When the two ends meet: an experiment on cooperation across the Italian North-South divide

When the two ends meet: an experiment on cooperation and social capital*

Pietro Battiston†, Simona Gamba‡

Abstract

We study the behaviour of individuals with different geographic origins interacting in a same public good game. We exploit the peculiar composition of the experimental sample to compare the performance of groups where individuals have mixed origins to homogeneous groups. We find that, despite the absence of any geographic framing, mixed groups exhibit significantly lower contributions. We also find that cooperation levels differ significantly across geographic origins, in line with the existing literature. This is explained by a different impact of coordination opportunities, such as communication, as we show by manipulating them. Our results point towards integration as a crucial aspect for the economic development of intercultural societies. They also confirm that, rather than being explained just by the differences in institutions and economic opportunities, the Italian North-South divide embeds elements of distrust, prejudice and a consequent path dependence in the level of social capital.

*We thank for comments Marco Casari, Alessandro Bucciol, Maria Bigoni, Valentina Rotondi, participants in the 2017 Royal Economic Society Conference, the CRIEP Workshop, the 14th University of Bologna Workshop on Social Economy for Young Economists, the 13th Maastricht Behavioral and Experimental Economics Symposium, the 58th Yearly Scientific Meeting SIE, together with participants in seminars at FBK-IRVAPP, University of Bolzano and University of Milan-Bicocca. We are grateful to Chiara Busnelli for her great support.

†Department of Economics and Management, University of Parma. Institute of Economics, Sant'Anna School of Advanced Studies. ORCID: 0000-0001-5343-7861. me@pietrobattiston.it

‡Department of Economics and Finance, Catholic University of Milan. ORCID: 0000-0002-5538-2873. gamba.simona@gmail.com

Keywords: public good, cooperation, social capital, cultural differences, lab experiment.
JEL classification: A13, C71, C92, H41.

1 Introduction

Almost since the dawn of experimental economics, researchers have looked with growing interest at what experiments ran in different locations and contexts could reveal about the specific characteristics of cultures and societies. This is particularly true for experiments focusing on traits related to *social capital*, such as trust (Croson and Buchan, 1999; Cassar and Wydick, 2010), cooperation (Cason et al., 2002; Finocchiaro Castro, 2008) and fairness (Oosterbeek et al., 2004), given the fundamental role that social capital bears in explaining variations in institutional organisation and economic outcomes (Knack and Keefer, 1997; Buonanno et al., 2009; Hoyman et al., 2016; Camussi et al., 2018).

Some countries have been shown to exhibit important *internal* differences in social capital across regions: apart from the case of Italy, discussed below, notable examples are represented by West and East Germany, also analyzed in the experimental literature (Ockenfels and Weimann, 1999; Brosig-Koch et al., 2011), and by heterogeneities across American States (Putnam, 2001). Analogous differences have been found between citizens of different European countries (Bornhorst et al., 2010; Finocchiaro Castro, 2008).

In Italy, profound internal differences in terms of social capital and cooperation have since long been identified as one of the causes of the North-South economic divide (Helliwell and Putnam, 1995; Leonardi, 1995; Guiso et al., 2004) that has characterised the country since its unification and has widened in the last decades. Although the historical origins of the gap remain a matter of debate, empirical evidence on its existence is overwhelming. Strong differences can be found not only in the values of economic indicators (GDP per capita, unemployment rate, internal migrations), but also looking at measures of quality of institutions (timeliness of budgets, legislative innovation, citizen satisfaction) and other indicators of social capital (frequency of blood donations, number of associations, voters turnout at elections, newspaper readership).[1]

While such differences are often explained on the ground of disparities in economic opportunities and quality of institutions, a stream of literature which can be traced back to the seminal work of Banfield (1967) focuses

[1]The empirical literature on the *"Questione meridionale"*, i.e. the North-South gap, is vast: see Helliwell and Putnam (1995); Ichino and Maggi (2000); Felice (2013) for a more comprehensive view.

instead on the individual determinants of the propensity to cooperate. Ichino and Maggi (2000), for instance, exploit the phenomenon of on-the-job movers inside a large Italian bank to compare individuals with different geographic backgrounds who face the same incentives at work. Bigoni et al. (2016) instead run a "laboratory-in-the-field" experiment in two cities located in the North and two located in the South of Italy, with experimental subjects being presented the exact same incentives and experimental conditions. Their results confirm that observed disparities in behaviour cannot be explained just by differences in the economic context, but rather are *"likely to derive from persistent differences in social norms"*.

In this paper, through an artefactual field experiment, we study the propensity of individuals from both the North and the South of Italy to contribute in a *same* repeated public good game. Compared to the literature on the North-South gap, the crucial novelty is therefore our ability to observe the interaction between individuals characterised by different geographic background.

Aside from the experimental literature, a growing stream of research focuses on the comparison of migrants and on-the-job movers to local populations (Ichino and Maggi, 2000; Gibson et al., 2015; Algan et al., 2016). Conversely, our setting abstracts from the determinants and effects of migration and integration (or segregation): our subjects were living in different cities *at the time of the experiment*, they moved to the location were the experiment took place only for few days, and except for the geographic background, they shared similar characteristics (such as age and education). In the experiment, we manipulate both the composition of the groups and the coordination opportunities (namely, identification and communication) available in specific phases of the public good game. Thus, we further improve upon the existing literature by exploring the effect that such opportunities have on contributions, and hence by shedding light on what determines differences in contributions between Northern and Southern citizens.

Our results confirm differences in reactions to identical incentives: specifically, lower contributions on behalf of Southern citizens. In a closely related study, Bigoni et al. (2018) separately elicit contributions, conditional contributions (i.e. as functions of peers' ones) and expectations regarding peers' contributions. This allows them to demonstrate that individuals from the South *do expect* lower contributions from their peers - while they do not exhibit a lower level of conditional cooperation. Relatedly, in our experiment, the difference in cooperation only emerges in the presence of coordination

opportunities: their positive effect is weaker for individuals from the South.

The aforementioned evidence emphasises the importance of behavioural aspects in explaining the North-South divide in Italy: although behavioural traits might have evolved as an adaptation to institutional characteristics (as already suggested by Putnam et al., 1993), it can be misleading to expect that changing such characteristics will have an immediate positive effect on cooperation levels. This argument can partly explain the failure of past measures adopted by Italian policymakers in order to close the North-South gap, and must be taken into account when planning further actions in this direction.

The comparison of individuals in mixed and homogeneous groups reveals that, once they get to know the composition of their group, the latter contribute more. The literature on performance in multi-ethnic groups reports diverse results: while insights from evolutionary psychology (Barrett, 2007), as well as experiments on trust games (Fershtman and Gneezy, 2001) would predict sub par performance, different results have been observed when *creativity* is involved (McLeod and Lobel, 1992), when one ethnicity has a more pro-social behaviour (Cox et al., 1991), or when conformism is actually detrimental (Levine et al., 2014).[2] In a study closely related to ours, Finocchiaro Castro (2008) runs a public good game where groups are composed by either Italian, UK, or mixed nationality subjects – participating in the experiment from separate laboratories – and finds that mixed groups, whose members interact remotely across labs, cooperate less. Our study benefits in multiple ways from the fact that subjects are regrouped in the same location. First and foremost, it allows us to avoid embedding in the design any explicit geographic framing. Second, it neutralises any credibility issue related to participants interacting between physically remote labs (Eckel and Wilson, 2006). Third, it allows for a more natural interaction between participants, and solves the problem of different location-specific effects which, if participants interact remotely, are indistinguishable from the effect of participants' geographic origin. At the same time, our sample selection allows us to cleanly distinguish the effect of geographic origin from the possible effect of a common affiliation, or pre-existing social ties, between students. It is worth emphasising that our subjects live in geographically separated areas but are not considered "ethnically diverse"; moreover, they have be-

[2]See Balliet et al. (2014) for a cross-country meta-analysis of in-group effects in cooperative games.

longed to the same nation for the past 150 years, sharing the same system of rights and laws. Under this aspect, our experiment is more similar to the one ran by Rustagi et al. (2016) with German and French speaking Swiss citizens, except that in their design, like in the one by Finocchiaro Castro (2008), subjects interacted remotely, and the geographic background of other participants was explicitly stated.[3]

Compared to the literature on (absence of) anonymity in public good games (Gächter and Fehr, 1999; Andreoni and Petrie, 2004; Rege and Telle, 2004; Bochet et al., 2006; Chaudhuri, 2011; Savikhin Samek and Sheremeta, 2014; Gaudeul and Giannetti, 2015), the novelty of our approach lies in the analysis of within-subject behaviour *across* rounds of a same game. Group composition and coordination opportunities are crucial aspects for the debate on economic inequalities across regions, since they highlight the possible role of prejudice and, more in general, lack of integration (Tajfel et al., 1979). In areas where social capital is scarce, economic development may also be hindered by the relatively difficult interactions with other regions: in absence of measures aimed at overcoming regional disparities, the latter might spontaneously deteriorate over time.

The following section describes the characteristics and design of the experiment, Section 3 summarises the game-theoretical analysis, Section 4 presents our hypotheses, Section 5 the results, and Section 6 concludes.

2 Experimental design

The experiment was ran on October 3, 2015, in Volterra (Italy), as part of a larger project, related to curriculum counselling and social mobility, organised by Sant'Anna School of Advanced Studies (Nuti and Ghio, 2017). It involved students from 13 high schools located in 6 different cities, part of 5 different Italian regions. All students were in their last year of school, hence 17 or 18 years old, and were chosen from the pre-selected schools according to a set of homogeneous requirements.[4] Namely, they shared an average so-

[3]Vice-versa, in the experiment by Cappelletti et al. (2015), participants came from the same geographic area and were differentiated based on their knowledge of a specific local language.

[4]In total, around 240 students were selected according to these criteria, and were distributed in different groups who participated to the larger project with a different timing. Hence, not all of them took part in the experiment being described.

cial background (which might make them more representative of the Italian population than samples of university students typically involved in experiments) and a track record of good grades.[5] Most importantly, the *geographic* representativeness of our sample is an exception in the experimental literature, since participants were living in the 6 cities of origin *at the time of the experiment*.[6] In total, the experiment involved 78 subjects (49 females and 29 males), grouped in four sessions of 19 or 20 subjects each. 42 subjects came from schools in the South of Italy and 36 from schools in the Center-North. [7] While the sample numerosity limits our ability to investigate interaction effects, the main results we later provide are robust to a range of specifications. It is important to consider that the setup we exploited was exogenously given (including its sample size, determined independently from the experimental design) and, to the best of our knowledge, unprecedented.

A crucial feature of our design is that, before participating in the actual randomized experiment, students involved in each session had spent two days living in the same facilities and carrying out structured group activities together (vice-versa, it is unlikely that subjects coming from different schools knew each other before then). The fact that, to Italian speakers, Southern and Northern accents are very easily distinguishable guarantees that, at the time our experiment was run, participants were roughly aware of each

[5]More precisely, our subjects all had a mother not holding a university degree. The sample selection criterion, which encompassed both merit and social background, is explained by the scope of the larger project: the literature on intergenerational transmission of education points at the mother's level of education as particularly relevant (Black et al., 2005; Pronzato, 2012).

[6]Bornhorst et al. (2010), Trifiletti and Capozza (2011) and Bigoni et al. (2018), for instance, look at samples of students originating from different cities/countries, but all enrolled in a same university. Vice-versa, in our case the geographic origin is *proxied* by the school attended – reflecting our desire to consider the environment where each participant was raised.

[7]The six cities involved in the experiment were Cagliari, Napoli, Palermo (including one school in Partinico, part of its metropolitan area) for the South, and Massa, Milano, Prato for the Center-North. Participants from the Center are pooled with those from the North in light of the characteristics of their cities of origin, both located in Toscana. Bigoni et al. (2016), in their selection procedure, classify Toscana in the North based on its latitude. Such choice is reinforced by a look at socioeconomic variables they adopt as proxies for social capital: when compared to the average for Northern Italy, Toscana has higher association density (68.44 per 100,000 inhabitants vs. 36.57, South is at 25.52) and electoral participation (86.67% vs. 86.04%, South is at 70.16%), while it is close to the North for blood donations (42.52 every 1000 inhabitants vs. 47.88, South is at 25.51). Statistics for the South include the island regions of Sicily and Sardinia.

other's origin. On the other hand, at no time during the experiment was any reference made to geographic origin, or to the North-South divide, in order to avoid any framing or Hawthorne effect.[8] It is worth observing that even short-lived social ties are known to be important for public good experiments (Goette et al., 2006, 2012). Social ties formed during the residence could possibly be affected by geographic origin, but given that assignment to groups was randomized in our experiment, this would only reinforce our ability to capture the real life consequences of geographic origin – which include mostly interacting with fellow citizens. Another important feature of our setup is that the students knew they would spend two more days together after the experiment, which is important given the role of reputation in public good games. These two features must be considered in contrast to the total anonymity and "absence of future" which characterise typical laboratory experiments on public good games (Gächter and Fehr, 1999).

Each of the four experimental sessions lasted between 20 and 30 minutes and involved 19 or 20 participants, who were regrouped into four groups. Of such groups, one was composed only by students coming from the schools in the South, one only by students coming from the schools in the Center-North, while the other two had mixed composition: *"being member of a homogeneous group"* is our main treatment variable. The groups were formed ex ante randomly, with the condition that no group would contain two or more students from the same school, and that the sizes of the groups were as homogeneous as possible, given the requirement defined above.[9]

In each session, six rounds of a linear public good game were played. At each round, each participant was given four playing cards, which only she or he could observe. Two of them were red, and were worth one point each; the other two were black, and were worth zero points. Two cards were then collected, covered, from each participant, who could therefore secretly decide to give zero, one or two points (red cards). The total amount of points collected within each group was multiplied by 2 and subdivided between participants of that group. Such points were then added to each participant's "private earnings" - the number of red cards she or he had decided to keep

[8]No reference of participants to geographic origin was recorded, neither during the experiment nor during the debriefing phase.

[9]All groups were designed to have five or six members, but five groups out of sixteen had only four members due to absences. No group had more than three members from a same city (the algorithm used for creating the groups is described in detail in Appendix A).

- so that total earnings for an individual i in a given round t would be:

$$\pi_{i,t} = \underbrace{2 - x_{i,t}}_{\text{Private earnings}} + \underbrace{\frac{2}{\mathcal{N}} \sum_{j=1}^{\mathcal{N}} x_{j,t}}_{\text{Public earnings}}$$

with $x_{i,t}$ being the individual contribution to the public good and $\mathcal{N}$ the group size.

After contributions were recorded, each participant received back her two cards,[10] still hidden from the view of others participants, and the next round began. Earnings were summed across all rounds,[11] and prizes were assigned, in each session, to the three players who had accumulated most points after the six rounds. This choice was made in order to make the tradeoff between contribution and personal gain more salient (see Section 3 for an analysis), and possibly increase the experiment power – in face of the impossibility to increase the sample size.

Instructions were read aloud (always by the same experimenter) at the beginning of each session. Participants were invited to ask questions in case any aspect was unclear. Prizes consisted in gadgets which were described to participants before the start of each session. The fact that they were indivisible objects (rather than, e.g., money) guarantees that agreements based on ex-post sharing of the prize were unfeasible.

2.1 Information and coordination

Information and coordination opportunities available to participants across the six rounds of the game are described below, and summarised in Figure 1.

[10] The mapping between cards and participants was fixed since the beginning, allowing the experimenters both to record private earnings, and to return to each player the contributed cards after each round. For practicality, each participant was assigned four cards with the same number or face, two from a black suit and two from a red suit, e.g. "10 of clubs and diamonds".

[11] By allowing for potential carry-over effects, we are able to study inter-temporal group dynamics more in depth, and in a more natural setting. This design choice is consistent with previous studies on repeated public good games (Andreoni, 1988; Andreoni and Petrie, 2004).

- Initially, the students were sitting in circle in an order, previously determined by the experimenters, satisfying the condition that neighbours were not in the same group. Students were told they had been subdivided in four groups of roughly equal size, which would have remained unchanged for the entire duration of the game, but they did not know who their groupmates were.

- After round 2, the names of members of each group were made public, ensuring, by asking them to raise their hands, that participants of each group had identified each other visually. Participants were then asked not to communicate in any way among them, until further notice.

- After round 4, participants were instructed to sit together with their groupmates, with each group in a different corner of the room, and were given two minutes to discuss among them. The same happened after round 5.[12]

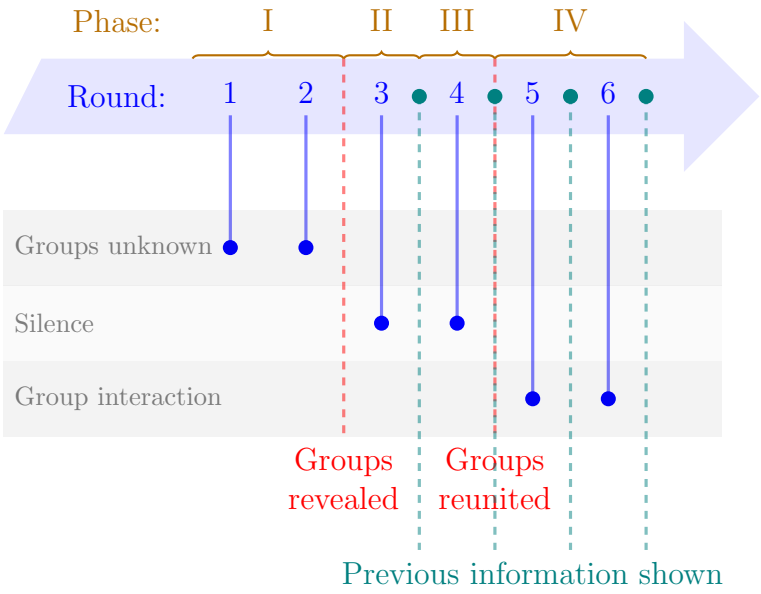
Moreover, after each round starting from the third, information about past contributions was released to participants in two ways: individual contributions from the previous round were read aloud but anonymously, i.e. by referring to the cards owned by each individual rather than to her name, and past results for each group were shown graphically to participants (for an example, see Figure 3 in Appendix C).[13]

These changes in design allow us to investigate the issues of information, anonymity and coordination. It is important to stress that participants were informed since the beginning that they would have received additional information along rounds of the game - but without any further detail, except for the fact that individual contributions would have remained anonymous across all rounds. They also ignored the number of rounds they would play, knowing only that they would be informed before playing the last round.

[12]See Bochet et al. (2006) for a between-subjects analysis of face-to-face interaction in public good games.

[13]This procedure incidentally allowed participants to verify that their contributions were recorded correctly: they were not allowed instead to reveal their cards to anybody else, hence preventing them from providing hard evidence concerning their contributions. Information during the first two rounds was actually more scarce than in most studies on anonymity in public games, in which participants know the composition of their group (Rege and Telle, 2004) or total contributions of their group in previous rounds (Gächter and Fehr, 1999).

Figure 1: Timeline of the experiment



3 Theoretical analysis

The tournament scheme used in the design of payoffs differs from the standard public good game experiment, in that it rewards participants according to their ranking. In what follows we show, through a game-theoretical analysis, that the tournament scheme does not affect the one shot unique optimal strategy - which remains full defection - but can facilitate cooperation in a repeated setting.

Our tournament scheme is a simplified version of the design implemented by Markussen et al. (2014), who introduce competition between groups in a public good game experiment. In their work, competition takes the form of a bonus which is proportional to the standing of the group in the groups ranking. The authors show that, in presence of a large enough bonus, full contribution is a Nash equilibrium, while for smaller values of the bonus (but larger than 0 - the "no competition" case), there are no pure Nash equilibria.

Our design is simpler in the fact that participants are rewarded according to the *individual*, rather than group, ranking. This implies that the tradeoff between improving the group standing and increasing private earnings dis-

appears: contributing is always a dominated strategy. Additionally, in our implementation, contributions are restricted to the set $\{0, 1, 2\}$ rather than being continuous, hence neutralising the possibility of infinitesimal deviations, which is crucial in the proof of absence of Nash equilibria. Summing up, in the one-shot version of our experiment, full defection is the only Nash equilibrium.

The above also applies to a repeated game in which participants receive no information concerning past contributions (i.e. as in the first three rounds of our experiment); moreover, the introduction of communication is irrelevant from a game-theoretical perspective, as long as it is non-binding. In our experiment, however, groups are kept unchanged across all rounds, and participants are provided information on previous contributions, starting from round 4. Thus, the Folk theorem (Friedman, 1971) guarantees the existence of a symmetric equilibrium as long as the condition

$$\delta \geq \frac{\pi^D - \pi^C}{\pi^D - \pi^P} \qquad (*)$$

holds, where δ is the discount factor,[14] and π^D, π^C, π^P denote the payoffs for defection, cooperation and punishment, respectively.

The set of existing equilibria is unchanged by the introduction of a "linear" tournament scheme (in which the ranking bonus is proportional to the individual ranking). This is true, again, because the tournament scheme *at the individual level* introduces no tradeoff, just like in the one-shot case previously described: the objective remains to maximise the "no-tournament" payoff, and condition (*) is unaffected. Instead, the non-linearity of our incentives scheme (the decision to award prizes *only to the three participants with the highest gains* in the session) could in principle matter: for members of under-performing groups, rational individual behaviour is not sufficient to guarantee any positive, albeit small, payoff.

Indeed, the effect of this non-linearity is to decrease π^P with respect to π^C (in our case, assuming that in at least one of the other groups all members are cooperating, $\pi^P = 0$). That is, punishment (non-cooperation) is harsher, making cooperation relatively more appealing. In fact, the proof of the Folk theorem holds *a fortiori* for our design: cooperative symmetric equilibria are qualitatively analogous to the standard public good game.[15]

[14] While in our experiment the game was finite, subjects were ex ante unaware of its length - see Section 2.1.

[15] Importantly, this comparison does not account for a taste for *reciprocity*. The tour-

4 Hypotheses

Design changes

We first analyse the effect of design changes on contribution levels: for this, we need to consider separately the different rounds of the game. Rounds 1 and 2 present the same information setting, and will be analysed together; the same can be said for rounds 5 and 6. Instead, rounds 3 and 4 differ, since before round 4 (but not before round 3) individuals were given aggregated information on their group's contributions history (and they knew that this information would be communicated after each of the following rounds). Hence, we will refer to rounds 1 and 2 as "phase I", round 3 as "phase II", round 4 as "phase III" and rounds 5 and 6 as "phase IV": each phase coincides with a different set of information/coordination opportunities.

In order to test the effect of such changes of setting, we estimate the following model:

$$\begin{aligned} x_{i,t} =& \alpha^f F_i + \alpha_I T_{t,I} + \alpha_{II} T_{t,II} + \alpha_{III} T_{t,III} + \alpha_{IV} T_{t,IV} + \epsilon_{i,t} \\ =& \alpha^f F_i + \sum_{P=I}^{IV} T_{t,P} \alpha_P + \epsilon_{i,t}, \end{aligned} \tag{1}$$

where each phase dummy $T_{t,P}$ takes value 1 if t is in phase P,[16] and F_i takes value 1 if individual i is a female.

In principle, a positive value for phases coefficients could be a spurious consequence of learning effects. However, this is categorically and consistently ruled out by a stylised fact coming from the literature on repeated public goods games: *"provision of the public good 'decays' toward the free riding level with each repetition"*, even when subjects do not know in ad-

nament scheme can cause a *decrease* in contributions should subjects dislike favouring groupmates who contribute strictly less than them. However, results from our experiment suggest that this phenomenon is scarcely relevant, as contributions increase in those rounds in which subjects receive information about their groupmates' (aggregate) contributions. We thank an anonymous reviewer for this observation.

[16] In this and subsequent models, we insert a dummy variable for *each* phase, including the first: coherently, we do not insert in the model a constant term, which would be colinear with them. This choice clearly does not affect the results (we will look at comparisons between coefficients α_P rather than at their individual values, and run significance tests in accordance), and it significantly simplifies the exposition.

vance the length of the game (Andreoni, 1988).[17] Hence, any significant increase in contributions across phases can be considered as (a lower bound to) the effect of the changes in design. Formally, we ascertain whether the changes in design affect contributions by testing the hypothesis

$$\mathcal{H}_0 : \alpha_P = \alpha_{P-1} \tag{HcP}$$

with $P = II, III, IV$.

Group composition

Concerning the treatment variable "belonging to a homogeneous group", denoted as HOM_i, we first test whether members of homogeneous groups exhibit a higher propensity to contribute to the public good: we denote such hypothesis as (Hh).[18]

We then analyse the treatment effect across the different settings by interacting it with phase dummies:

$$x_{i,t} = \beta^f F_i + \sum_{P=I}^{IV} T_{t,P}(\beta_P + \beta_P^h HOM_i) + \epsilon_{i,t}. \tag{2}$$

During phase I subjects do not have any information on their group, and so a treatment effect can be excluded. For each phase P from II to IV, instead, we can first check the effect of design changes on contributions, for mixed groups:

$$\mathcal{H}_0 : \beta_P = \beta_{P-1}, \tag{HmP}$$

and for homogeneous groups:

$$\mathcal{H}_0 : \beta_P + \beta_P^h = \beta_{P-1} + \beta_{P-1}^h. \tag{HhP}$$

We can then check whether phase P features a higher level of contributions in homogeneous compared to mixed groups:

$$\mathcal{H}_0 : \beta_P^h = 0. \tag{HdP}$$

[17] Consistent results, with and without anonymity, are provided by Andreoni and Petrie (2004); Savikhin Samek and Sheremeta (2014).

[18] In this test, we will exclude observations from phase I, when the groups composition is still unknown to participants.

Geographic origin

We want to test whether the propensity to contribute to the public good is related to geographic origin, as predicted by the existing literature. To this aim, we test the hypothesis that the average contribution of individuals from the North and from the South differ. We refer to such hypothesis as (Hn).

In the spirit of phase-treatment interactions presented in Equation (2), we also look at interactions between phases and geographic origin:

$$x_{i,t} = \gamma^f F_i + \sum_{P=I}^{IV} T_{t,P}(\gamma_P + \gamma_P^n N_i) + \epsilon_{i,t}. \tag{3}$$

where the dummy N_i takes value 1 if subject i is from the North. Equation (3) allows us to test the hypothesis $\mathcal{H}_0 : \gamma_P^n = 0$, denoted as (HnP), answering the question of whether the effect of coordination opportunities is heterogeneous across geographic origins.

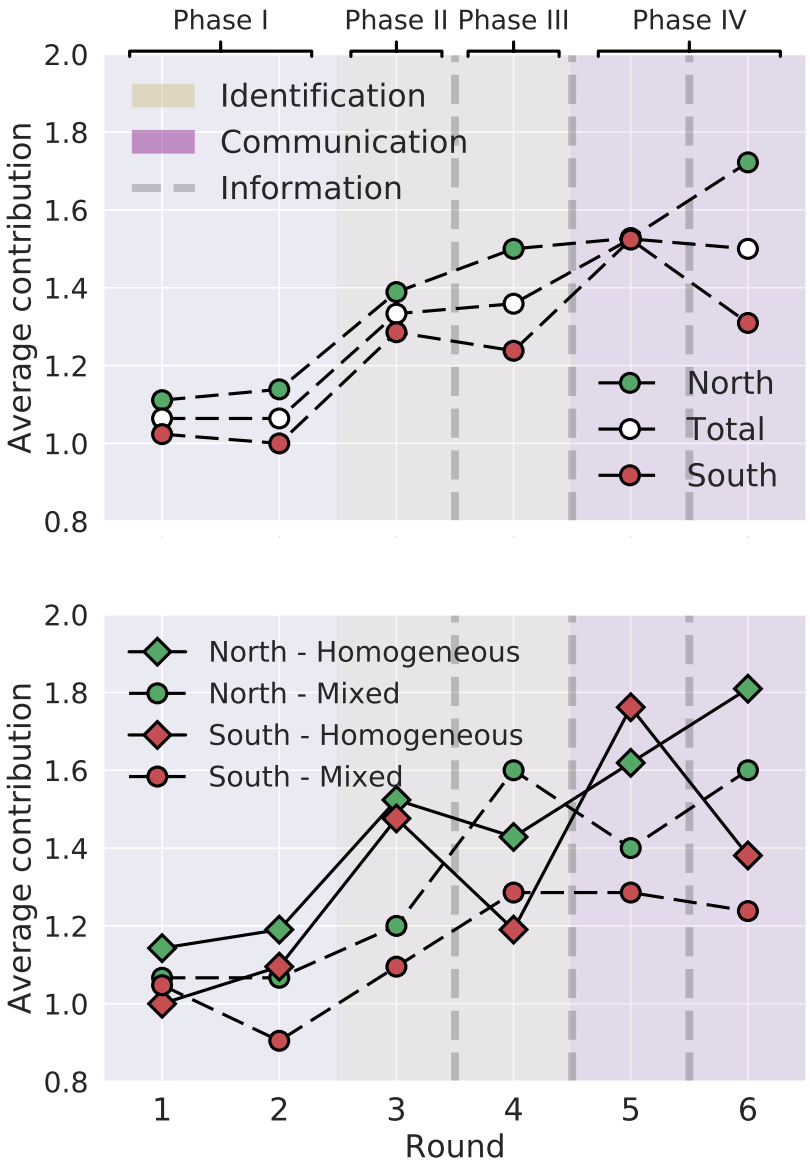
5 Results

The average contribution across all sessions and rounds was 1.308. Figure 2 (top, white dots) displays average contributions in each round: the effect of changes in coordination possibilities is evident between phases I and II, and between phases III and IV. Instead, no evident change can be detected between phases II and III, possibly because the effect of information about group contributions depends on such contributions (e.g. due to conditional cooperation or, conversely, incentives to free ride). In fact, the disaggregation by geographic origin shows that, when transitioning from phase II to phase III, the behaviour differs between individuals from the North and from the South (red and green dots), while the disaggregation by treatment status (Figure 2, bottom) shows an even more pronounced difference: contributions increase in mixed groups and decrease in homogeneous ones.

In what follows, we substantiate these findings by systematically testing hypotheses formulated in Section 4. Estimated coefficients for equations (1), (2) and (3) are presented in Table 1, together with group-level clustered standard errors.[19]

[19]In light of the possibility, for subjects, to influence each others' decisions via communication, or information on past contributions, we also allow for non-independence by

Figure 2: Average contributions per round



Note: Top: average contributions by geographic origin. Bottom: average contributions by geographic origin and group composition.

5.1 Treatment effect

We start by testing the treatment effect (Hh) on individual averages over rounds ($\bar{x}_i$)[20] with a Mann-Whitney test. Analogously to Finocchiaro Castro (2008), we find that $\{\bar{x}_i\}_{HOM_i=1} > \{\bar{x}_i\}_{HOM_i=0}$ (participants in homogeneous groups contribute more), and that the difference is significant ($p = 0.040$).

> **Result 1** *Groups composed by members sharing the same geographic origin contribute to the public good more than mixed groups.*

We show in Section 5.4 that the MW test above is not robust to the exclusion of three female-only groups, but alternative tests of the same hypothesis are.[21]

The already mentioned increase in contributions across phases, which is evident in Figure 2, is per se a nontrivial finding, given the decay in contributions over time consistently observed by the experimental literature on public good games (Andreoni, 1988). Hence, we can infer that changes of setting have an effect in increasing contributions: we now proceed to a more formal analysis of such effect.

What follows is the summary of results concerning the identification of groupmates, which happens in phase II.

- From Equation (1):
  - (HcII): $\alpha_{II} > \alpha_I$ ($p = 0.015$)
- From Equation (2):
  - (HmII): $\beta_{II} = \beta_I$ not rejected ($p = 0.450$)
  - (HhII): $\beta_{II} + \beta^h_{II} > \beta_I + \beta^h_I$ ($p = 0.001$)
  - (HdII): $\beta^h_{II} > 0$ ($p = 0.048$)

running the analysis at the *group*, rather than individual, level (where feasible), and obtain analogous results. See Appendix D.1.

[20] We exclude from this test phase I, when participants did not know the composition of their group: we analyse this phase in Section 5.4 as a robustness test on the randomisation process.

[21] Also see Section 5.3, where we look for, and do not find, a gender effect in contributions.

Table 1: Main results

		Eq. (1) (α)	Eq. (2) (β)	Eq. (3) (γ)
	Female	0.133	0.082	0.149
		(0.089)	(0.081)	(0.086)
Phase	I	0.981***	0.977***	0.913***
		(0.095)	(0.118)	(0.097)
	II	1.250***	0.102***	1.186***
		(0.112)	(0.144)	(0.123)
	III	1.275***	1.380***	1.139***
		(0.098)	(0.103)	(0.107)
	IV	1.429***	1.324***	1.317***
		(0.094)	(0.140)	(0.111)
Phase interacted with treatment	h,I		0.065	
			(0.117)	
	h,II		0.333**	
			(0.155)	
	h,III		-0.135	
			(0.148)	
	h,IV		0.254	
			(0.157)	
Phase interacted with North	n,I			0.126
				(0.081)
	n,II			0.116
				(0.152)
	n,III			0.274*
				(0.131)
	n,IV			0.221**
				(0.102)
	N	468	468	468
	R^2	0.790	0.795	0.794

Note: OLS estimation, dependent variable: $x_{i,t}$ (individual contributions). Row labels indicate coefficients subscripts: phase dummies are in the first block, phase-treatment interactions in the second block, phase-origin interactions in the third block. E.g. the bottom right estimate refers to of γ^n_{IV}, relative to the interaction of dummy variables $T_{t,IV}$ (fourth phase) and N_i (North) in Equation 3. Group-level clustered standard errors in parentheses.

***p< 0.01, **p< 0.05, *p< 0.10.

Identification of group members has a positive and significant effect on contributions (HcII): this is driven by subjects in homogeneous groups (HhII), who contribute significantly more than subjects in mixed groups (HdII), for which no significant change is observed (HmII). Notice that, for members of mixed groups, the identification of groupmates does not reveal, at an aggregate level, new information concerning the group composition: on average, the group has the same share of participants from the South and from the North as the entire population of participants in the session. This can partly explain why the transition from phase I to phase II does not affect mixed groups.

> **Result 2** *Identification of groupmates significantly increases contributions only in groups composed by members sharing the same geographic origin.*

What follows is the summary of results concerning the transition from phase II to phase III, when information on previous contributions is provided.

- From Equation (1):
  - (HcIII): $\alpha_{III} = \alpha_{II}$ not rejected ($p = 0.818$)
- From Equation (2):
  - (HmIII): $\beta_{III} > \beta_{II}$ ($p = 0.075$)
  - (HhIII): $\beta_{III} + \beta^h_{III} = \beta_{II} + \beta^h_{II}$ not rejected ($p = 0.126$)
  - (HdIII): $\beta^h_{III} = 0$ not rejected ($p = 0.374$)

As already suggested by Figure 2 (top, white dots), observing past group performance does not significantly affect *average* contributions (HcIII). In the bottom plot of the figure, we can however observe a sort of rebound effect: contributions in round 4 decrease for homogeneous groups, which were the best performers in round 3 (while the opposite stands for mixed groups). Indeed, while contributions in homogeneous groups decrease, although not significantly (HhIII), they increase significantly for mixed groups (HmIII), reversing the gap between the two categories, which changes sign and is now not significant (HdIII).

There are two plausible explanations for the rebound effect observed in homogeneous groups: best contributors at round 3 (when 52% of individuals in homogeneous groups contributed 2, the maximum possible amount)

may discover that their cooperative effort was not matched, and reduce it (exhibiting conditional cooperation); group members may free-ride on what they see as a sufficiently cooperative group (exhibiting the *opposite* of conditional cooperation). These two explanations are non-exclusive, as they could each apply to different individuals (Fischbacher et al., 2001). This said, the second one seems more relevant. Firstly, it is consistent also with the significant increase observed for *mixed* groups – where members increased their contributions in response to relatively bad group performances. Secondly, a more disaggregated analysis of the data provides support for it: in mixed groups, 76.9% of those who had contributed 1 at round 3 increase their contributions, while in homogeneous groups only 36.8% do so. The behaviour of individuals who had contributed 2 at round 3 is instead similar across the two groups. Thirdly, the second explanation is also consistent with the result of Cárdenas and Mantilla (2015) that, in a public good game also involving between-groups competition, observed group performance is negatively correlated with subsequent individual behaviour. It is important to mention that feedback provided to participants (average group contributions) might have interacted with the payoff scheme (tournament) in ways that our experiment is not designed to entirely unveil. Other groups' contributions were particularly relevant to one's probability to win,[22] and this might have influenced not only the observed rebound, but also the rounds that followed. That is, with our scheme we do not study group behaviour in isolation, but within an environment in which – as in many real world contexts – multiple groups compete among them.

We finally verify how the transition to phase IV (characterised by the possibility to communicate) affects contributions.

- From Equation (1):
  - (HcIV): $\alpha_{IV} = \alpha_{III}$ not rejected ($p = 0.240$)
- From Equation (2):
  - (HmIV): $\beta_{IV} = \beta_{III}$ not rejected ($p = 0.796$)
  - (HhIV): $\beta_{IV} + \beta^h_{IV} > \beta_{III} + \beta^h_{III}$ ($p = 0.012$)
  - (HdIV): $\beta^h_{IV} = 0$ not rejected ($p = 0.127$)

[22] We thank an anonymous reviewer for this observation.

Comparing phase IV with phase III, we do not find a significant increase in average contributions (HcIV). We do find a positive variation for homogeneous groups (HhIV), but not for mixed ones (HmIV). The effect of treatment in phase IV is also non significant (HdIV): in fact, if we rerun our analysis by splitting phase IV into rounds 5 and 6, we find a significant difference ($p = 0.005$) for round 5 only.[23] This is also clearly visible in Figure 2 (bottom) where the transition to round 5 compensates the already mentioned rebound effect for round 4. The suggestive hypothesis that communication in homogeneous groups may be more effective than in mixed groups is consistent with contribution levels observed in phase IV, but not supported by conclusive evidence. Indeed, our estimates do not allow us to causally distinguish the effect of communication from the effect of additional past information. The increase in contributions for homogeneous groups (HhIV) might represent a further rebound rather than actual evidence of a communication effect.

5.2 Effect of geographic origin

The average contribution is 1.398 for Northern subjects and 1.230 for southern subjects. By testing (Hn), we ascertain whether this difference is significant. A Mann-Whitney test on average contributions yields $\{\bar{x}_i\}_{N_i=1} > \{\bar{x}_i\}_{N_i=0}$ ($p = 0.010$). While this result clearly evidences that subjects from the North contribute more than subjects from the South to the public good, its interpretation is nontrivial. Indeed, being in a "North-only group" can have a different effect than being in a "South-only group" on the propensity to contribute, and to the extent to which this is true, the result can be affected by the treatment. In order to isolate the individual-level geographic effect, we hence run the same test restricting the sample to mixed groups (members of which are not affected by differences in treatment), again rejecting the null hypothesis ($p = 0.047$). This evidence is in line with the available experimental literature on the North-South gap in Italy (Bigoni et al., 2016).

Result 3 *Subjects from the North contribute to the public good more than subjects from the South.*

[23] Instead, removing round 6 from phase IV does not affect results for hypotheses (HcIV), (HmIV), and (HhIV). Results are available upon request.

Concerning the analysis of the geographic effect across phases (HnP), we find that γ_P^n is (positive and) significantly different from zero in the last two phases only ($p = 0.141, 0.459, 0.054, 0.048$, respectively). We can hence state the following:

> **Result 4** *The higher level of contributions of subjects from the North is explained by a positive reaction to the introduction of coordination opportunities rather than by a higher propensity to contribute since the first rounds.*

While the first two rounds of our experiment might seem very similar to the experiment of Bigoni et al. (2016), who ran public good game experiments in two cities in the North and two in the South of Italy, an important difference is that, while in their design participants knew they were facing individuals with the same geographic background, this is not true in our case. The distinction is particularly important in light of findings by Bigoni et al. (2018) concerning expectations and conditional cooperation.

Interaction coefficients from Equation 3 could again be affected by a potentially asymmetric treatment effect (being part of a South-/North-only group). Further disaggregating the analysis, by combining the two aspects of group homogeneity and geographic origin, would allow us to tackle this issue and also to verify whether the treatment effect itself (Result 1) is to be attributed in larger part to North-only (South-only) groups. We do so in Equation 5, presented in Appendix D.2; however, the increase in the number of regressors can by itself justify the mostly non-significant results.[24] The two significant interaction coefficients (evidencing higher contributions, in mixed groups, of Northern subjects) are in line with Result 3, and the fact that they refer to the last phases is a confirmation of Result 4.

It might be worth mentioning that the last round taken in isolation features a clear difference ($p = 0.076$) between Northeners and Southeners: the latter feature the typical drop in contributions characterising the last round of public good experiments, where incentives to reputation building disappear (Gächter and Fehr, 1999), while the *opposite* effect can be seen for Northeners. Recall that in our setup, individuals would interact for two

[24]For instance, since there is a significant effect of being assigned to a homogeneous group (results 1 and 2) at the aggregate level, then this must also hold in either North-only or South-only groups (or both), but neither of the respective coefficients is significant.

more days after the experiment, so it is not necessarily surprising that reputation concerns do not disappear at the last round (as typically reported in the literature). A conclusive interpretation of the last round difference across geographic origin would require additional data, but it is tempting to relate it to the stronger betrayal aversion measured by Bigoni et al. (2018) for Southerners. Betrayal aversion has been shown to be related to conditional cooperation (Cubitt et al., 2017). This could also explain why the negative reaction to past information (phase III) seems to be stronger for Southerners than for Northerners.

5.3 Contributions and gender

In the literature on public good games, there is some evidence of a higher propensity to contribute on behalf of females (Nowell and Tinkler, 1994). Estimates reported above have been obtained controlling for a potential gender effect; in the present section, we explicitly verify whether females and males have a different propensity to contribute to the public good, and whether they react differently to the treatment and/or to the changes in information setting.

We can test the presence of a gender effect by running a Mann-Whitney test on average contributions of females versus males: the result is not significant ($p = 0.106$). To control for the possibility of asymmetric treatment effects, we run the same test restricting to mixed groups (like we did for hypothesis (Hn)): the result is again not significant ($p = 0.180$).

For what concerns the gender component *of the treatment effect*, we can estimate the following equation (analogous to Equation 5 in Appendix D.2, but with geographic origin replaced by gender):

$$x_{i,t} = \sum_{P=I}^{IV} T_{t,P}(\delta_P + \delta_P^h HOM_i + F_{i,P}(\delta_P^f + \delta_P^{hf} HOM_i)) + \epsilon_{i,t}. \qquad (4)$$

See Table 6 in Appendix D.2 for estimation results. By testing $\mathcal{H}_0 : \delta_P^f + \delta_P^{hf} = 0$ for $P = I, II, III, IV$, we verify whether, in homogeneous groups, females behave differently than males. Results are never significant ($p = 0.484, 0.739, 0.604, 0.514$): that is, we find no evidence of a gender difference in the effect of the treatment.

5.4 Robustness

Result 1, concerning the effect of the treatment, is supported by a non-parametric Mann-Whitney test: alternatively, we can test parametrically the joint significance of phase-treatment interaction dummies (excluding phase I) in Equation 2: we do so through a Wald test, and again reject the null hypothesis of no difference ($p = 0.079$). Analogously, Result 3, concerning the effect of geographic origin, is confirmed by looking at the joint significance of phase-origin interaction dummies in Equation 3 ($p = 0.054$). The same analysis yields similar results when run at the group level (see Appendix D.1).

As already mentioned in Section 4, Equation 2 is not expected to yield interesting insights concerning phase I: at that time, subjects did not know who their groupmates were, and hence their contribution could not be affected by being in a homogeneous or mixed group. If homogeneous and mixed group members had differed in their contribution levels already before the group composition was made public, this would have represented an alarming signal of ex ante differences between the two samples. Instead this is not the case, as the null hypothesis that $\beta_I^h = 0$ cannot be rejected ($p = 0.586$).

Female participants outnumbered male participants in all sessions. We both allowed for a gender effect in our estimates, and explicitly looked at a gender component of the treatment effect in Section 5.3, without finding any. This said, the randomisation algorithm resulted for session 2 in a significantly unbalanced composition of homogeneous groups, which included no male participants (see Appendix C). We hence re-estimate hypothesis (Hh), again with a Mann-Withney test on individual averages, but this time excluding such groups from the analysis, and still find a significantly positive treatment effect ($p = 0.079$). If we also drop a third female-only group (South-only group in session 3), the result of the MW test becomes non-significant ($p = 0.173$), while the analysis at the group level mentioned above still finds the phase-interaction dummies to be jointly significant. This discrepancy is likely to result from the sample reduction due to excluding 3 homogenous groups out of 8: indeed, the "Female" dummy is never significant.

As mentioned in Section 2, the size of groups varied from four to six. In principle, the size of a participant's group can influence her behaviour, as the individual share of "public earnings" resulting from each contributed red card is, according to Equation 2, $\frac{1}{\mathcal{N}}$ (or in other terms, the marginal loss from each contributed red card is $\frac{\mathcal{N}-1}{\mathcal{N}}$). While in the first phase individuals

ignored the composition of their groups, we verify that this does not affect our results for the following phases by rerunning equations (1), (2) and (3) including the group size as regressor. Such variable is never significant, and other results are qualitatively unchanged.

Finally, we re-run our tests with random effects estimation and, when meaningful (that is, for those hypotheses, based on equations (1) and (2), which involve a cross-period comparison), fixed effects. Results are virtually unchanged.

6 Conclusions

We run a repeated public good game with participants coming from different Italian cities. Differently from the existing experimental literature on inter-regional gaps in social capital, our analysis focuses on the *interaction* between subjects with different geographic background and *only temporarily* abstracted from their respective cities of origin. By manipulating the composition of groups, we compare the level of cooperation of those composed only by individuals sharing the same origin, on the one side, and mixed ones, on the other. At the same time, we explore the extent to which changes in coordination opportunities can lead an individual to contribute more. Moreover, we compare the contribution patterns of individuals with different geographic backgrounds.

We find that groups composed both by subjects from the North and from the South of Italy perform significantly worse than homogeneous groups (Result 1), mainly because of the different impact of coordination opportunities, and in particular of identification (Result 2). As already reported in the literature, individuals from the North contribute more than individuals from the South (Result 3): this is explained by a different reaction to information concerning past contributions and to the possibility to communicate (Result 4). Instead, geographic origin does not significantly predict contributions in the first phase of the experiment: hence, there is no evidence of a difference in the ex ante propensity to contribute. In general, the introduction of coordination opportunities has a strongly positive effect, which more than counterbalances the expected decay of contributions over time. Finally, we find no significant difference between males and females, neither in average contributions nor in their reaction to the treatment.

Our results shed new light on the problem of the North-South divide in

Italy. Result 3 reinforces the conclusions of Bigoni et al. (2016) that the gap *"appears to lie in the ability to cooperate"*. In particular, Result 4 points out that the level of contributions crucially depends on the effectiveness of coordination: since geographic origin is not a significant predictor of contributions *at the beginning* of the experiment (when participants act in isolation), we suggest that a different propensity to *build* mutual trust might play a fundamental role in the North-South divide. At the same time, Results 1 and 2 highlight the difficulty in cooperation *across* the North-South divide: such difficulty could represent a cause of path dependence (historically, the gap in socioeconomic indicators does not seem to vanish over time, rather the opposite) and a further obstacle to economic and social development.

Summing up, our results reinforce the view that the North-South gap in social capital cannot be imputed only to differences in institutions and opportunities, given the different reactions observed to the same incentives.[25] Hence, such gap cannot be levelled by only focusing on institutional settings: in the long term, behaviour can certainly react to institutional determinants, but such reactions might be too slow. Most importantly, institutions themselves are composed of citizens, and any attempt at shaping them must take this into account. Interestingly, like most of the Italian population, the typical participant in our experiment had relatively few occasions to enter in relation with compatriots from the other side of the peninsula: the literature on the positive effect of diversity on group performance (McLeod and Lobel, 1992; Lazear, 1999; Hong and Page, 2001) is well aware of the problem of communication costs, and suggests that policymakers should work in the direction of integration and mutual knowledge. These should be considered among the main objectives when dealing with socioeconomic differences across regions, and as a viable way to increase the level of social capital in countries characterised by strong heterogeneities.

Further studies could analyse more in depth the interaction of the treatment with geographic origin. Indeed, we do not find a significant effect of being in a homogeneous group *conditional* on being from the North/South: such level of detail might be achieved with additional experimental evidence. Such evidence would allow for instance to state whether the worse performance of mixed groups can be imputed more to Northerners or to Southerners, or whether individuals from the North contribute less when they are in

[25] While the incentives might have a different desirability for different groups of subjects, this does not alter the policy implications.

mixed groups than when they are in North-only groups. We think that these are important issues to consider for the understanding of the North-South economic divide, and that they are an interesting venue for future research, together with the study of other brackets of the population, and with the possible interplay between geographic origin and competitive attitude.

References

Algan, Y., C. Hémet, and D. D. Laitin (2016). The social effects of ethnic diversity at the local level: A natural experiment with exogenous residential allocation. *Journal of Political Economy 124*(3), 696–733.

Andreoni, J. (1988). Why free ride?: Strategies and learning in public goods experiments. *Journal of Public Economics 37*(3), 291–304.

Andreoni, J. and R. Petrie (2004). Public goods experiments without confidentiality: a glimpse into fund-raising. *Journal of Public Economics 88*(7), 1605–1623.

Balliet, D., J. Wu, and C. K. De Dreu (2014). Ingroup favoritism in cooperation: A meta-analysis. *Psychological Bulletin* (140), 1556–1581.

Banfield, E. C. (1967). *The moral basis of a backward society*. Free Press.

Barrett, L. (2007). *Oxford Handbook of Evolutionary Psychology*. Oxford University Press, USA.

Bigoni, M., S. Bortolotti, M. Casari, and D. Gambetta (2018). At the root of the north–south cooperation gap in italy: Preferences or beliefs? *The Economic Journal 129*(619), 1139–1152.

Bigoni, M., S. Bortolotti, M. Casari, D. Gambetta, and F. Pancotto (2016). Amoral Familism, Social Capital, or Trust? The Behavioural Foundations of the Italian North-South Divide. *The Economic Journal 126*(594), 1318–1341.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005). Why the apple doesn't fall far: Understanding intergenerational transmission of human capital. *The American Economic Review 95*(1), 437–449.

Bochet, O., T. Page, and L. Putterman (2006). Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior & Organization 60*(1), 11–26.

Bornhorst, F., A. Ichino, O. Kirchkamp, K. H. Schlag, and E. Winter (2010). Similarities and differences when building trust: the role of cultures. *Experimental Economics 13*(3), 260–283.

Brosig-Koch, J., C. Helbach, A. Ockenfels, and J. Weimann (2011). Still different after all these years: Solidarity behavior in East and West Germany. *Journal of Public Economics 95*(11), 1373–1376.

Buonanno, P., D. Montolio, and P. Vanin (2009). Does social capital reduce crime? *Journal of Law and Economics 52*(1), 145–170.

Camussi, S., A. L. Mancini, and P. Tommasino (2018). Does Trust Influence Social Expenditures? Evidence from Local Governments. *Kyklos 71*(1), 59–85.

Cappelletti, D., L. Mittone, M. Ploner, et al. (2015). Language and intergroup discrimination. evidence from an experiment. *CEEL Working Paper N. 1504, University of Trento, Department of Economics*.

Cárdenas, J. C. and C. Mantilla (2015). Between-group competition, intragroup cooperation and relative performance. *Frontiers in Behavioral Neuroscience 9*, 33.

Cason, T. N., T. Saijo, and T. Yamato (2002). Voluntary participation and spite in public good provision experiments: an international comparison. *Experimental Economics 5*(2), 133–153.

Cassar, A. and B. Wydick (2010). Does social capital matter? evidence from a five-country group lending experiment. *Oxford Economic Papers 62*(4), 715–739.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics 14*(1), 47–83.

Cox, T. H., S. A. Lobel, and P. L. McLeod (1991). Effects of Ethnic Group Cultural Differences on Cooperative and Competitive Behavior on a Group Task. *The Academy of Management Journal 34*(4), 827–847.

Croson, R. and N. Buchan (1999). Gender and culture: International experimental evidence from trust games. *The American Economic Review 89*(2), 386–391.

Cubitt, R., S. Gächter, and S. Quercia (2017). Conditional cooperation and betrayal aversion. *Journal of Economic Behavior & Organization 141*, 110–121.

Eckel, C. C. and R. K. Wilson (2006). Internet cautions: Experimental games with internet partners. *Experimental Economics 9*(1), 53–66.

Felice, E. (2013). *Perché il Sud è rimasto indietro*. Il Mulino.

Fershtman, C. and U. Gneezy (2001). Discrimination in a segmented society: An experimental approach. *The Quarterly Journal of Economics 116*(1), 351–377.

Finocchiaro Castro, M. (2008). Where are you from? cultural differences in public good experiments. *The Journal of Socio-Economics 37*(6), 2319–2329.

Fischbacher, U., S. Gächter, and E. Fehr (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economics Letters 71*(3), 397–404.

Friedman, J. W. (1971). A non-cooperative equilibrium for supergames. *The Review of Economic Studies 38*(1), 1–12.

Gächter, S. and E. Fehr (1999). Collective action as a social exchange. *Journal of Economic Behavior & Organization 39*(4), 341–369.

Gaudeul, A. and C. Giannetti (2015). Privacy, trust and social network formation. *Working Paper 269, University of Goettingen, Department of Economics*.

Gibson, J., D. McKenzie, H. Rohorua, and S. Stillman (2015). The Long-Term Impacts of International Migration: Evidence from a Lottery. *World Bank Economic Review 32*(1), 127–147.

Goette, L., D. Huffman, and S. Meier (2006). The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *American Economic Review 96*(2), 212–216.

Goette, L., D. Huffman, and S. Meier (2012). The impact of social ties on group interactions: Evidence from minimal groups and randomly assigned real groups. *American Economic Journal: Microeconomics 4*(1), 101–15.

Guiso, L., P. Sapienza, and L. Zingales (2004). The role of social capital in financial development. *The American Economic Review 94*(3), 526–556.

Helliwell, J. F. and R. D. Putnam (1995). Economic Growth and Social Capital in Italy. *Eastern Economic Journal 21*(3), 295–307.

Hong, L. and S. E. Page (2001). Problem solving by heterogeneous agents. *Journal of Economic Theory 97*(1), 123–163.

Hoyman, M., J. McCall, L. Paarlberg, and J. Brennan (2016). Considering the Role of Social Capital for Economic Development Outcomes in US Counties. *Economic Development Quarterly 30*(4), 342–357.

Ichino, A. and G. Maggi (2000). Work environment and individual background: explaining regional shirking differentials in a large Italian firm. *The Quarterly Journal of Economics 115*(3), 1057–1090.

Knack, S. and P. Keefer (1997). Does social capital have an economic payoff? a cross-country investigation. *The Quarterly Journal of Economics 112*(4), 1251–1288.

Lazear, E. P. (1999). Globalisation and the market for team-mates. *The Economic Journal 109*(454), 15–40.

Leonardi, R. (1995). Regional Development in Italy: Social Capital and the Mezzogiorno. *Oxford Review of Economic Policy 11*(2), 165–179.

Levine, S. S., E. P. Apfelbaum, M. Bernard, V. L. Bartelt, E. J. Zajac, and D. Stark (2014). Ethnic diversity deflates price bubbles. *Proceedings of the National Academy of Sciences 111*(52), 18524–18529.

Markussen, T., E. Reuben, and J.-R. Tyran (2014). Competition, cooperation and collective choice. *The Economic Journal 124*(574), F163–F195.

McLeod, P. L. and S. A. Lobel (1992). The effects of ethnic diversity on idea generation in small groups. In *Academy of Management Proceedings*, Volume 1992, pp. 227–231. Academy of Management.

Nowell, C. and S. Tinkler (1994). The influence of gender on the provision of a public good. *Journal of Economic Behavior & Organization 25*(1), 25–36.

Nuti, S. and A. Ghio (2017). *Obiettivo mobilità sociale*. Il Mulino.

Ockenfels, A. and J. Weimann (1999). Types and patterns: an experimental East-West-German comparison of cooperation and solidarity. *Journal of Public Economics 71*(2), 275–287.

Oosterbeek, H., R. Sloof, and G. Van De Kuilen (2004). Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics 7*(2), 171–188.

Pronzato, C. (2012). An examination of paternal and maternal intergenerational transmission of schooling. *Journal of Population Economics 25*(2), 591–608.

Putnam, R. (2001). Social capital: Measurement and consequences. *Canadian Journal of Policy Research 2*(1), 41–51.

Putnam, R. D., R. Leonardi, and R. Y. Nanetti (1993). *Making democracy work: Civic traditions in modern Italy*. Princeton University Press.

Rege, M. and K. Telle (2004). The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics 88*(7), 1625–1644.

Rustagi, D., M. Veronesi, et al. (2016). Social identity, attitudes towards cooperation, and social preferences: Evidence from Switzerland. *Working Paper N. 01/2016, University of Verona, Department of Economics*.

Savikhin Samek, A. and R. M. Sheremeta (2014). Recognizing contributors: an experiment on public goods. *Experimental Economics 17*(4), 673–690.

Tajfel, H., J. C. Turner, W. G. Austin, and S. Worchel (1979). An integrative theory of intergroup conflict. *Organizational identity: A reader*, 56–65.

Trifiletti, E. and D. Capozza (2011). Examining group-based trust with the investment game. *Social Behavior and Personality: an international journal 39*(3), 405–409.

A Algorithm for the creation of groups

The following algorithm was implemented to subdivide participants of each session in four groups. Importantly, in each session, each school was represented by a maximum of 3 students.

1. Create three empty lists: $\mathcal{S}$(outh) with 6 slots, $\mathcal{N}$(orth) with 6 slots, $\mathcal{M}$(ixed) with 12 slots. A slot is occupied whenever a student is appended to a list.
2. If the session has strictly less participants from the North (South), remove one slot to the $\mathcal{N}$ ($\mathcal{S}$) list, respectively.
3. Let I be the school with the most students among schools still not processed.
4. Let $\mathcal{L}$ be the list $\mathcal{S}$ if the school is from the South, $\mathcal{N}$ otherwise.
5. If $\mathcal{L}$ has a free slot, append a randomly selected student from I to it.
6. If there are still students to be placed from I, append them to $\mathcal{M}$.
7. If there are still schools to be processed, go back to point 3.
8. Create two lists $\mathcal{M}1$ and $\mathcal{M}2$ from elements of $\mathcal{M}$ in odd and even positions, respectively.

The rationale for ordering schools by size was to guarantee that no two students from the same school would end up in the same group (i.e. that schools with more students, and hence more difficult to place, would get their students assigned first).

For completeness, the following table presents a brief overview of the schools involved:

Table 2: Subdivision of participants across schools and cities

City	Area	Students
Massa	North	8
Massa	North	4
Milano	North	12
Milano	North	10
Prato	North	8
Prato	North	9
Prato	North	12
Cagliari	South	13
Cagliari	South	10
Napoli	South	14
Napoli	South	8
Palermo	South	11
Palermo (Partinico)	South	12

B Instructions

Participants in each session spent the entire time of the experimental session sitting in circle in the same room. In order to limit communication to the phases designed for this (see Section 2.1), subjects were, since the beginning, instructed not to talk among them about the experiment, and to direct any question to the experimenter in charge of explaining the activity (the experimenter was always the same across all sessions). For the same reason, instructions and control questions were not provided in written form, and were instead verbally issued (in Italian).

- Initially, subjects entered the room together. The experimenter and a helper (the same across all sessions) were present. A number of chairs, one for each participant, were distributed in circle around the room, except for a portion of one wall, where a table was placed, holding the material used for the experiment. Such material included a computer connected to a projector (where no image was projected initially), the playing cards, and an empty box with a small opening in its top.

  "On each of these chairs you will find a note with a name written on

it: please find your name and sit on the corresponding chair. Please do not discuss this game with other participants until its end. If you have any questions, please ask them to me. You can do so at any time."

- After all participants had sat down, the playing cards used for communicating their decision were distributed, according to a predefined random mapping. The helper went distributing the cards in circle, verifying the names of participants against the list he had.

  "My colleague will now assign four cards, covered, to each of you. We will explain you in few moments what use you will make of these cards. It is important that they remain secret: one basic rule of this game is that nobody, at no time, must see your cards."

- After all playing cards had been distributed, the functioning of the activity was explained.

  "Please now look at your cards, while keeping their face hidden from the view of other participants. You will notice that you have two cards with a red suit, and two with a black suit. The red cards are worth one point, the black cards are worth zero points. You will also notice that the two cards of a black suit come from different decks: one has a blue back, one has a red back. The same applies to the cards of a red suit."

  "The game will be composed of several rounds. At each round, we will collect two cards, covered, from each of you. You all have the possibility, and can freely decide, to give zero, one or two cards from a red suit, that is, to contribute zero, one or two points. The points you will not contribute will remain to you. Please give one card with a blue back and one with a red back."

  "Although their composition is, at this point, unknown to you, participants in this room have been subdivided in four groups of approximately the same size. The points you decide to contribute when we collect the cards will flow into a common fund – one for each group. The content of this fund will be multiplied by two and subdivided in equal shares between members of the group. So, at the end of each round, each of you will obtain the points he or she decided to keep, plus the total number of points that your team members decided to contribute, multiplied by two, and divided by the number of team members."

- Subjects were then proposed the following three test questions in order to ensure they had correctly understood. For each question, after one or more participants had answered correctly, the experimenter would explain the answer to all participants, and the steps (calculations) required to obtain it.

    1. *"Let us assume that you are in a group of four people and that, at a given round, each of you decides to keep both points: how many points does each of you get for that round?"*
    2. *"What if, in this same group, each of you decides to contribute both points?"*
    3. *"What if in this same group you decide to keep both points, but the three other group members decide each to contribute both points?"*

- Finally, information was given about the rounds and payoffs.

  "You now know the basic functioning of the game. The points you make in each round will be added up and, at the end of the game, the three people with most points will win a prize. You don't know now the number of rounds which you will play, but I will inform you before playing the last round. Moreover, I will provide you with other pieces of information in the following rounds, but details will be provided at that time only."

- The experiment then started. Each step of the game was guided by the experimenter. *"Now my colleague will pass among you with this box. Please insert in its opening the two cards you decide to give."*

  "Now I will register your choices on the computer."

  (Notice that the spreadsheet on which this was done was *not* shown on the projector screen.)

  "Now my colleague will return you your two cards: you can verify they are indeed the two you gave. We can then proceed to the next round."

- After the first two rounds, absolute silence was requested:

  "From this moment, and until further notice, please abstain from any talking. If you have any question, please raise your hand an I will come and answer it."

Immediately after, the composition of the groups was revealed, one at a time:

"Please Eeeeee Ffffff, Ggggg Hhhhhh, Iiiiii Jjjjjj and Kkkkkk Llllll now raise your hand. You are members of group X. Please make sure you all identify each other, while remaining silent. Please now lower your hand."

(notice participants were called by name and surname, while X, the number of the group, was one of A, B, C, or D).

- After each round starting from the third, information about past contributions was shown on the projector screen (see Figure 3 for an example referring to round 5).

  "These are the average contributions so far: each line refers to a group"

- After round four, members of each group were invited to discuss among them:

  "From now on, you can talk again. Please, members of group A move to this corner of the room, members of group B to this other corner, members of group D to this other corner, and members of group D to this last corner. I will give you exactly two minutes, starting from now, to freely discuss with your group members, then we will proceed with the next round of the game. Remember that it is still forbidden to show your cards to any other participant."

- Finally, before round six, subjects were informed that it would be the last round.

  "We are now going to play round six, which will be the last round of this game."

- A debriefing would follow, in which participants were invited to comment the experiment and the effects of differences across rounds. The experimenter would mention some real world examples of public goods, and comment on the fact that even non-binding mechanisms such as identification and communication can have important consequences on economic decisions made by individuals within societies.

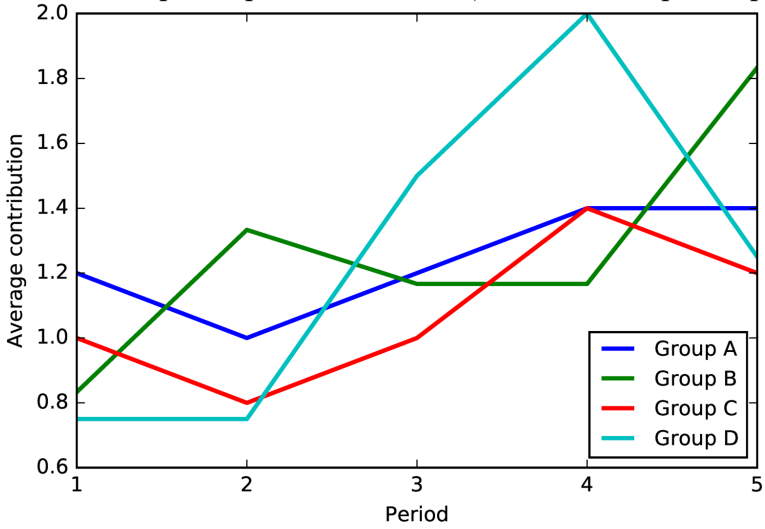
C Additional material

Figure 3 features an example of how information about past group contributions was shown to participants (from round 3 onwards).

Table 3 provides some descriptive statistics for participants: for each session, we show the distribution of individual characteristics (geographic origin/gender) based on the assignment of individuals to the treatment. T-tests ran on the each session fail to reject the null of identical distribution between the two categories, with the exception of Session 2 ($p = 0.001$), in which homogeneous groups were composed only of female participants (we take this into account in Section 5.4).

Table 4 provides information about the 12 prize winners (three for each session). For comparison, the minimum possible earning was 2 and the maximum 32 (see Equation (2)). The signs of deviations between the shares of winners and the shares of sample presenting each feature are in line with results presented in the main text (females contribute more, although not significantly, "North-only" groups perform better, although not significantly, homogeneous groups perform better).

Figure 3: Example of past information, as shown to participants



Note: Information shown to participants of session 1 before the last round (labels translated from Italian).

Table 3: Descriptive characteristics

		Female		**North**	
Session	**Treatment**	**0**	**1**	**0**	**1**
1	0	5	4	6	3
	1	3	8	6	5
2	0	6	3	4	5
	1	0	10	4	6
3	0	5	4	6	3
	1	3	8	6	5
4	0	4	5	5	4
	1	3	7	5	5

Table 4: Descriptive characteristics of winners

Session	**Rank**	**Female**	**North**	**Treatment**	**Total gain**
1	1	0	0	1	22
1	2	1	1	1	21.4
1	3	0	1	0	21
2	1	1	0	1	24
2	2	1	1	1	23
2	3	0	1	0	22.5
3	1	0	0	0	23.6
3	2	1	0	1	22.7
3	3	1	0	1	22.7
4	1	1	1	1	22.8
4	2	0	1	1	22.8
4	3	1	0	0	22.5
Winners (share):		**7 (58%)**	**6 (50%)**	**8 (67%)**	
Share of all participants:		63%	46%	54%	

D Supplementary results

D.1 Analysis of group averages

The following table replicates the analysis of phase- and treatment-effects of Table 1 (columns (1) and (2), respectively) at the group, rather than individual, level (analysis of gender and geographic origin is omitted because these characteristics vary within groups).

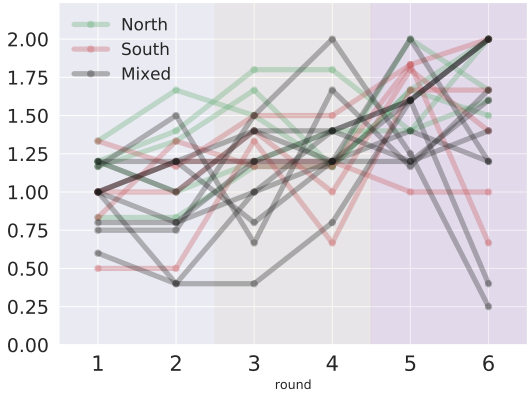
Table 5: Main results, group averages

		Eq. (1) (α)	Eq. (2) (β)
Phase	I	1.057***	1.009***
		(0.0667)	(0.102)
	II	1.327***	1.144***
		(0.0944)	(0.143)
	III	1.365***	1.419***
		(0.0944)	(0.106)
	IV	1.494***	1.350***
		(0.0667)	(0.150)
Phase interacted with treatment	h,I		0.0948
			(0.130)
	h,II		0.367**
			(0.166)
	h,III		-0.108
			(0.149)
	h,IV		0.287
			(0.169)
	N	96	96

Note: OLS estimation, dependent variable: average of $x_{i,t}$ at the group/period level.
Row labels indicate coefficients subscripts: phase dummies are in the first block, phase-treatment interactions in the second block.
Group-level clustered standard errors in parentheses.
***p< 0.01, **p< 0.05, *p< 0.10.

As a robustness test for results 1 and 2, we also replicate tests presented

Figure 4: Average contributions per round, disaggregated by group



in Section 5, confirming all findings (with the only – unrelated – exception of (HmIII), for which we now obtain $p = 0.137$).

Figure 4 is the equivalent of Figure 2, disaggregated at the group level. While variability increases significantly (each point only represents the mean of four to six individual contributions), it allows to observe that all groups are generally increasing , and that no obvious outliers emerge – with the exception of the last round (discussed at the end of Section 5.2), in which a few groups drop their contribution levels significantly, as compared to the other groups.

D.2 Phase-treatment interaction

In what follows, we combine Equations 2 and 3, interacting phase and treatment dummies with the geographic origin of participants.

$$x_{i,t} = \zeta^f F_i + \sum_{P=I}^{IV} T_{t,P}(\zeta_P + \zeta_P^h HOM_i + N_{i,P}(\zeta_P^n + \zeta_P^{hn} HOM_i)) + \epsilon_{i,t}. \quad (5)$$

Hypotheses (HdII), (HdIII) and (HdIV) allowed us to investigate whether being in a homogeneous group (instead of a heterogeneous one) has an effect on contributions. The estimation of Equation (5) can help us verify if there is a treatment effect *conditional on the geographic origin of individuals*.

Table 6: Additional estimation results

		Eq. (4) (δ)			Eq. (5) (ζ)
				Female	0.104
					(0.076)
Phase	I	0.950***	Phase	I	0.926***
		(0.179)			(0.108)
	II	1.250***		II	1.046***
		(0.195)			(0.195)
	III	1.350***		III	1.236***
		(0.198)			(0.152)
	IV	1.225***		IV	1.212***
		(0.158)			(0.155)
Phase interacted with treatment	h,I	0.272	Phase interacted with treatment	h,I	0.032
		(0.263)			(0.156)
	h,II	0.194		h,II	0.341
		(0.277)			(0.224)
	h,III	-0.128		h,III	-0.135
		(0.312)			(0.203)
	h,IV	0.331		h,IV	0.270
		(0.232)			(0.203)
Phase interacted with Female	f,I	0.144	Phase interacted with North	n,I	0.098
		(0.229)			(0.088)
	f,II	-0.250		n,II	0.113
		(0.227)			(0.258)
	f,III	0.150		n,III	0.322*
		(0.293)			(0.155)
	f,IV	0.306**		n,IV	0.246*
		(0.115)			(0.134)
Phase interacted with F. and t	hf,I	-0.290	Phase interacted with N. and t.	hn,I	0.036
		(0.307)			(0.172))
	hf,II	0.321		hn,II	-0.050
		(0.308)			(0.309)
	hf,III	-0.039		hn,III	-0.069
		(0.361)			(0.244)
	hf,IV	-0.195		hn,IV	-0.088
		(0.202)			(0.201)
	N	468		N	468
	R^2	0.797		R^2	0.798

Note: OLS estimation, dependent variable: $x_{i,t}$ (individual contributions). Row labels indicate coefficients subscripts: see the description of each block for the interpretation of the coefficients. Group-level clustered standard errors in parentheses.

***p< 0.01, **p< 0.05, *p< 0.10.

Namely, we can answer such question by running the following joint tests on coefficients presented in Table 6:

- $\mathcal{H}_0 : \zeta_P^h + \zeta_P^{hn} > 0$ for individuals from the North,
- $\mathcal{H}_0 : \zeta_P^h > 0$ for individuals from the South,

for each phase $P = II, III, IV$. From such tests, no significant differences emerge ($p = 0.183, 0.239, 0.259$ for the North, $0.149, 0.517, 0.204$ for the South, respectively).

By exploiting the disaggregation along the dimension of geography, we can also compare North-only and South-only groups between them. This is done by testing $\mathcal{H}_0 : \zeta_P^n + \zeta_P^{hn} > 0$ for each phase $P = II, III, IV$.[26] Results do not suggest that people from the North act differently from people from the South in homogeneous groups ($p = 0.703, 0.191, 0.306$, respectively).

By running the same analysis for mixed groups, we can instead compare the behaviour of Southerners and Northerners subject to the *same* treatment (i.e. being in a mixed group) in each phase.[27] Namely, we test $\mathcal{H}_0 : \zeta_P^n > 0$ for each phase $P = II, III, IV$: in line with Result 3, in mixed groups we find a higher level of contributions on behalf of Northeners compared to Southeners, for two phases out of three ($p = 0.669, 0.055, 0.087$, respectively).

In conclusion, while we confirm the higher level of contributions of Northeners (Result 3) in mixed groups, we find no evidence that the treatment effect is related to the geographic origin of subjects. That is, we cannot explain Results 1 and 2 as the consequence of an interaction between the treatment and the geographic origin. However, we cannot exclude the possibility that such "non-result" is due to the low numerosity of observations in each of the subsamples considered.

[26] It is worth stressing that such tests pool together an intrinsic feature (the geographic origin) and a possible treatment effect (being in a North-only or South-only group).

[27] We had already tested Hypothesis (Hn) on such a subsample, but pooling together all phases, that is, not looking for an effect of design changes.