cooperators and, conditional on the others' behavior, "defect" paying substantially more than "cooperate." After making their decision, each player chose whether to be paid based on the outcome of the game or to receive a stochastic payoff that did not depend on the behavior of the participants. They hypothesized that cooperators would be more inclined to exit than defectors, given that the expected outside option payoff is relatively more attractive to those choosing "cooper-

[16] A precursor to this experiment is Miller and Holmes [1975], which introduced an "Expanded Prisoner's Dilemma" game that includes a third possible action for the players. This additional "defensive" or "withdrawal" option is a best response to one's partner defecting, but does not also reduce a cooperative partner's payoff. Although not a pure exit option, since payoffs to those electing to withdraw vary slightly depending on whether one's partner also withdrew, this expansion provides subjects with an option beyond simple cooperation or defection. They found that would-be cooperators are more likely to choose the withdrawal option than revert to defecting when facing an uncooperative partner. This experiment was not conducted under standard experimental economics procedures (participants actually played against a computer program rather than an actual partner and do not appear to have been incentivized).

ate" (who always earn lower payoffs from staying than defectors). In contrast, they find that those choosing the cooperative option were weakly less likely to exit. Though they find that cooperators have higher expectations regarding the cooperation of others, which could make staying more attractive, there is no evidence that this explains the difference in exit and they instead conclude that cooperators are simply more willing to be part of a group.

Orbell and Dawes [1993] consider one-shot, two-player prisoner's dilemma games and directly compare games in which participation is required to games in which participants could opt-out for a zero payoff – which was more attractive than the mutual defection outcome. They find that participants earned significantly more when participation was voluntary. Average cooperation was higher, conditional on playing the game, in the voluntary condition, as defectors tended to assume that their partners would defect as well and thus were more likely than cooperators to choose not to play.

While these games were one-shot, Hauk [2003] considers a 10-period repeated prisoner's dilemma game in which participants can choose in each period whether to exit, defect, or cooperate. This experiment replicates the Orbell and Dawes [1993] finding that cooperation is higher when association is voluntary; the difference is that participants have knowledge of their prospective partners' history, and thus exit can be used to avoid future interactions with partners who have previously defected.

In the experiments discussed thus far, the payoff from exiting is typically higher than that from full defection, making universal non-entry and defection the subgame perfect equilibrium. Therefore, anyone who does enter the game, either with cooperative or competitive intentions, must do so expecting some degree of cooperation: an expectation consistently shown to be more prevalent among cooperators. When the outside option is worse than mutual defection, however, everyone will prefer to enter and thus exit provides a severe punishment for an uncooperative partner. This is explored in more detail in a third treatment reported by Hauk [2003], in which the mutual defection outcome is more attractive than not playing the game. While she finds evidence that subjects use the unattractive outside option as punishment for defection, cooperation is far lower than when participation was mandatory or the outside option was attractive, likely in part due to the fact that payoffs from (Defect, Defect) were positive in this treatment and negative in the others.

Hauk and Nagel [2001] further expand the investigation of exit with an attractive outside option. They consider scenarios in which participation does not require mutual agreement by both partners but, rather, one player can unilaterally force the game to be played – a setup that supplies data on how those who prefer to exit will play the game when required. They find that defection rates are lowest when participants can opt-out, such that, conditional on being matched, cooperation rates are high. At the same time, however, participants who are involuntarily forced to play the game frequently end up cooperating eventually. As a result, they suggest that unilateral matching is most favorable for cooperation rates: mutual defectors opt-out while many of those who are forced to play by their partner eventually learn to cooperate and can be "reformed."

Keser and Montmarquette [2011] consider a chosen-effort team production experiment with convex effort costs, in which participants are either required to participate in teams of two (where they face a social dilemma) or are paid only based on their own effort unless both players agree to form a team for the period. As in the "attractive outside option" experiments described above, participants should not voluntarily form teams if they anticipate players choosing their dominant strategy effort in the team remuneration setting – but opting into team production could be profitable if players are sufficiently cooperative. They find that players regularly opt-in to the team remuneration arrangement and, after voluntarily joining a team, exert effort close to the Pareto optimal levels – which is significantly greater than effort in the mandatory team treatment.

Boun My and Chalvignac [2010] conduct a five-person 20-period public goods game, in which participants can (temporarily) opt-out of the group in each period and instead receive an outside option payoff that is either just slightly above the per-person endowment or somewhat higher. They find little effect of voluntary participation on average contributions. However, they find some evidence that the decay in contributions over time is mitigated in the condition with a higher outside option, and they attribute this finding to a pattern in which higher contributors opt-out after observing free-riding, causing lower contributors to increase the contributions to attract the cooperators back to the group.

Nosenzo and Tufano [2017] study a two person, one-shot public goods game, in which they vary whether voluntary participation can influence contributions either through assortative matching or through threat of exit. They compare a baseline treatment, in which participation is mandatory, with treatments where participation occurs only by mutual agreement. In the unconditional treatment, participants first decide whether to participate or take an outside option payoff that is just slightly above the mutual free-riding payoff. In the conditional treatment, both players first decide on contributions and then decide whether to continue with the partnership (and receive the resulting payoffs) or to withdraw such that both are instead paid the outside option. They find a strong effect of the conditional participation treatment, with contributions more than doubling (even including those who don't participate) relative to the baseline. However, there is no difference between contributions in the baseline and unconditional treatments. To assess whether sorting was limited by the relatively low outside option – which caused most players to opt in – they conducted an additional robustness check treatment with a higher outside option, but still find no difference in contributions compared to baseline. Finally, they also use a sequential prisoner's dilemma game to elicit beliefs and cooperative preferences, and do observe a false consensus effect (with cooperators believing cooperation is more likely), but these beliefs do not seem to influence the choice to opt-in to the unconditional public goods game.

With respect to use of exit as a threat or punishment, Wilson and Wu [2017] study an infinitely repeated joint production task with imperfect monitoring, varying whether players can unilaterally and permanently terminate the game and the value of the outside option. They observe higher cooperation when termination is possible, regardless of the outside option value.

Yamagishi [1988] and Herbst et al. [2015] each investigate voluntary participation in a team work task and both find that stronger performers prefer to work as individuals rather than be compensated based on the combined efforts of a team. In the experiment reported in Yamagishi [1988], teams of three participated in a real effort (decoding) task for which they were paid an equal share of the team's output. Prior to each work period, team members could exit the group and instead be paid based on their own performance. As expected, the most productive team members were more likely to exit the group compensation scheme, even when there was a cost to doing so. Likewise, Herbst et al. [2015] find that participants who exert high effort in a chosen effort Tullock contest, either due to intrinsic competitiveness or lower induced effort costs, are less likely to join an alliance with another player. Despite this, they find that participants who voluntarily join alliances supply greater effort than those who were exogenously assigned to participate in an alliance, indicating a positive effect of opting-in to a team.

While the games discussed thus far only allowed the participant to avoid interaction with a fixed partner – but not acquire a new one – a theoretical literature on prisoner's dilemma simulations (beginning with Schuessler [1989] allows agents who sever their interactions with one partner to randomly select a different partner from the population. In an experiment that provides a link between this theoretical work on random re-matching and the experiments on exit discussed thus far, Barclay and Raihani [2015] study a prisoner's dilemma game with costly punishment, in which participants could either: (1) withdraw from the game for a single period (receiving a payoff below the mutual defection payoff) and then return to the same partner; (2) withdraw for a single period and then be randomly matched with a new partner upon their return; or (3) pay a cost to be immediately matched with a new partner. They find that cooperation is lower when the participant can only withdraw and not switch to a new partner, in which case they are more likely to respond to a defection by punishing than using the withdrawal option. When changing partners is possible, participants are equally likely to switch partners or punish their existing partner.[17]

To summarize, we have seen evidence that voluntary association, or the ability to exit, often promotes cooperation in social dilemmas among those who opt-in to the game or partnership. In games where participants can opt not to play (either at all or temporarily), the outside option is typically attractive relative to the Nash outcome of the social dilemma. Therefore, there is a selection effect: cooperators have higher expectations about the likelihood that others will cooperate and are more likely to opt-in to a social dilemma. The experiments reported by Orbell and Dawes [1993], Hauk [2003], Hauk and Nagel [2001], and Keser and Montmarquette [2011] all exhibit evidence of this type of selection or "forward-induction" argument. The downside is that those who opt-out do not learn to cooperate. There is additional indication that the positive effect of an exit option depends on the relative attractiveness of the outside option, although outside options that are less attractive than the unique Nash equilibrium have not yet been thoroughly studied. In cases where participants differ in their abilities or costs, there is considerable evidence that weaker players are more likely to opt-in to team remuneration schemes Yamagishi [1988]; Herbst et al. [2015]; Keser and Montmarquette [2011] also find indication of this, but the result is not significant at the 10% level; Hamilton et al. [2003] observe a similar phenomenon in a field study of a garment factory). Finally, we also observe preliminary evidence that voluntary association in itself can increase team effort Herbst et al. [2015] and that the option to exit could increase cooperation further if exiting agents have the opportunity to be re-matched with a new partner, at least when punishment is also available

[17] Boone and Macy (1998) report an experiment in which subjects play a prisoner's dilemma-like online card game against simulated partners, with some players given the option to exit a relationship and be randomly paired with a new partner at any time. (Participants were not paid their payoffs, but, rather, the top performers received a fixed cash prize.) They do not find a difference in cooperation compared to the no-exit condition, and in a follow-up study they propose that this is due to exit having an opposing effect on the behavior of cooperative and uncooperative individuals: competitive players take advantage of the ability to escape retribution and use a "hit-and-run" strategy, while cooperators use the re-matching to form stable relationships with one another Boone and Macy [1998].

Barclay and Raihani [2015].

4 Exogenous sorting

The bottom line from the endogenous sorting literature is that the presence of subjects whose subjective payoffs do not coincide with the material payoffs of the game is not sufficient for groups to sustain high cooperation levels. Cooperative behaviors can survive over time only if the freedom to form and leave groups is coupled with a pre-determined rule that allows conditional cooperators to stay away from free-riders. This prompts the question of whether there is an intrinsic value to the freedom to choose one's partners. After all, the separation of cooperators from free-riders could also be done exogenously, e.g. by the experimenter. The exogenous sorting of participants into groups has been the focus of much experimental research over the last decade.

Gunnthorsdottir et al. [2007] experimentally investigate the way individual disposition and history (experience playing with subjects having different preferences) interact to affect a subject's cooperative decision-making. The authors implement a public goods game with groups of 4 playing for 10 rounds. Each session includes 12 subjects. Their experiment has two main design features. First, subjects are grouped according to two different rules: (1) in the *baseline* condition, in each round, each subject has an equal chance of being grouped with any three other subjects, while in the (2) *sorted* condition, after subjects have made their contribution decisions, the four highest contributors are placed into one group, the fifth to the eighth highest are placed into another group and the four lowest into a third group. To avoid strategic behavior, subjects are not informed about the sorting mechanism. The second design feature of their experiment is that the two group assignment rules are crossed with three MPCR levels: 0.3, 0.5, and 0.75.

The authors find that within each MPCR condition, aggregate contribution levels in the sorted condition always exceed those in the baseline. Moreover, cooperative decay is slower in the sorted condition. To better understand the effect of history on individual decision-making in the sorted condition compared to the baseline, subjects were classified into *free-riders* if they contributed 30% or less of their endowment to the public account in the first round,[18] and into *cooperators* otherwise. An analysis by types reveals that cooperator contributions in the sorted condition exceed cooperator contributions in the

[18] Although the subjects were informed that the experiment had 10 rounds, the authors contend that first-round decisions reveal subjects' predisposition to cooperate because they are not influence by the history of the game. However, as noted by Kreps and Wilson [1982], forward looking subjects may behave strategically starting from the first round with the intention to influence others' beliefs about the composition of the group and their subsequent behavior in the game.

baseline no later than the fourth round, and continue to do so until round 10. Accordingly, the authors conclude that cooperators behave differently in the sorted condition than in the baseline condition because in the former there are less encounters with free-riders than in the latter.

Nax et al. [2017] generalize Gunnthorsdottir et al. [2007]'s design to include noise in how precisely each individual's contribution is detected. In doing so, they vary the degree to which group assignment accurately reflects earlier contributions, ranging from a situation in which groups are not based on earlier contributions (reflecting a "perfect strangers" public goods game) to the "perfect meritocracy" situation of Gunnthorsdottir et al. [2007]. With no or low levels of noise, consistent with Gunnthorsdottir et al. [2007], contributions are close to the social optimum. Furthermore, even with substantial noise, such that the unique Nash equilibrium is full free-riding, contributions tend to stabilize at intermediate levels.

Gunnthorsdottir et al. [2007] approach is carried further by Ones and Putterman [2007], who control group formation in a more complex collective action environment. In their experiment, subjects have two decision variables under their control: (1) decisions on how much to contribute to the public account and (2) decisions on whether and by how much to impose costly punishment on other group members after learning of their contributions. More specifically, in each session, 16 subjects make 25 sets of contribution and punishment decisions. The 25 rounds are divided into 4 segments. That is, subjects are informed that they would be put in one group of 4 in the first round, a possibly different group whose members would not change during rounds 2–5, another fixed group for rounds 6–15, and a final fixed group for rounds 16–25. Ones and Putterman [2007] look at two different grouping rules: purposeful and random grouping.

The purposeful grouping treatment attempted first to identify types. In the first round, subjects made contribution decisions and were informed about the decisions made by the other three members of the group. However, without subjects' knowledge, groups were formed only after the first contribution decisions. This was done in order to create similarly diverse groups composed of one of the four highest first-round contributors in the session, one of the four lowest contributors, and so on. The assignment to groups for rounds 2-5 is similar to the first round. In addition, the computer also calculates a punishment index based on how much subjects punished low contributors at the end of the first round. Thus, groups are composed of both high and low contributors, both aggressive punishers of low contributors and non-punishers or perverse punishers. At the end of the fifth round, the subjects were ranked by their average contribution over rounds 1–5 as a whole and by their average punishment index over periods 1–5 as a whole. The contribution and punishment ranks were added together. Subjects with the lowest summed rank (high contributors who aggressively punished low contributors) were grouped *together* for rounds 6–15. The group composition is therefore homogeneous during these rounds.[19] Finally, in rounds 16–25 subjects were placed in randomly formed groups. In the random grouping treatment, subjects were placed in randomly created groups in rounds 1, 2, 6, 16.

Ones and Putterman [2007] confirm the findings from Gunnthorsdottir et al. [2010] that in the purposeful grouping treatment highly cooperative groups contribute more than their counterparts in the random grouping treatment. The authors also find that the behaviors displayed in rounds 1–5 are somewhat predictive of behaviors in rounds 6–15 and in rounds 16–25. The persistence of types over time holds for contribution and punishment decisions.

Although the classification method in Ones and Putterman [2007] and in Gunnthorsdottir et al. [2007] seem to be validated experimentally by the observed differences in behavior between the sorted and the unsorted groups, the effect size is not large. One of the possible reasons may lie in their sorting method that rests on repeated interactions which may introduce all sorts of strategic behavior. For instance, Ockenfels and Weimann [1999] report an experiment in which they sort people into "cooperative" and "less cooperative" groups on the basis of observing the subjects' contributions over ten rounds of a repeated public goods game. They do not observe any effect of this sorting on cooperation levels.

Gächter and Thöni [2005] implement an experimental protocol aimed at overcoming the possible bias introduced by measuring agents' types within repeated strategic interactions. They use a one-shot linear public goods game as the measurement instrument for cooperative attitudes. More specifically, subjects first participate in a "ranking" experiment, which consists of playing a one-shot linear public goods game with an MPCR = 0.6 in randomly formed groups of 3. Subjects were informed that they would play the "ranking" experiment just once and that some other part of the experiment would follow, but were not given further details. This was done in order to ensure that subjects' decisions in the ranking experiment are not strategically biased and reflect people's cooperative attitudes. Additionally, they did not receive immediate feedback on the outcome of the "ranking" experiment game.

Subjects then received the instructions for the second part of the experiment, called

[19] The authors wanted to have diverse groups in rounds 1–5 in order to control for differences in behavior stemming from differences in the kind of group one may find oneself in.

the "sorted" experiment, which consists of playing a ten periods linear public goods game with fixed membership. Before playing the "sorted" experiment, subjects were informed that they had been ranked according to their contribution to the public account in the first part of the experiment, i.e. in the ranking experiment. They were also publicly informed that the three highest contributors in the ranking experiment were put together in one group, the next three in the second group and so on to the three lowest contributors who form the last group. Subjects were then informed about their new group members contributions in the ranking experiment. There is also a control treatment, called the "unsorted" experiment, where after playing the ranking experiment, the groups are formed randomly and each participant is informed about their new group mates' contributions in the ranking experiment. The authors also combine the two treatments – sorted and unsorted – with the opportunity to punish group members at a cost.

In the sorted experiment, the subjects who were sorted in the TOP contributor groups invested on average 18.1 ECU in the ranking experiment, MIDDLE contributors invested 10.1 ECU and LOW contributors invested 0.8 ECU. The sorting mechanism implemented by Gächter and Thöni [2005] led to a substantial increase in cooperation. While average contributions were 9.5 ECU in the unsorted experiment, they amounted to 13.9 ECU in the sorted experiment. At the individual level, the authors find that the top third of contributors in the sorted experiment invested significantly more in the public account than the most cooperative third of subjects in the unsorted experiment. Even without any punishment opportunities, the difference in cooperation levels is quite substantial (18.4 vs 14.1 ECU).[20]

The same year as the publication of Gächter and Thöni [2005] paper, there appeared Burlando and Guala [2005] experiment that combines multiple sources of evidence in order to refine the categories of agents in social dilemma games. Their experiment consists of two parts, with exactly a one week interval between them. The first part is composed of four different tasks: the strategy method task used by Fischbacher et al. [2001] to elicit participants' full schedule of contributions conditional on the cooperation of others; the decomposed game technique used by Offerman et al. [1996] to elicit social value orientation; a linear repeated public goods game for 20 rounds; and a questionnaire. The second part of the experiment consists of other 20 rounds of a repeated linear public goods game.

In the first part of the experiment, the assignment of subjects to groups was random.

[20] In fact, the availability of costly punishment does not result in higher cooperation levels compared to the sorted experiment without punishment. In the latter, the groups manage to sustain the same level of cooperation as in the former.

At the end of this part and before playing the second part of the experiment a week later, subjects had not been given any information about the results of the various tasks. They were also unaware of the fact that their behavior in the four different tasks would be used to classify them into "types" and that similar types would end up in the same group for the second part of the experiment.[21]

With the data from the four different tasks, Burlando and Guala [2005] classify the subjects into three main categories: 32% are classified as free riders, 18% as (unconditional) cooperators, and 35% as reciprocators, with the remaining 15% classified in a "noisy" group. By focusing on the behavior of pre-determined types in the second part of the experiment, the authors notice that free-riders start with a lower level of contribution, which tends to decay very quickly. Cooperators and reciprocators start with a very high level of contribution, which remains high throughout most of the game. The behavior of reciprocators is particularly impressive, with constant (and almost full) contribution until round 19.

However, it should be noted that the difference in cooperation levels between cooperators and reciprocators is small. As emphasized by the authors, this is probably due to the fact that they were not entirely successful in forming perfectly homogeneous groups. Given that the sorting method implemented by Burlando and Guala puts a high weight on people's behavior in the repeated public goods game, it may be that their groups of cooperators were "infiltrated" by other types. This is not the only difference between their design and Gächter and Thöni [2005] experiment. Contrary to the later, in Burlando and Guala [2005] the subjects are not aware of the group composition.

de Oliveira et al. [2015] study how providing information about the group composition affects the behavior of different social preference types. Their experiment consists of two waves of experimental sessions. In the first wave participants are invited via email to participate in an internet experiment. They play a one-shot linear public goods game with strategy method as in Fischbacher et al. [2001]. The subjects' decisions are used to classify them into two types: (1) "Selfish" and (2) "Conditional Cooperators."[22] The second wave of the experiment took place on a different day. Participants were invited to the laboratory and played a linear public goods game for 15 rounds in groups of 3. The

[21] More precisely, Burlando and Guala [2005] assigned weights to the four classification methods. Given that the ultimate goal was to understand behavior in a repeated public goods game, they decided to put more weight on the public goods game data, according to the following formula: repeated public goods game 40%; strategy method 20%; decomposed game 20%; questionnaire 20%. When no classification reached the 50% level (and therefore in all cases of tie), they assigned the subjects to a "noisy" group.

[22] Subjects who never give more than five are classified as selfish. The threshold is similar to the one used by Gunnthorsdottir et al. [2007] to identify the selfish types.

laboratory experiment has two main design features. First, participants are grouped into either (i) homogeneous groups of all conditional cooperators (C, C, C), (ii) homogeneous groups of all selfish players (S, S, S), or (iii) heterogeneous groups of two of one type and one of another (S, S, C) and (C, C, S). The second design feature of de Oliveira et al. [2015], participants are explicitly told about the group composition prior to starting the experiment.[23] In the "No Information" treatment, participants are not informed about the group composition. In both treatments, participants see information about their own type. The composition of the group remains unchanged for the 15 rounds.

The first surprising result is that under no information about the group composition contributions appear similar in the (C, C, C) grouping compared with the (C, C, S), but higher compared with the (C, S, S). Given the small size of the group, the similar levels of cooperation in groups of three and groups of two conditional cooperators is indeed a puzzling finding. A plausible explanation may be that the classification method adopted by de Oliveira et al. [2015] does not distinguish between reciprocators and unconditional cooperators. The latter type of players would not respond to observed free-riding by reducing their own contributions. Another explanation may be that in small groups of three, the free-rider would still contribute positive amounts for strategic reasons.

The second original finding is that information about the group composition affects people's behavior solely in the (C, C, C) grouping. This suggests that for conditional cooperators, knowing that the other group mates are of the same type increases their individual investment in the group account. In addition to the actual presence of subjects with prosocial preferences, the existence of *common knowledge* that all the group members are of the same social preference type increase cooperation among like-minded people.[24]

To summarize, we have seen that the exogenous formation of groups by the experimenter, either based on contributions in earlier rounds under complete information (as in Gunnthorsdottir et al. [2007], Nax et al. [2017], and Ones and Putterman [2007]) or based on earlier tasks that participants did not know would influence later groupings (as in Gächter and Thöni [2005], Burlando and Guala [2005], and de Oliveira et al. [2015]), regularly increase and sustain contributions in public goods games, particularly among the most cooperative participants. It turns out that cooperation is higher when there is

[23] To be more precise, they are first told the meaning of the different types, their own type and the type of their mates.

[24] Other experiments where subjects are sorted out exogenously include van den Berg, van den Berg et al. [2015] who find that when grouped together frequency-based learners cooperate more than success-based learners, Kimbrough and Vostroknutov [2015] Kimbrough and Vostroknutov [2016] introduce an incentivized method of eliciting individual norm-sensitivity and show that when grouped together the subjects who suffer more disutility from violating norms also behave more prosocially wherever there is a norm of prosocial behavior, e.g. in public goods games and common pool resource games.

common knowledge about the cooperative attitudes of the group mates. Additionally, the method of measuring people's cooperative attitudes (i.e., the dimension on which they are sorted) plays a fundamental role in how successful groups are in sustaining cooperation. Some methods that elicit cooperativeness in non-repeated interactions (e.g. Gächter and Thöni [2005]) achieve a higher degree of homogeneity than others (e.g. Ockenfels and Weimann [1999]).

5 Meta-analysis of cooperation in public goods games with endogenous and exogenous sorting

In this section we analytically synthesize the experimental evidence on the effect of different sorting mechanisms on cooperation levels. Specifically, we compare the exogenous and the endogenous mechanisms in the context of linear public goods games. Following the existing surveys (Ledyard [1995], Chaudhuri [2011]) and meta-studies on contribution decisions in linear public goods games in selecting the variables of interest (Croson and Marks [2000], Zelmer [2003] and Fiala and Suetens [2017]) our meta-analysis mainly asks whether the endogenous sorting leads to higher cooperation levels than the exogenous ones and how the design (punishment) and environmental variables (group size, MPCR, number of rounds) affect the functioning of the two sorting mechanisms.

Selection criteria We include papers that comply with the following criteria:

- Incentivized laboratory experiments in controlled environment;
- Linear public good games such that:
  - The game is symmetric (homogeneity in the endowment);
  - The treatments concern the mechanism of sorting subjects into groups (endogenous or exogenous mechanism);
  - The subgame perfect Nash equilibrium is unique and Pareto dominated;
  - The group size > 2;
  - Decisions are taken simultaneously.

Searching, selection and data-abstraction: The search of the economics literature for studies meeting the above criteria was undertaken through Google Scholar, Internet Documents in Economics Access Service (IDEAS), references cited in Ledyard [1995]

Papers	**Journal**	**Observations**	**Mechanism**
Aimone et al. 2011	*Review of Economic Studies*	88	Endogenous
Brekke et al. 2011	*Journal of Public Economics*	58	Endogenous
Burlando and Guala 2005	*Experimental Economics*	17	Exogenous
Cabrera et al. 2013	*Experimental Economics*	20	Endogenous
de Oliveira et al. 2015	*Experimental Economics*	33	Exogenous
Gächter and Thöni 2005	*Journal of the European Economic Association*	42	Exogenous
Gunnthorsdottir et al. 2007	*Journal of Economic Behavior and Organization*	33	Exogenous
Gürerk et al. 2006	*Science*	14	Endogenous
Gürerk 2013	*Journal of Economic Psychology*	28	Endogenous
Kimbrough et al. 2016	*Journal of the European Economic Association*	18	Exogenous
Sääskvuori 2014	*Behavioral Ecology and Sociobiology*	80	Endogenous

Table 1: Papers included in the Meta-Regression

Variable	Obs	Mean	Std. Dev.	Min	Max
Contributions as % of endowment	431	2.42	2.23	0	11.39
MPCR	431	0.48	0.11	0.14	0.75
Group Size	431	4.00	1.81	2.20	11.50
dsorting	431	0.66	0.47	0	1
dpun	431	0.10	0.31	0	1
rounds	431	13.41	8.67	1	30

Table 2: Descriptives of dependent and independent variables.

and Chaudhuri [2011], and posting messages on the Economic Science Association (ESA) Google group. Once the list of all the eligible studies was completed, we sent emails to the respective authors asking for the raw data from the experiment. We collected 431 group-level observations of 11 studies (out of 14 that satisfy our selection criteria) involving 18 treatments overall (see Table 1). For each of these studies we observe: the MPCR, group size, the sorting mechanism implemented, total number of rounds, initial endowment, presence of punishment and the amount of group contributions in each round.

Descriptives The variable of interest of our analysis is the average of the group contribution as percentage of the initial endowment. We also include both design and environmental variables (Ledyard [1995]): the MPCR, the dimension of groups, a dummy variable *dsorting* capturing the mechanism employed (0 if exogenous, 1 if endogenous), a dummy variable *dpun* indicating whether subjects can monetarily punish each other (0 if not, 1 if punishment is allowed), and the total number of rounds.

Table 3 shows pairwise correlation among regressors. Variables involved are not highly

	MPCR	dsorting	dpun	rounds
MPCR	1			
dsorting	-0.30*	1		
dpun	-0.20*	-0.14*	1	
rounds	-0.07	0.08	0.23	1

Table 3: Pairwise Correlation. * indicates significance at 5%

correlated. However, as we show below the correlation between the MPCR and the variable *dsorting* is of particular relevance when analyzing their impact across different model specifications.

Quantitative Results We employ a WLS cluster-robust variance estimation, where clusters are defined at the treatment level, having so 18 clusters[25]. Table 4 shows the results obtained. Since the unit of observation in our dataset is at the group level, the total number of data points in the meta-analysis is equal to the number of studies times the number of groups in each study, across treatments. Each observation is weighted by the standard deviation of contributions in the same treatment.[26]

Fig. 1 shows a higher efficiency level in experiments using an endogenous sorting mechanism than in experiments using an exogenous sorting. The average contribution in experiments with an endogenous sorting is 37% higher than in experiments implementing an exogenous sorting. Indeed, as shown in the model specification (4) (table 4), *dsorting* has a significant impact on contribution levels[27] controlling for MPCR, punishment and rounds[28].

In the first section of this paper, we highlighted the fact that under free-migration cooperation is higher when punishment of free-riders is possible. Here we investigate the role of monetary punishment under the two sorting mechanisms, endogenous and exogenous. On one hand, Gächter and Thöni [2005] is the paper that implements an exogenous sorting mechanism with monetary punishment. On the other hand, Gürerk et al. [2006] and Gürerk [2013] are the papers that implement monetary punishment in an endogenous sorting setting. Model (3) and (4) in Table 4 shows that the possibility to punish clearly increases overall contributions under endogenous sorting. However, punishment does not significantly affect contribution decisions in the exogenous mechanism.[29] This result is in

[25] We also run OLS model as robustness check obtaining similar results.

[26] Weighting observations by the standard error of contributions over rounds cannot be done as some of the papers considered in this study implement a one-shot public goods game.

[27] The OLS model shown in table 5 presents similar results.

[28] When considered by its own, *dsorting* has no significant impact on contribution levels. The reason of this shift when controlling for MPCR and *dpun* is because of their correlation (3). Therefore, including *dsorting* š by its own, would lead to biased estimates.

[29] The interaction *dpun*dsorting* is always significantly positive, while *dpun* is only significant in speci-

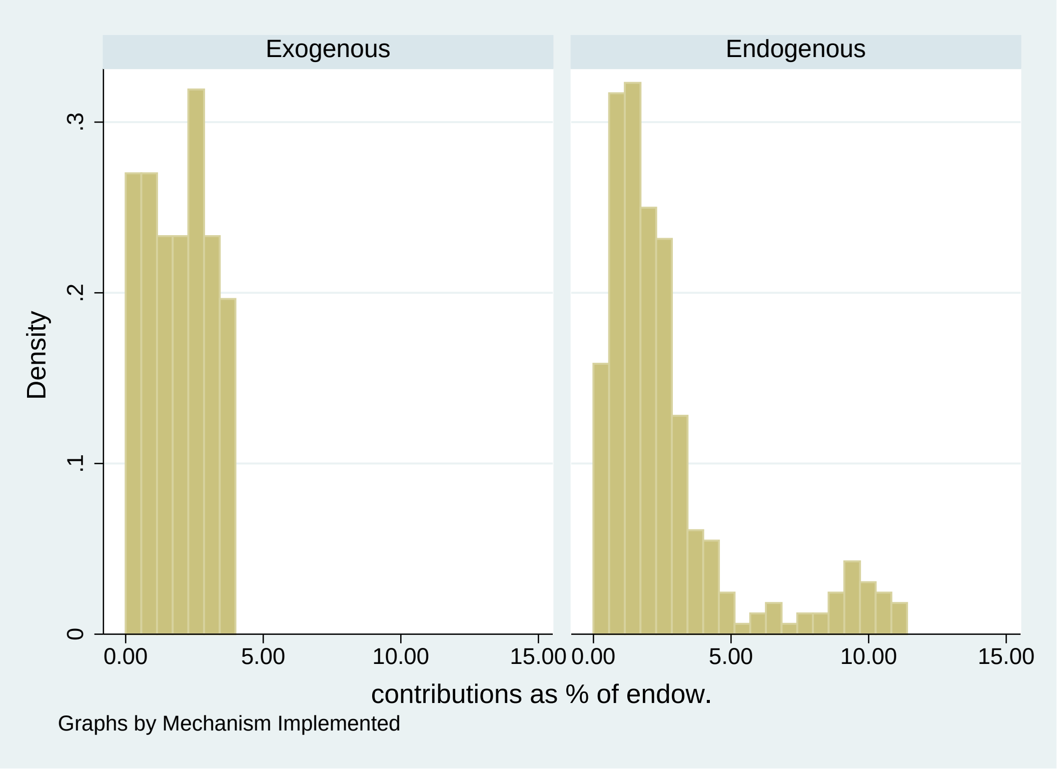

Figure 1: Efficiency by sorting mechanism

line with the afore-mentioned papers implementing punishment.

Previous studies have shown that the MPCR affects cooperation levels under random sorting (Zelmer [2003]). Figure 2 shows that the relationship between the MPCR and group contributions differs in our two sorting mechanisms. In studies implementing an endogenous sorting mechanism, the correlation between the MPCR and group contributions is negative because the MPCR often depends on the group size.[30] In fact, in most of the experiments where subjects are free to create or leave groups, the size of the group is positively correlated with the efficiency levels (% of endowment contributed to the group account). The relationship between the size of the group and the level of the MPCR in the studies implementing an endogenous sorting is presented in fig.3. People are attracted to a specific community (e.g. one where punishment is possible) because efficiency there is higher. Consequently, the MPCR is lower over time in successful communities.

To summarize, the collected data suggest that contribution levels are higher in the endogenously sorted groups than in the exogenously sorted ones when we control for some

fication (3).

[30] The MPCR in these instances is computed as an inverse function of the group size, which is why we decided to keep only one variable in the regression, namely MPCR. However, not every setting implementing an endogenous mechanism is associated with varying MPCR, since not all allow the group size to vary. In our dataset the exceptions are Aimone et al. [2013], Brekke et al. [2011] and Cabrera et al. [2013], which keep the group size constant over the experimental session.

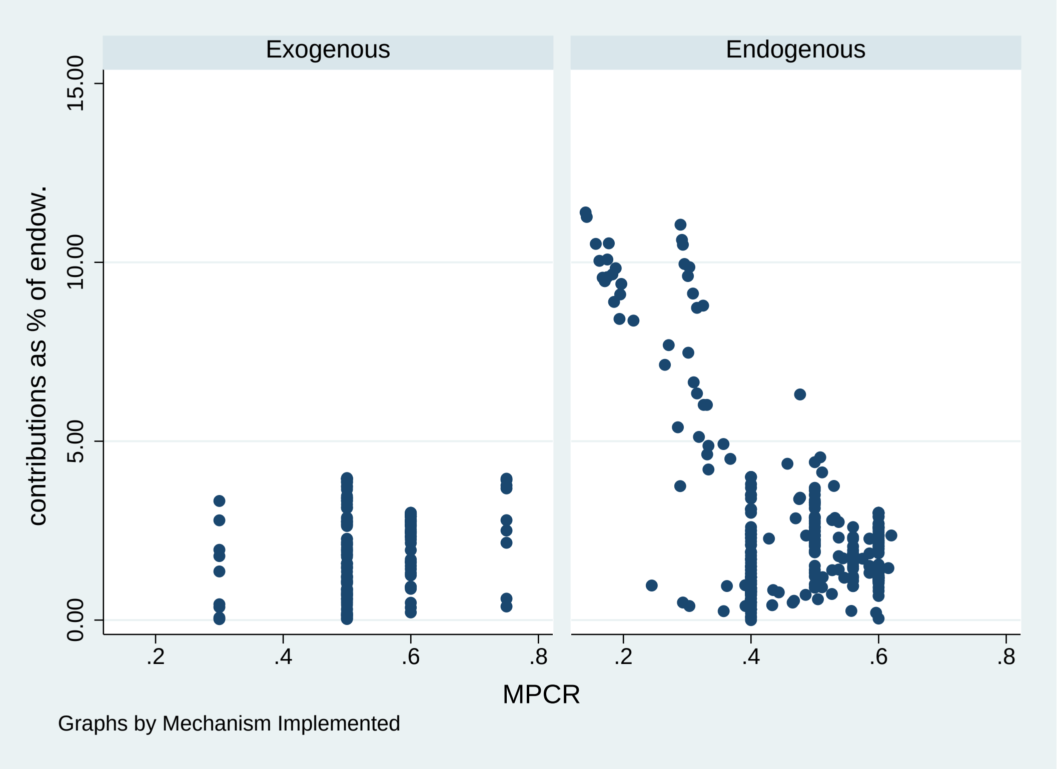

Figure 2: Scatter-plot of the relation between contribution levels and the MPCR by sorting mechanism

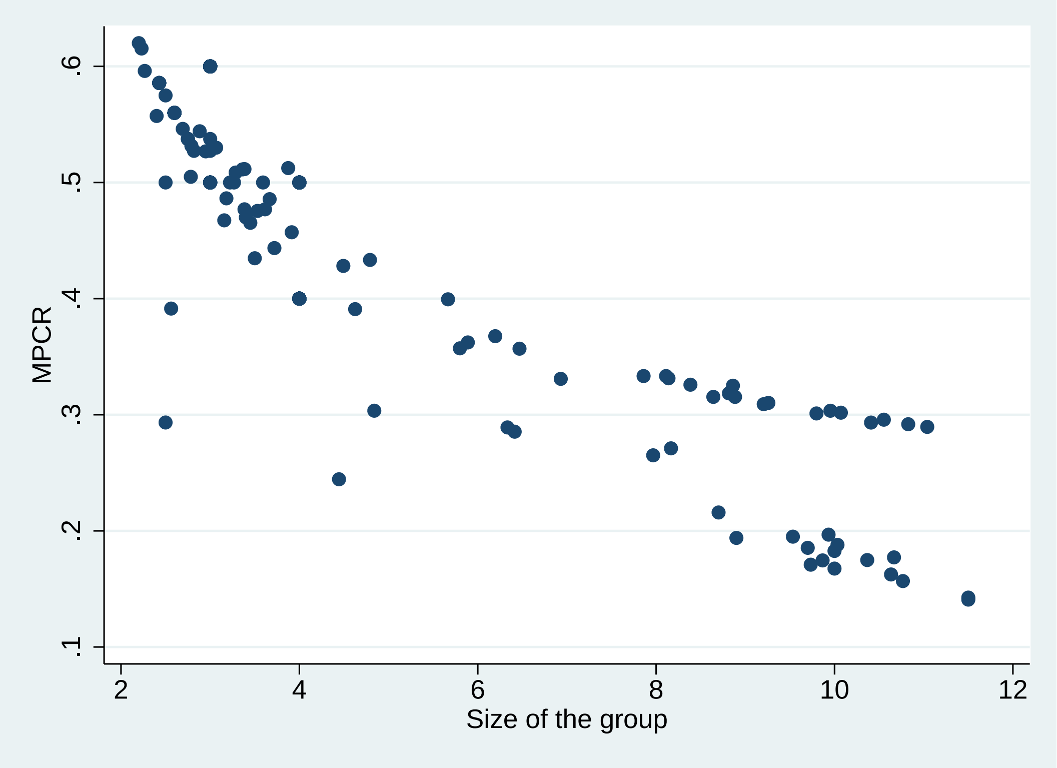

Figure 3: Scatter-plot of MPCR and group size

of the environmental and design variables that might have affected cooperative decisions. With respect to the impact of punitive measures in the two sorting mechanisms, the meta-analysis confirms that punishment is more effective in the endogenously formed groups than in the exogenously formed ones. There is also indication that in endogenously formed groups without a cap on the group size the success of a specific institution is accompanied by an increase in the group size and consequently a decrease in the MPCR. This is a plausible explanation for a more stable level of cooperation in studies that exogenously fix the group size, yet allowing subjects to voluntarily choose a specific group before playing the social dilemma Aimone et al. [2013] or during the game Brekke et al. [2011].

6 Conclusion

There are several important results that emerge from the surveyed literature on group formation. First, Ehrhart and Keser [1999] seminal finding that under free migration defectors join highly cooperative groups and cooperators try to flee them has been replicated as a baseline in other experiments with the same result – individual strategies rapidly converge to the Nash equilibrium.

Second, cooperation improves with the use of punitive and non-punitive mechanisms. On the punitive side, peer monetary punishment leads to greater cooperation levels and higher efficiency over time. On the non-punitive side, many experimental studies find that allowing people to undertake costly actions, such as sacrificing one's private return, successfully separates cooperators from defectors and encourages the former to contribute more than in randomly composed groups. Similarly, the option to enter/exit a social dilemma game often promotes cooperation because of a selection effect: cooperators have higher expectations about the likelihood that others will cooperate and are more likely to opt-in to a social dilemma. The downside from allowing subjects to voluntarily form or leave groups is that free-riders do not learn to cooperate without punishment.

The third important lesson that emerges from the surveyed literature is that with exogenous sorting the level of cooperation varies with the dimension on which subjects are sorted into groups – eliciting people's cooperative attitudes in non-repeated games results in more homogeneous groups – and with the information that subjects have about the type of their group mates – cooperation rates are higher with common knowledge about the cooperative type of the group members. To conclude with the exogenous formation of groups, mixed groups (composed of free-riders and cooperators) do not promote cooperation among free-riders, unless there is only one free-rider in a group of many cooperators.

Table 4: Results from the WLS regression

	(1)	(2)	(3)	(4)
dsorting	0.318			5.181*
	(0.360)			(0.022)
MPCR		-3.137		3.887*
		(0.247)		(0.019)
MPCR*dsorting		0.0857		-9.407*
		(0.893)		(0.019)
dpun			0.362*	0.407
			(0.037)	(0.119)
dpun*dsorting			6.536***	3.564*
			(0.000)	(0.039)
rounds				0.0629
				(0.078)
const.	1.804***	3.560*	1.887***	-1.119
	(0.000)	(0.020)	(0.000)	(0.362)
N	431	431	431	431
R^2	0.011	0.040	0.252	0.283
adj-R^2	0.009	0.036	0.249	0.274
Note:	*p<0.05; **p<0.01; ***p<0.001			

Table 5: Results from the OLS regression

	(1)	(2)	(3)	(4)
dsorting	0.722			6.088*
	(0.106)			(0.019)
MPCR		-9.415*		3.211***
		(0.017)		(0.001)
MPCR*dsorting		0.588		-11.820*
		(0.506)		(0.015)
dpun			0.162*	0.273
			(0.454)	(0.269)
dpun*dsorting			6.614***	3.017*
			(0.000)	(0.144)
rounds				0.059
				(0.078)
const.	1.944***	7.152*	2.087***	-0.543
	(0.000)	(0.001)	(0.000)	(0.565)
N	431	431	431	431
R^2	0.023	0.241	0.428	0.523
adj-R^2	0.021	0.237	0.425	0.517
Note:	*p<0.05; **p<0.01; ***p<0.001			

References

T. K. Ahn, R. Mark Isaac, and Timothy C. Salmon. Endogenous group formation. *Journal of Public Economic Theory*, 10(2):171–194, 2008. doi: 10.1111/j.1467-9779.2008.00357.x.

T. K. Ahn, R. Mark Isaac, and Timothy C. Salmon. Coming and going: Experiments on endogenous group sizes for excludable public goods. *Journal of Public Economics*, 93 (1-2):336–351, 2009. doi: 10.1016/j.jpubeco.2008.06.007.

Jason A. Aimone, Laurence R. Iannaccone, Michael D. Makowsky, and Jared Rubin. Endogenous group formation via unproductive costs. *Review of Economic Studies*, 80(4): 1215–1236, 2013. doi: 10.1093/restud/rdt017.

Pat Barclay and Nichola Raihani. Partner choice versus punishment in human Prisoner's Dilemmas. *Evolution and Human Behavior*, 172:263–271, 2015. doi: 10.1016/j.evolhumbehav.2015.12.004.

Ralph Bayer. Cooperation in Partnerships: The Role of Breakups and Reputation. *Journalof Institutional and theoretical Economics*, 172:615–638, 2016. doi: 10.1628/093245616X14610627109836.

Iris Bohnet and Dorothea Kübler. Compensating the cooperators: Is sorting in the prisoner's dilemma possible. *Journal of Economic Behavior and Organization*, 56(1):61–76, 2005. doi: 10.1016/j.jebo.2003.04.002.

R. Thomas Boone and Michael W. Macy. Dependence and Cooperation in the Game of Trump. *Advances in Group Processes*, 15:161–185, 1998.

Kene Boun My and Benoît Chalvignac. Voluntary participation and cooperation in a collective-good game. *Journal of Economic Psychology*, 31(4):705–718, 2010. doi: 10.1016/j.joep.2010.05.003.

Kjell Arne Brekke, Karen Evelyn Hauge, Jo Thori Lind, and Karine Nyborg. Playing with the good guys. A public good game with endogenous group formation. *Journal of Public Economics*, 95(9-10):1111–1118, 2011. doi: 10.1016/j.jpubeco.2011.05.003.

Roberto M Burlando and Francesco Guala. Heterogeneous agents in public goods experiments. *Experimental Economics*, 8:35–54, 2005.

Susana Cabrera, Enrique Fatas, Juan A Lacomba, and Tibor Neugebauer. Vertically splitting a firm: promotion and relegation in a team production experiment. *Experimental Economics*, 16:426–441, 2013.

Gary Charness and Chun Lei Yang. Starting small toward voluntary formation of efficient large groups in public goods provision. *Journal of Economic Behavior and Organization*, 102:119–132, 2014. doi: 10.1016/j.jebo.2014.03.005.

Ananish Chaudhuri. Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, 14(1):47–83, 2011. doi: 10.1007/s10683-010-9257-1.

Matthias Cinyabuguma, Talbot Page, and Louis Putterman. Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics*, 89(8 SPEC. ISS.):1421–1435, 2005. doi: 10.1016/j.jpubeco.2004.05.011.

Giorgio Coricelli, Dietmar Fehr, and Gerlinde Fellner. Partner Selection in Public Goods Experiments. *Journal of Conflict Resolution*, 48(3):356–378, 2004. doi: 10.1177/0022002704264143.

Rachel T A Croson and Melanie Beth Marks. Step returns in threshold public goods: A meta- and experimental analysis. *Experimental Economics*, 2(3):239–259, 2000. doi: 10.1007/BF01669198.

Angela C.M. de Oliveira, Rachel T.A. Croson, and Catherine Eckel. One bad apple? Heterogeneity and information in public good provision. *Experimental Economics*, 18 (1):116–135, 2015. doi: 10.1007/s10683-014-9412-1.

Karl-Martin Ehrhart and Claudia Keser. Mobility and Cooperation: On the Run. *Working papers, Cirano*, 1999.

Lenka Fiala and Sigrid Suetens. Transparency and cooperation in repeated dilemma games: a meta study. *Experimental Economics*, pages 1–17, 2017. doi: 10.1007/s10683-017-9517-4.

Urs Fischbacher, Simon Gächter, and Ernst Fehr. Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3):397–404, 2001. doi: 10.1016/S0165-1765(01)00394-9.

Simon Gächter and Christian Thöni. Social Learning and Voluntary Cooperation Among Like-Minded People. *Journal of the European Economic Association*, 3:303–314, 2005. doi: 10.1162/jeea.2005.3.2-3.303.

Veronika Grimm and Friederike Mengel. Cooperation in viscous populations-Experimental evidence. *Games and Economic Behavior*, 66(1):202–220, 2009. doi: 10.1016/j.geb.2008.05.005.

Anna Gunnthorsdottir, Daniel Houser, and Kevin McCabe. Disposition, history and contributions in public goods experiments. *Journal of Economic Behavior and Organization*, 62(2):304–315, 2007. doi: 10.1016/j.jebo.2005.03.008.

Anna Gunnthorsdottir, Roumen Vragov, Stefan Seifert, and Kevin McCabe. Near-efficient equilibria in contribution-based competitive grouping. *Journal of Public Economics*, 94 (11-12):987–994, 2010. doi: 10.1016/j.jpubeco.2010.07.004.

O Gurerk, B Irlenbusch, and B Rockenbach. Voting with feet: community choice in social dilemmas. *Uni Erfurt Working Paper*, (4643):1–46, 2010.

Özgür Gürerk. Social learning increases the acceptance and the efficiency of punishment institutions in social dilemmas. *Journal of Economic Psychology*, 34:229–239, 2013. doi: 10.1227/01.NEU.0000349921.14519.2A.

Özgür Gürerk, Bernd Irlenbusch, and Bettina Rockenbach. The competitive advantage of sanctioning institutions. *Science*, 312:108–111, 2006.

Özgür Gürerk, Bernd Irlenbusch, and Bettina Rockenbach. On cooperation in open communities. *Journal of Public Economics*, 120:220–230, 2014. doi: 10.1016/j.jpubeco.2014.10.001.

Barton H. Hamilton, Jack A. Nickerson, and Hideo Owan. Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation. *Journal of Political Economy*, 111(3):465–497, 2003. doi: 10.1086/374182.

Esther Hauk. Multiple prisoner's dilemma games with(out) an outside option: An experimental study. *Theory and Decision*, 54(3):207–229, 2003. doi: 10.1023/A:1027385819400.

Esther Hauk and Rosemarie Nagel. Choice of Partners in Multiple Two-Person Prisoner's Dilemma Games: An Experimental Study. *Journal of Conflict Resolution*, 45(6):770–793, 2001. doi: 10.1177/0022002701045006004.

Luisa Herbst, Kai A. Konrad, and Florian Morath. Endogenous group formation in experimental contests. *European Economic Review*, 74:163–189, 2015. doi: 10.1016/j.euroecorev.2014.12.001.

Laurence R. Iannaccone. Sacrifice and Stigma: Reducing Free-riding in Cults, Communes, and Other Collectives. *Journal of Political Economy*, 100(2):271–291, 1992. doi: 10.1086/261818.

Claudia Keser and Claude Montmarquette. Voluntary versus Enforced Team Effort. *Games*, 2(3):277–301, 2011. doi: 10.3390/g2030277.

Claudia Keser and Frans van Winden. Conditional Cooperation and Voluntary Contributions to Public Goods. *The Scandinavian Journal of Economics*, 102(1):23–39, 2000. doi: 10.1111/1467-9442.00182.

Erik O. Kimbrough and Alexander Vostroknutov. The social and ecological determinants of common pool resource sustainability. *Journal of Environmental Economics and Management*, 72(430):38–53, 2015. doi: 10.1016/j.jeem.2015.04.004.

Erik O. Kimbrough and Alexander Vostroknutov. Norms Make Preferences Social. *Journal of the European Economic Association*, 14(3):608–638, 2016. doi: 10.1111/jeea.12152.

David M. Kreps and Robert Wilson. Sequential Equilibria. *Econometrica*, 50(4):863—-894, 1982. doi: 10.2307/1912767.

John O Ledyard. Public Goods: A Survey of Experimental Research. In *The Handbook of Experimental Economics*, pages 111–194. 1995. ISBN 069104290X (acid-free paper). doi: 10.3987/Contents-12-85-7.

Frank P. Maier-Rigaud, Peter Martinsson, and Gianandrea Staffiero. Ostracism and the provision of a public good: experimental evidence. *Journal of Economic Behavior & Organization*, 73(3):387–395, 2010. doi: 10.1016/j.jebo.2009.11.001.

David Masclet. Ostracism in work teams: a public good experiment. *International Journal of Manpower*, 24(7):867–887, 2003. doi: 10.1108/01437720310502177.

D T Miller and J G Holmes. The role of situational restrictiveness on self-fulfilling prophecies: A theoretical and empirical extension of Kelley and Stahelski's triangle hypothesis. *Journal of Personality and Social Psychology*, 31:661–673, 1975. doi: 10.1037/h0077081.

Heinrich H Nax, Stefano Balietti, Ryan O Murphy, and Dirk Helbing. A noisy institution : An experimental welfare investigation of the contribution-based grouping mechanism. *Social Choice and Welfare*, 2017. doi: 10.1007/s00355-017-1081-5.

Daniele Nosenzo and Fabio Tufano. The Effect of Voluntary Participation on Cooperation. *Journal of Economic Behavior & Organization*, 142:307–319, 2017. ISSN 01672681. doi: 10.1016/j.jebo.2017.07.009.

Axel Ockenfels and Joachim Weimann. Types and patterns: an experimental East-West-German comparison of cooperation and solidarity. *Journal of Public Economics*, 71(2): 275–287, 1999. doi: 10.1016/S0047-2727(98)00072-3.

Theo Offerman, Joep Sonnemans, and Arthur Schram. Value Orientations, Expectations and Voluntary Contributions in Public Goods. *The Economic Journal*, 106(437):817, 1996. doi: 10.2307/2235360.

Umut Ones and Louis Putterman. The ecology of collective action: A public goods and sanctions experiment with controlled group formation. *Journal of Economic Behavior and Organization*, 62(4):495–521, 2007. doi: 10.1016/j.jebo.2005.04.018.

John M Orbell and Robyn M. Dawes. Social Welfare, Cooperators' Advantage, and the Option of Not Playing the Game. *American Sociological Review*, 58(6):787–800, 1993. doi: 10.2307/2095951.

John M Orbell, P Schwartz-Shea, and Randy T Simmons. Do Cooperators Exit More Readily Than Defectors? *American Political Science Review*, 76(1):753–766, 1984. doi: 10.2307/1961254.

Andrea Robbett. Local institutions and the dynamics of community sorting. *American Economic Journal: Microeconomics*, 6(3):136–156, 2014. doi: 10.1257/mic.6.3.136.

Andrea Robbett. Community dynamics in the lab. *Social Choice and Welfare*, 46(3): 543–568, 2016. doi: 10.1007/s00355-015-0928-x.

Lauri Sääksvuori. Intergroup conflict, ostracism, and the evolution of cooperation under free migration. *Behavioral Ecology and Sociobiology*, 68(8):1311–1319, 2014. doi: 10.1007/s00265-014-1741-8.

Rudolf Schuessler. Exit Threats and Cooperation under Anonymity. *The Journal of Conflict Resolution*, 33(4):728–749, 1989. doi: 10.1177/0022002789033004007.

Charles M. Tiebout. A Pure Theory of Local Expenditures. *Journal of Political Economy*, 64(5):416–424, 1956. doi: 10.1086/257839.

Pieter van den Berg, Lucas Molleman, and Franz J. Weissing. Focus on the success of others leads to selfish behavior. *Proceedings of the National Academy of Sciences*, 112 (9):2912–2917, 2015. doi: 10.1073/pnas.1417203112.

Alistair J. Wilson and Hong Wu. At-will relationships: How an option to walk away affects cooperation and efficiency. *Games and Economic Behavior*, 102:487–507, 2017.

Toshio Yamagishi. Exit from the group as an individualistic solution to the free rider problem in the United States and Japan. *Journal of Experimental Social Psychology*, 24 (6):530–542, 1988. doi: 10.1016/0022-1031(88)90051-0.

Jennifer Zelmer. Linear public goods experiments: A meta-analysis. *Experimental Economics*, 6(3):299–310, 2003. doi: 10.1023/A:1026277420119.